UNIVERSITY OF CALIFORNIA
RIVERSIDE

Exploring Immune Cell Heterogeneity Through Single-Cell RNA Sequencing Analysis

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics & Bioinformatics

by

Xinru Qiu

June 2023

Dissertation Committee:

Dr. Adam Godzik, Chairperson
Dr. Meera G. Nair
Dr. Zhenyu Jia

The Dissertation of Xinru Qiu is approved:

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

Throughout my doctoral studies and the writing of this dissertation, I have received a great deal of support and assistance from my mentors, collaborators, family and friends.

I would like to express my deepest appreciation to my principal investigator (PI), Dr.Adam Godzik, for his unwavering support and guidance throughout my PhD journey. Their expertise, dedication, and patience have been instrumental in shaping my research and nurturing my development as a scholar. I am truly fortunate to have had the opportunity to work under their mentorship.

My heartfelt thanks also go out to my lab members, who have not only been exceptional colleagues but also friends throughout this journey. Dr.Lukasz Jaroszewski, Mayya Sedova, Dr.Arghavan Alisoltani, Zhanwen Li, Dr. Mallika Iyer, the camaraderie and collaboration have made my time in the lab an enjoyable and enriching experience.

I would like to extend my appreciation to my thesis committee members, Dr.Adam Godzik, Dr. Meera G. Nair, and Dr. Zhenyu Jia, for their insightful feedback and valuable suggestions. The expertise and constructive criticism have played a pivotal role in refining my research and ensuring its success.

I am grateful to my collaborators, for their contributions and dedication to our joint projects. Our collaborations have not only expanded my knowledge but have also provided me with the opportunity to learn from your perspectives and experiences.

Last but certainly not least, I would like to express my eternal gratitude to my parents, for their support, and encouragement throughout my life. The faith in my abilities has been my constant source of strength and motivation, and I dedicate this thesis to you.

ABSTRACT OF THE DISSERTATION

Exploring Immune Cell Heterogeneity Through Single-Cell RNA Sequencing Analysis

by

Xinru Qiu

Doctor of Philosophy, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, June 2023
Dr. Adam Godzik, Chairperson

Over the past decade, single-cell RNA sequencing (scRNA-seq) has revolutionized the field of transcriptomics, enabling the acquisition of unprecedented insights and fostering research that was previously unattainable. This advanced technology allows scientists to investigate the gene expression patterns of individual cells, providing unprecedented insight into cellular differences, changes, and functions. scRNA-seq enables researchers to examine the unique gene expression patterns of each cell, revealing the true extent of cellular heterogeneity. By revealing the molecular signatures of different cell types, scRNA-seq helps researchers understand the specific roles and functions of cells within a tissue or organism. This knowledge can be used to investigate how cells interact with each other, communicate, and influence their microenvironment.

The technology has also shed light on identifying rare and previously unknown cell types. These rare cells could have crucial functional roles in development, tissue homeostasis, or disease progression, which has advanced our understanding into biological

processes. Furthermore, scRNA-seq has provided valuable information about disease mechanisms by revealing the molecular underpinnings of complex diseases by investigating differences in gene expression between healthy and diseased cells. This information can be used to identify potential therapeutic targets, develop new treatments, and better understand disease progression. In cancer research, it has deepened our understanding of tumor heterogeneity, immune cell infiltration, and the discovery of new cellular subpopulations linked to drug resistance or metastasis. Additionally, scRNA-seq has been used to study individual cell responses to drug treatments, revealing molecular-level mechanisms of drug resistance and laying the groundwork for personalized medicine.

Despite these advances, challenges remain with scRNA-seq, such as technical issues concerning sensitivity, scalability, and data analysis. However, as experimental techniques and computational methods continue to improve, scRNA-seq is expected to become even more powerful and useful in the future.

In this dissertation, we discuss scRNA-seq analysis and its application in various contexts. Chapter 1 serves as an introduction to scRNA-seq analysis, detailing current protocols, technologies, and computational methods. Chapter 2 focuses on the use of scRNA-seq in studying the immune system, examining different types of immune cells and their roles in health and disease. Chapter 3 presents our findings on abnormal immune cell subsets, functional pathway changes, and molecular signatures associated with sepsis patient outcomes. In Chapter 4, we compare single-cell transcriptomics data from sepsis, COVID-19, and SLE patients, exploring molecular pathways and potential biomarkers related to disease outcomes. We also investigate platelet-immune cell interactions and their

implications for disease severity. In Chapter 5, we examine the impact of smoking history on the tumor immune microenvironment (TIME) in lung cancer patients. Our findings reveal that smoking exacerbates T cell heterogeneity and alters gene expression patterns in immune cells, which may have implications for the efficacy of immune-based cancer treatments.

In conclusion, this dissertation discusses the computational and statistical methods for scRNA-seq data analysis and its application in studying the immune system. Our research highlights the potential of scRNA-seq in understanding immune system diversity and its implications for patient prognosis, offering valuable insights that may lead to the development of new diagnostic tools and treatments.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION TO SINGLE-CELL RNA SEQUENCING ANALYSIS 1

1.1 INTRODUCTION 1

1.2 CURRENTLY AVAILABLE SINGLE-CELL RNA-SEQUENCING PROTOCOLS AND TECHNOLOGIES 3

1.2.1 Tissue preparation and quality control 4

1.2.2 Single-cell isolation 5

1.2.3 Library preparation and quality control 7

1.3 STATISTICAL AND COMPUTATIONAL METHODS FOR SINGLE-CELL TRANSCRIPTOMICS ANALYSIS 9

1.3.1 Data preprocessing 10

1.3.2 Quality Control of scRNA-seq data 11

1.3.3 Normalization and imputation 12

1.3.4 Data integration and batch correction 14

1.3.5 Dimensionality reduction and clustering methods 16

1.3.6 Cell type prediction and annotation 18

1.3.7 Differential expression analysis and gene set analysis 20

1.3.8 Cell-cell communication networks 22

1.3.9 Inferring time dynamics from single-cell expression data 24

1.3.10 Gene regulatory networks 26

1.4 DEEP LEARNING APPLICATIONS IN SCRNA-SEQ 28

1.4.1 Dimension reduction and cluster 28
1.4.2 Cell type prediction and annotation 30
1.4.3 Treatment response prediction 32
1.5 Complementary Approaches for Studying Immune Cell Heterogeneity: Beyond Single-Cell RNA Sequencing 33
1.6 Discussion 35
CHAPTER 2 UNRAVELING IMMUNE SYSTEM COMPLEXITY THROUGH SINGLE-CELL RNA SEQUENCING 38
2.1 Introduction 38
2.2 Balancing Immune Homeostasis through Immune Cell Diversity 39
2.3 Immune Cell Profiling and Heterogeneity 41
2.3.1. Identifying immune cell types and subtypes 41
2.3.2. Characterizing immune cell states and activation 43
2.4 Inflammatory Responses in Infectious and Chronic Diseases 46
2.4.1. Dynamics of immune cell populations during infectious diseases 47
2.4.2. Characterizing inflammatory responses in chronic diseases 50
2.5 Cancer Immunology and Immuno-oncology 55
2.5.1. Tumor microenvironment and immune cell infiltration 55
2.5.2. Immune checkpoint blockade and personalized immunotherapy 57
2.6 Discussion 59
CHAPTER 3 DYNAMIC CHANGES IN HUMAN SINGLE CELL TRANSCRIPTIONAL SIGNATURES DURING FATAL SEPSIS 62

3.1 INTRODUCTION 62
3.2 MATERIALS AND METHODS 65
3.2.1 Human blood collection and harvest of PBMCs. 65
3.2.2 PBMC analysis 65
3.2.3 10X Genomics 66
3.2.4 Process and quality control of the single-cell RNA-seq data 66
3.2.5 Dimensionality reduction, clustering, and consensus-based cell-type annotation 67
3.2.6 Differential gene expression analysis and functional annotation of genes 68
3.2.7 Comparison of module scores 69
3.2.8 Statistics 69
3.3 RESULTS 70
3.3.1 Subject characteristics 70
3.3.2 Single-cell transcriptomics identify immune cell subsets associated with sepsis severity 73
3.3.3 Platelet responses are a hallmark of fatal sepsis with similar transcriptional pathways to severe COVID-19 disease 75
3.3.4 Hypoxic stress is a driving factor for erythropoiesis in sepsis with shared pathways in COVID-19 infection 79
3.3.5 Monocyte transcriptional changes occur within hours of sepsis recognition and reflect immunosuppression 81

3.3.6 CD52 is a prognostic biomarker for beneficial outcomes in sepsis and is associated with lymphocyte activation 84
3.4 DISCUSSION 88
3.5 ACKNOWLEDGMENTS 92
CHAPTER 4 PLATELETS IN IMMUNE DYSREGULATION: IDENTIFYING NOVEL PLATELET SUBTYPES IN SEVERE AND FATAL CASES OF COVID-19 AND SEPSIS 94
4.1 INTRODUCTION 94
4.2 MATERIALS AND METHODS 95
4.2.1 Integrated Single-cell Transcriptome Atlas of Peripheral Blood Mononuclear Cells (PBMCs) from COVID-19, Sepsis, and Systemic Lupus Erythematosus Patients 95
4.2.2 Integrating the Datasets and Identifying the Cell Types 96
4.2.3 Differential Expression Analysis 97
4.2.4 Comparison of Module Scores 97
4.2.5 Pathway Enrichment Analysis 98
4.2.6 Calculation of Ligand-Receptor Interaction Scores Between Platelets and Other Cell Types 98
4.2.7 Trajectory Inference of Transition in Platelets Subclusters 99
4.3 RESULTS 99
4.3.1 PBMC Composition Changes with Patient Disease Severity and Outcome. 99

4.3.2 Platelets Amplify Endotheliopathy and Lead to Potential Increase in Disseminated Intravascular Coagulation in Fatal Patients 101
4.3.3 Platelet Subpopulations Associated with Disease Severity 103
4.3.4 Pseudotime Trajectory Analysis Identifies Platelet Signature Dynamics in Survival or Fatal Disease Outcomes 108
4.3.5 Unique and Shared Gene Expression Changes in Platelets from COVID-19, SSH, Sepsis, and SLE Samples 111
4.3.6 Pathway Enrichment Related to Disease Severity in Platelets 115
4.3.7 Platelet Crosstalk with Monocytes and Lymphocytes 119
4.4 DISCUSSION 121

CHAPTER 5 IMPACT OF SMOKING ON IMMUNE CELL POPULATIONS IN LUNG CANCER PATIENTS 126

5.1 INTRODUCTION 126
5.2 METHODS 127
5.3 RESULTS 129
5.3.1 Increased T cell heterogeneity in smokers 129
5.3.2 Smoking effects on tumor-immune microenvironment at the gene expression level 131
5.4 DISCUSSION 131
5.5 ACKNOWLEDGEMENTS 133

APPENDIX A SUPPLEMENTARY FIGURES 141

APPENDIX B SUPPLEMENTARY TABLES 150

REFERENCES..... ERROR! BOOKMARK NOT DEFINED.

LIST OF FIGURES

FIGURE 1.1 OVERVIEW OF SINGLE-CELL RNA SEQUENCING WORKFLOW.....3

FIGURE 1.2: SCHEMATIC OF SINGLE-CELL RNA-SEQ ANALYSIS WORKFLOW.....9

FIGURE 2.1 OVERVIEW OF THE IMMUNE SYSTEM. 39

FIGURE 3.1 FLOW CYTOMETRIC ANALYSIS OF PBMC FROM HEALTHY CONTROL (HC), NON-SURVIVOR (NS), AND SURVIVOR (S) SEPSIS PATIENTS AT FIRST BLOOD COLLECTION (T0). 72

FIGURE 3.2 SINGLE-CELL TRANSCRIPTIONAL PROFILING OF PBMC FROM HEALTHY CONTROLS AND GRAM-NEGATIVE SEPSIS PATIENTS.....74

FIGURE 3.3 PLATELET TRANSCRIPTIONAL CHANGES OVER 6 H ARE ASSOCIATED WITH SEPSIS SEVERITY. 77

FIGURE 3.4 ELEVATED ERYTHROID PRECURSOR CELLS ARE ASSOCIATED WITH HYPOXIC STRESS. 80

FIGURE 3.5 FATAL SEPSIS PATIENTS EXHIBIT IMMUNOSUPPRESSIVE PATHWAYS IN MONOCYTES. 85

FIGURE 3.6 CD52 EXPRESSION CORRELATES WITH LYMPHOCYTE ACTIVATION. 87

FIGURE 4.1 PBMC PROFILING FROM HEALTHY CONTROLS, SEPSIS, SIMILAR SYMPTOM HOSPITALIZED, COVID-19 AND SLE PATIENTS.....100

FIGURE 4.2 DIFFERENTIAL EXPRESSION OF PLATELETS AFFECTING ENDOTHELIOPATHY ACROSS DISEASE SEVERITY STATES.....102

FIGURE A.1 UMAP REPRESENTATION OF ALL MERGED SAMPLES. 141

FIGURE A.2 ABERRANT PLATELET FUNCTION CHANGES OBSERVED IN FATAL SEPSIS. 142

FIGURE A.3 IMMUNOSUPPRESSIVE TRENDS IN MONOCYTES FROM SEPSIS NON-SURVIVORS. 143

FIGURE A.4 THE ASSOCIATION BETWEEN CD52 EXPRESSION AND B CELL ACTIVATION. ..144

FIGURE A.5 PBMC PROFILING FROM HEALTHY CONTROLS, SEPSIS, SIMILAR SYMPTOM HOSPITALIZED, COVID-19 AND SLE PATIENTS.....145

FIGURE A.6 CLUSTERED PLATELETS AND THEIR UNIQUE PATHWAY EXPRESSION CHANGES. 146

FIGURE A.7 PSEUDO-TIME PLOT OF PLATELETS FROM ALL CLUSTERS EXHIBITING TRAJECTORY FATES 147

FIGURE A.8 PLATELET'S EXPRESSION CHANGES AMONG HEALTHY CONTROLS, SEPSIS, SIMILAR SYMPTOM HOSPITALIZED, COVID-19 AND SLE PATIENTS 148

FIGURE A.9 PLATELET TRANSCRIPTIONAL CHANGES UNDER DIFFERENT DISEASE SEVERITY RELATED TO NEURODEGENERATION DISEASES. 149

LIST OF TABLES

TABLE 1.1 BRIEF OVERVIEW OF SCRNA-SEQ APPROACHES 7

TABLE 3.1 CHARACTERISTICS OF ENROLLED NON-SEPSIS VOLUNTEERS AND SEPSIS PATIENTS AT SEPSIS RECOGNITION (T0). 70

TABLE B.1 ABSOLUTE CELL NUMBERS FOR EACH CELL TYPE FROM ALL SAMPLES. 150

TABLE B.2 GENES USED FOR GENE MODULE SCORING. 151

TABLE B.3 COUNT AND FREQUENCY OF EACH CLUSTER OF PLATELETS COMPOSED OF EACH DISEASE SEVERITY. 157

TABLE B.4 COUNT AND FREQUENCY OF EACH CLUSTER OF PLATELETS COMPOSED OF EACH OUTCOME. 162

TABLE B.5 UP-REGULATED GENES IN EACH PLATELET CLUSTER COMPARED TO OTHER CLUSTERS OF PLATELETS. 165

TABLE B.6 CONSISTENTLY UP/DOWN-REGULATED GENES IN COVID-19, SIMILAR SYMPTOM HOSPITALIZED (SSH), SEPSIS, AND SLE PATIENTS COMPARED TO HEALTHY CONTROLS (HC). 168

CHAPTER 1 Introduction to Single-cell RNA Sequencing Analysis

1.1 Introduction

Transcriptomics is the study of the entire set of RNA transcripts produced by the genome of an organism, and it has become a vital tool for understanding gene expression and regulation. Over the years, several techniques have been developed for detecting and quantifying RNA transcripts. One of the first approaches that allowed scientists to detect and quantify RNA transcripts was a Northern Blot technique. It involved transferring RNA from a gel to a membrane, hybridizing it with a labeled probe, and visualizing the results using autoradiography (1). In the late 1990s, microarrays were developed, which enabled the simultaneous measurement of the expression of first hundreds and then thousands of genes. They used small pieces of DNA or RNA probes that were printed onto a glass slide and hybridized with labeled cDNA or RNA (2). RNA sequencing (RNA-Seq) is a more recent technology that uses next-generation sequencing to detect and quantify RNA transcripts (3). Since then, bulk RNA-Seq technologies have played a significant role in transcriptome profiling, enabling researchers to study transcriptional structures by simultaneously mapping transcribed regions, analyzing gene expression, and assessing dynamic range to determine the extent of gene expression. These technologies facilitate the discrimination between isoforms and alleles (4). However, bulk RNA-seq data provides average measurement of gene expression across all cells in a sample, which can obscure or mask the unique functions of individual cells (5). This limitation is particularly pronounced when cellular behavior varies greatly or is governed by rare cell types. Moreover, there may be considerable variation among individual cells, and a more detailed understanding

of cell heterogeneity is crucial for a full picture of the disease process. Investigating how metabolic, signaling, and transcriptional networks influence the interactions between individual cells and cell groups in various diseases is essential for understanding the underlying mechanisms of disease progression and developing effective therapies. Before the advent of single-cell transcriptomics, bulk transcriptomics served as the main technique for studying gene expression patterns in biological specimens. However, this method struggled to accurately reveal cellular heterogeneity, as it masked the unique transcriptomic information of individual cells. As a result, rare or transient cell states often went undetected. This limitation made it difficult to understand the intricate processes through which immune cells process information, react to infections or other stimuli, and display varied responses when confronted with different threats.

In 2009, the pioneering study on single-cell RNA-seq (scRNA-seq) was published by Tang et al., in which they developed a method for sequencing the transcriptome of individual blastomeres and oocytes (6). Since then, a growing number of modified and improved single-cell RNA sequencing technologies have come out. These have made substantial improvements in areas like sample collection, single-cell capture, barcoded reverse transcription, cDNA amplification, library preparation, sequencing, and streamlined bioinformatics analysis (7). Nature Methods named scRNA-seq "Method of the Year" in 2013, which shows how important it is in the field. These single-cell approaches have provided unprecedented molecular detail, revealing insights into the inner workings of multicellular systems and previously unattainable processes.

1.2 Currently Available Single-cell RNA-sequencing Protocols and Technologies

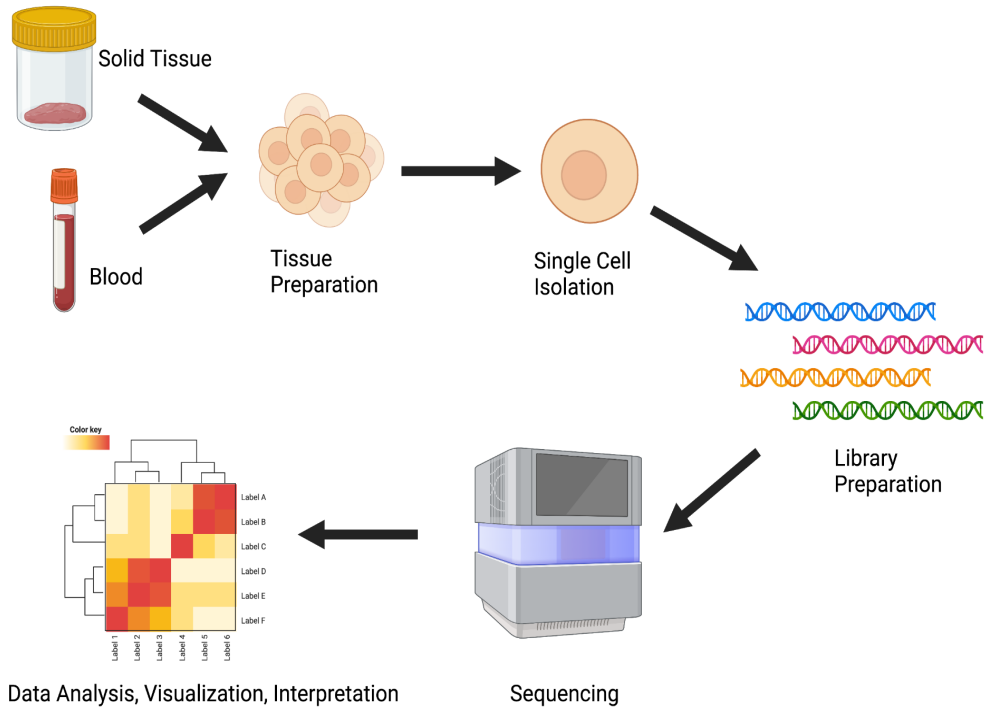

Figure 1.1 Overview of single-cell RNA sequencing workflow.

A schematic diagram illustrating the significant steps in the scRNA-seq process, such as cell isolation, library preparation, sequencing, and data analysis. Figure generated by using BioRender.

Several scRNA-seq protocols have been developed, each with their own strengths and weaknesses. The main steps in scRNA-seq are preparing the tissue, isolating and capturing a single cell, preparing the library, sequencing and primary analysis, and then visualizing and interpreting the data (Figure 1.1) (8, 9). The protocols can differ in one or more aspects (Table 1.1) (10).

1.2.1 Tissue preparation and quality control

Tissue preparation and cell isolation are the initial steps in preparation prior to library preparation. Clumped cells or cells with high death rates can confound data and lead to wrong conclusions from the experiment. Non-adherent cells, such as peripheral blood mononuclear cells (PBMCs), can be easier to handle for single-cell processing than adherent cells from organ tissues. Dissociation protocols include mechanical dissection, which involves cutting, dicing, and pipetting the tissue to break it up and separate single cells (11). Enzymatic protocol, is a method in which tissue is incubated with various enzymes, such as collagenase, trypsin, dispase, and elastase, to cleave protein bonds and extracellular matrix (12).

After the tissue is ready, density gradient centrifugation and filtration can be used to separate cells by size, shape, and density (13). The presence of dead cells can introduce confounding results to the analysis, and various fluorescence dyes are available to remove dead cells using flow cytometry (14). Flow cytometry or magnetic bead-based isolation can be used with antibody labeling for positive and negative selection to pick out certain cell subpopulations or cell types (15).

Single-cell sequencing experiments require a significant investment in time, money, sample material, and resources. Quality control measures throughout the tissue preparation process can ensure the following steps of the experiment are accomplished smoothly. Visual inspection of cell suspensions after tissue dissociation by brightfield microscopy can reveal debris, cell doublets, and larger aggregates (16). Flow cytometry can be used to

simultaneously analyze multiple metrics, including cell size, concentration, viability, cell doublets, and aggregates (17).

1.2.2 Single-cell isolation

Single-cell isolation protocols include high throughput single-cell profiling where researchers can examine up to tens of thousands of cells per experiment like droplet fluidics (18, 19), microwells (20, 21), and fluorescence activated cell sorting (FACS) (22). Low throughput single-cell profiling methods include robotic micromanipulation (23), flow sorting technologies. Low-cost methods include manual cell picking/micromanipulation (24), and serial dilution. The method that can preserve tissue spatial context is laser capture microdissection (LCM) (25) (Table 1.1).

Different scRNA-seq technologies use different isolation protocols, which each have their own pros and cons. For example, the droplet fluidics platform uses compartmentalization of individual cells in droplets using a microfluidics device followed by lysis and capture of target DNA and RNA, and it uses unique molecular identifiers (UMIs), and cell barcodes to enable cell and gene-specific identification, low-cost per cell. UMIs are short, random nucleotide sequences that are incorporated into the reverse transcription primers or adapters used during the library preparation process (18, 19). In microwell-based methods, cells are isolated using microscopic wells etched into a solid surface, such as a microarray or a microplate. Each well is designed to capture a single cell, and the wells are filled with the necessary reagents for reverse transcription and cDNA synthesis. It supports imaging and short-term cell culture, and it is ideal for adherent cells.

Microwell-based methods can also utilize UMIs to enable cell and gene-specific identification, similar to droplet-based methods (20, 21). However, both droplet fluidics and microwells require specialized equipment. Another method is fluorescence-activated cell sorting (FACS), which applies microdroplets with single cells isolated using electric charge. It uses specific immuno-tagging of cell-surface markers to improve accuracy and has high throughput, but compared to other protocols, it requires specific antibodies/markers. Robotic micromanipulation uses robotic micropipettes to isolate single cells. It has high accuracy but low throughput.

Flow sorting uses electric charge to isolate microdroplets containing single cells, so cell types are selected accurately by size, morphology, internal complexity, and protein expression by antibody labeling. Still, flow sorting depends on expensive, specialized equipment, and cells tend to be exposed to high pressure. Micromanipulation can separate different types of cells from mixed populations, but it is slow and needs a lot of cells to start with. LCM was applied as cells were cut from tissue sections with a laser under a microscope. It can preserve spatial context but is technically challenging and can potentially cause UV damage to RNA/DNA.

Table 1.1 Brief overview of scRNA-seq approaches

Platform	Isolation Strategy	Throughput (Cell numbers)	Full-Length Coverage	UMIs	Reference
Fluidigm C1 (SMART-seq)	Microfluidics	10s - 100s	Yes	No	(26)
CEL-seq	Plate-based/microfluidics	100s - 1000s	No	Yes	(27)
Quartz-Seq	Plate-based	100s - 1000s	Yes	No	(28)
SMART-seq2	Plate-based	100s - 1000s	Yes	No	(29)
MARS-seq	Plate-based/microfluidics	100s - 1000s	No	Yes	(30)
10x Genomics Chromium	Droplet	1000s - 10,000s	No	Yes	(31)
inDrop-seq	Droplet	1000s - 10,000s	No	Yes	(19)
MATQ-seq	Plate-based	100s - 1000s	Yes	No	(32)
ddSEQ	Droplet	1000s - 10,000s	No	Yes	(18)
BD Rhapsody	Microwells	1000s - 10,000s	No	Yes	(33)
Seq-Well	Nanowell array	1000s	No	Yes	(34)
SPLIT-seq	Plate-based	1000s - 10,000s	No	Yes	(35)

1.2.3 Library preparation and quality control

Library preparation is the next critical step in the single-cell sequencing workflow. There are now more amplification techniques for single-cell profiling. Amplification techniques

are crucial for converting minute amounts of RNA into detectable signals, with each technology employing unique methods. Techniques range from PCR-based amplification (Fluidigm C1, SMART-seq2, ddSEQ, BD Rhapsody, Seq-Well, SPLIT-seq) to the use of the T7 RNA polymerase to perform linear isothermal amplification of cDNA by in vitro transcription (IVT) (CEL-seq, MARS-seq, Quartz-seq, inDrop-seq) and combinations of droplet-based cell isolation, barcoding, and PCR-based amplification (10x Genomics Chromium). Another method MATQ-seq, uses a multiple annealing and looping based amplification cycles (MALBAC), including quasilinear amplification and PCR (36).

Full-length transcript scRNA-seq technologies, such as Fluidigm C1 (SMART-seq), Quartz-Seq, and SMART-seq2, are advantageous for capturing complete gene expression profiles and enabling the analysis of alternative splicing events. In contrast, UMI-based technologies like CEL-seq, MARS-seq, 10x Genomics Chromium, inDrop-seq, ddSEQ, BD Rhapsody, Seq-Well, and SPLIT-seq provide improved quantification accuracy by mitigating amplification biases and enabling the detection of PCR duplicates (Table 1.1).

qPCR is typically used to quantitatively measure the quality and quantity of the prepared libraries in order to maximize the quality and output of the sequencing data. After the size and quality of the libraries have been measured, the right amount can be loaded for sequencing. The exact amount depends on the sequencing platform, library size, and flow cell (37).

1.3 Statistical and Computational Methods for Single-cell transcriptomics Analysis

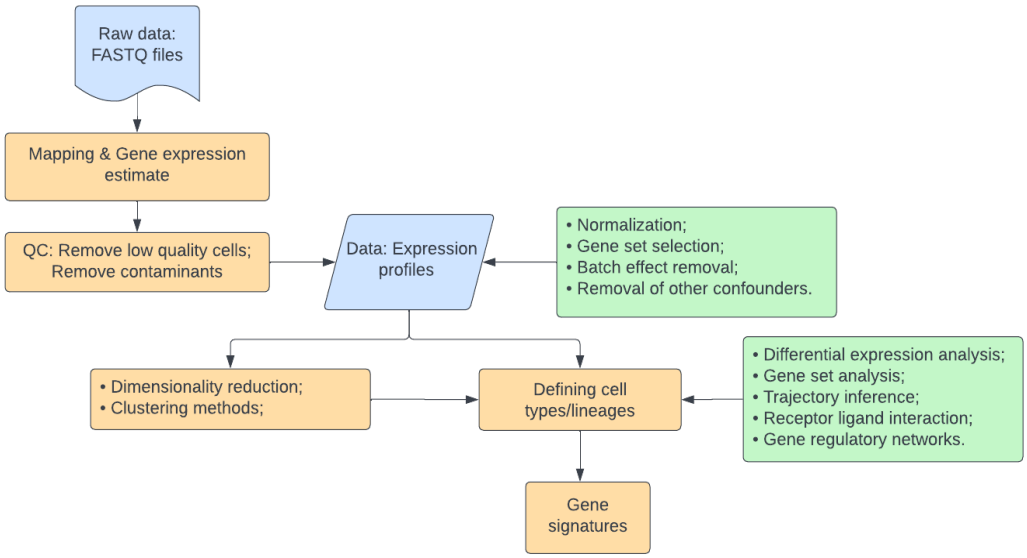

Figure 1.2: Schematic of single-cell RNA-seq analysis workflow

The initial step in the workflow involves processing and aligning raw sequencing data to generate count matrices. These count data are then subjected to pre-processing and subsequent downstream analysis.

Upon completion of the sequencing, the data undergoes a multi-step analysis process. In the case of single-cell sequencing, this process consists of primary, secondary, and tertiary analysis stages. The primary analysis involves data preprocessing steps for file conversion. The secondary analysis encompasses demultiplexing, alignment, and quality control (QC) measures. The tertiary analysis is concerned with data processing, including visualization and interpretation of the data (Figure 1.2).

1.3.1 Data preprocessing

Raw sequencing data are typically generated in the form of FASTQ files, which contain the sequence reads and corresponding quality scores (38). Before aligning the reads to a reference genome or transcriptome, quality control (QC) steps are essential to assess the overall quality of the sequencing data and remove potential sources of error (39). Common QC procedures include filtering low-quality reads, trimming adapter sequences, and removing PCR duplicates (40).

Following QC, the processed reads are aligned to a reference genome or transcriptome using specialized aligners. There are various aligners available, such as STAR (41), HISAT2 (42), and Kallisto (43), which use different algorithms and strategies to optimize alignment accuracy and speed. The choice of aligner depends on the specific scRNA-seq technology used, as some aligners are tailored to handle unique molecular identifiers (UMIs) or specific library preparation methods (44).

Once the reads are aligned, gene expression quantification is performed to generate count matrices. The count matrix represents the number of reads mapped to each gene for each individual cell. Tools such as featureCounts (45) or HTSeq (46) are commonly used to quantify gene expression from aligned reads. For UMI-based scRNA-seq data, tools like UMI-tools (47) or Drop-seq tools (18) can be employed to generate count matrices while accounting for UMIs to reduce PCR duplicates and amplification biases.

The resulting count matrices are the foundation for downstream analyses such as normalization, batch effect correction, dimensionality reduction, and differential

expression analysis, which enable the characterization of cell populations and the identification of cellular subtypes and gene regulatory networks.

1.3.2 Quality Control of scRNA-seq data

The main purpose of quality control (QC) in scRNA-seq analysis is to remove low-quality cells and doublets (48). To identify low-quality cells, we should examine the expected library size and the number of expressed genes. All cell types have an expected library size and a typical number of expressed genes. Cells that fall outside the expected range may be either poor-quality cells or technical artifacts such as dead or dying cells or multiple-cell aggregates that should be excluded from downstream analysis (49). To detect doublets, the number of genes per cell reflects the situation. For any given cell type, there is a typical expected number of expressed genes (n). While most viable single cells may fall into a natural distribution around this number, cells observed outside that distribution, e.g., with roughly twice that number (2n), may represent cells of interest warranting further investigation and characterization. Another QC metric is the proportion of reads that mapped to genes in the mitochondrial genome or reads that mapped to ribosomal RNAs (50). High proportions of mitochondria and ribosomes are a sign of low-quality cells, most likely due to more apoptosis. These cells may be excluded from downstream analysis. A third metric involves plotting genic unique molecular identifier (UMI) counts in descending order against cell barcodes, enabling the statistical identification of real cells and excluding noncellular barcodes (18). Cell barcodes above the threshold have genic

UMIs that represent real cells, while those below the threshold, typically representing empty beads, have genic UMI counts of 1-100.

In addition to removing cells with poor quality and doublets, it is important to examine genes that might be artifacts. We further examine the UMI and gene distribution in the datasets and check the UMI distribution of individual genes in each dataset. In most cases, the proportion is < 0.1 (i.e., the UMIs of any gene are less than 10% of the total UMIs). If any gene has extremely high UMIs, we will consider removing it from the gene-cell matrix as it is most likely artificial (51).

1.3.3 Normalization and imputation

In scRNA-seq analysis, it is important to take into account different types of differences, such as biological and technical factors, in order to correctly identify the heterogeneity of cell populations, find rare cell types, and find out how cells decide what to do with their lives (52). Biological differences can be caused by differences in cell types, cell states, cell sizes, or samples like age, gender, disease, or treatment. Technical factors can also introduce bias into the raw read counts, such as gene length, GC content, sequencing depth, and dropouts. Various normalization tools are commonly used in scRNA-seq data analysis to overcome these challenges. Transcripts per million (TPM) is a normalization method that accounts for differences in sequencing depth and library size between cells (53). TPM normalizes by dividing the count for each transcript by the total number of transcripts in that cell, then multiplying by the 10^6 scaling factor.

$$TPM = \frac{Number\ of\ reads\ mapped\ to\ the\ transcript / Length\ of\ the\ transcript\ in\ kilobases}{\sum(Number\ of\ reads\ mapped\ to\ the\ transcript / Length\ of\ the\ transcript\ in\ kilobases)} * 10^6$$

Fragments per kilobase of transcript per million (FPKM) is another normalization method that accounts for sequencing depth and gene length (53). FPKM normalizes by calculating the number of fragments mapping to each gene, dividing by the gene or transcript length, and multiplying by the 10^6 scaling factor.

$$FPKM = \frac{Number\ of\ raw\ reads\ or\ fragment\ counts}{Length\ of\ the\ transcript\ in\ kilobases\ * \sum(Number\ of\ raw\ reads\ or\ fragment\ counts) * 10^6}$$

DESeq2 is a popular normalization tool that can also be applied to scRNA-seq data, the normalization method is based on the negative binomial distribution. It estimates size factors from the median-of-ratios method to normalize the counts (54). SCnorm is an R package and a normalization method that calculates scaling factors for each cell by fitting a quantile regression model (55). Seurat is a popular scRNA-seq analysis tool that includes normalization methods such as "LogNormalize" and "SCTransform" (56). LogNormalize is a global-scaling normalization method. In order to use this method, first divide the raw counts for each cell by the total counts for that cell, and then multiply the result by a scaling factor (the default is 105). Finally, the natural logarithm of the resulting value is computed.

$$LogNormalized = ln\left(\frac{The\ read\ count\ of\ gene_x\ in\ the\ cell}{Total\ counts\ per\ cell} \times Scaling\ factor\ + 1\right)$$

While SCTransform uses regularized negative binomial regression to learn gene-group specific factors (57).

The drop-out effect is a common technical artifact in scRNA-seq. It happens when there isn't enough RNA in some cells to detect genes. This can lead to missing data and

inaccurate estimates of how much a gene is being expressed. It can be a big problem for scRNA-seq analysis. The drop-out effect happens because scRNA-seq usually only sequences a small amount of the RNA molecules in a cell, and the number of molecules from a gene that are sequenced determines whether or not that gene is found (49). To mitigate the effects of dropout, various computational methods have been developed, including imputation methods such as MAGIC (Markov Affinity-based Graph Imputation of Cells), which utilizes data diffusion to share information across similar cells and successfully denoise the cell count matrix, fill in missing transcripts, recover gene-gene relationships, and additional structures (58). SAVER (Single-cell Analysis Via Expression Recovery), applies a regularized regression prediction and empirical Bayes technique to extract the actual gene expression pattern from scRNA-seq data that is both sparse and noisy (59). And scImpute, a regularized regression prediction and empirical Bayes technique to extract the actual gene expression pattern from scRNA-seq data that is both sparse and noisy (60).

1.3.4 Data integration and batch correction

Analyzing scRNA-seq data can be challenging due to technical variability and differences between experimental conditions. Technical variability can be caused by several factors, including the use of different sequencing platforms, library preparation methods, and sequencing depth, which can introduce biases and affect the accuracy of gene expression quantification (61). Moreover, different experimental conditions, such as different cell types, cell states, and treatments, can result in significant differences in gene

expression patterns, making it difficult to compare data across different conditions (56). To address these challenges, integration and batch correction methods are used to combine data from multiple experiments or sources. This process helps to improve the resolution and sensitivity of scRNA-seq analyses by identifying shared variation across different datasets or samples and removing or reducing technical variability between experiments or samples (62).

The principles that underlie many of the integration and batch correction methods involve identifying cell types, genes, or other features that are expressed similarly across different experiments or sources and removing the effects of technical variability (63). The ultimate goal of integration and batch correction methods is to improve the accuracy and sensitivity of downstream analyses, such as clustering, differential gene expression analysis, and pathway analysis.

There are several scRNA-seq integration methods available, each with its own strengths and limitations. One popular method is Seurat Integration, which uses a canonical correlation analysis (CCA) approach to identify shared variation across different datasets and integrate them into a single dataset (56). Another integration method is Seurat Anchors, which uses a mutual nearest neighbor approach to identify shared cell types between a reference dataset and other datasets, and then integrates the data using CCA (64). Harmony is another integration method that employs an unsupervised approach, using a novel iterative algorithm that aligns single cell data in a shared low-dimensional space, thus preserving the inherent structure and relationships within the data (65). Scanorama aligns datasets based on the intersection of all genes, a conservative approach that minimizes

differences due to expression quantification methods (66). Lastly, MNN (mutual nearest neighbors) is a batch correction method that identifies and aligns similar cells from different batches by matching their nearest neighbors and adjusts the expression values to remove batch effects (67).

1.3.5 Dimensionality reduction and clustering methods

Dimensionality reduction and clustering are two important steps in the analysis of scRNA-seq data (68). scRNA-seq produces high-dimensional data with thousands of genes and hundreds to thousands of cells, making it difficult to visualize and interpret. Dimensionality reduction methods are used to reduce the complexity of the data by projecting it into a lower-dimensional space while retaining as much information as possible. Clustering methods are used to group similar cells into distinct clusters based on their gene expression profiles. These two steps help to identify cell types and subpopulations, and reveal the heterogeneity and complexity of the biological system being studied.

Principal Component Analysis (PCA) is one of the most widely used dimensionality reduction methods (69). PCA works by finding the most important directions of variation in the data and projecting the data onto a lower-dimensional space while retaining the maximum amount of variance. This method is simple and computationally efficient, and it is often used as a preprocessing step for downstream analyses. t-SNE (t-Distributed Stochastic Neighbor Embedding) is a nonlinear dimensionality reduction method that is often used for visualizing scRNA-seq data (70).

t-SNE (t-Distributed Stochastic Neighbor Embedding) (71) and UMAP (Uniform Manifold Approximation and Projection) (72) are both non-linear, graph-based methods for dimensionality reduction in scRNA-seq analysis. However, t-SNE seeks to preserve local structures in the data by minimizing the divergence between two probability distributions: one that measures pairwise similarities in the high-dimensional space and another in the low-dimensional space. The algorithm uses a Student's t-distribution in the low-dimensional space to model pairwise similarities. As for UMAP, UMAP aims to preserve both local and global structures in the data by approximating the manifold on which the data lie. It does this by constructing a graph representation of the high-dimensional data, followed by optimizing an embedding in the low-dimensional space that preserves the topological structure of the graph. UMAP is generally faster than t-SNE and is better at preserving global structures, making it useful for a wider range of applications.

Clustering methods are used to group similar cells into distinct clusters based on their gene expression profiles. K-means is a popular clustering method that aims to partition the data into K distinct clusters, where K is a user-defined parameter. K-means works by iteratively assigning each cell to the closest centroid and then updating the centroids based on the mean of the assigned cells (73). Hierarchical clustering is another clustering method that builds a tree-like structure of the cells based on their pairwise distances. It can be used to identify clusters at different levels of granularity and is often visualized as a dendrogram (74). Density-based clustering methods, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), work by

identifying dense regions in the data and clustering cells that belong to these regions. These methods can be more robust to noise and able to identify clusters of different shapes and sizes. Graph-based clustering methods, such as Louvain clustering and Leiden clustering, use a graph representation of the data to identify communities or clusters of cells. These methods work by identifying densely connected regions in the graph and grouping cells that belong to these regions (56). Graph-based clustering methods can be more efficient and flexible than other clustering methods and can identify rare or transient cell types.

In summary, dimensionality reduction and clustering are crucial steps in scRNA-seq data analysis, allowing researchers to visualize, interpret, and identify cell types and subpopulations within complex biological systems. Methods such as PCA, t-SNE, and UMAP are commonly used for dimensionality reduction, each with its own advantages and limitations. Clustering methods, including K-means, hierarchical clustering, density-based clustering (DBSCAN and HDBSCAN), and graph-based clustering (Louvain and Leiden), help group cells based on their gene expression profiles, revealing the structure and relationships between different cell types and subpopulations.

1.3.6 Cell type prediction and annotation

Cell type prediction and annotation tools are very important for scRNA-seq data analysis because they let researchers find and describe cell types based on how their genes are expressed. This information can be used to understand the functional roles of different cell types, compare gene expression patterns across cell types, and link cell types to specific biological functions. Methods and algorithms for cell type prediction and annotation use

machine learning and statistical methods to figure out what kind of cell each cell is based on the way its genes are expressed.

SingleR is a method for cell type prediction, utilizes a reference dataset to annotate cells in a target dataset. It first applies dimensionality reduction and clustering to both datasets and then matches the cell types in the target dataset to those in the reference dataset by correlating the gene expression profiles. This approach enables researchers to accurately identify cell types and uncover the underlying biological processes associated with them (75). scCATCH, is another tool for cell type annotation specifically designed for scRNA-seq data. scCATCH employs an automatic cell type annotation algorithm that uses a priori gene sets to identify cell type-specific marker genes. It then classifies cells into known cell types based on these marker genes, allowing for precise cell type predictions (76). Seurat v4 also allows for reference mapping to annotate cell types, which involves mapping query datasets to annotated references in order to interpret scRNA-seq data. Seurat v4 also has new features that let you project query cells onto a UMAP visualization that has already been made (77). CellAssign is a method for cell type prediction that automates the annotation process by computing a probabilistic assignment for each cell to a cell type defined by a set of marker genes or to an "unassigned" class. The marker genes can be found with the help of expert knowledge from literature or databases like CellMarker, or they can be found directly in data from sources like PanglaoDB (78). Another tool for cell type prediction is LIGER, which is a method that takes in multiple single-cell datasets and identifies shared and dataset-specific latent factors, or metagenes, that correspond to biological or technical signals. Through integrative nonnegative matrix factorization, these

metagenes are calculated. They can be used to find groups of cells and label them with known marker genes or statistical models (79). Lastly, CellTypist, which is an automated tool for annotating cell types in scRNA-seq datasets using logistic regression classifiers trained on either built-in or user-generated models, The tool uses majority voting to improve the prediction results and gives each subcluster a distinct label based on the type of cell that is predicted to be most common (80).

1.3.7 Differential expression analysis and gene set analysis

Differential expression analysis (DEA) is a commonly used method in scRNA-seq data analysis to identify genes that are specifically expressed in certain cell types or are differentially expressed, potentially contributing to functional differences between cell types. DEA helps researchers identify potential biomarkers for disease diagnosis or drug response, understand gene regulation, and identify genes involved in specific biological processes. Furthermore, DEA validates the accuracy of cell type assignments obtained from unsupervised clustering methods (81).

EdgeR, DESeq2, and MAST are popular methods and tools for performing DEA in scRNA-seq data. EdgeR and DESeq2 use negative binomial distributions to model differences in gene expression and find genes that are significantly different in how they are expressed in different conditions. In contrast, MAST addresses the sparsity and heterogeneity of scRNA-seq data by using a zero-inflated negative binomial model to identify differentially expressed genes. These important tools and methods enable

researchers to better understand the mechanisms controlling gene expression and functional differences among distinct cell types (82).

Upon obtaining DEA results, gene set analysis (GSA) is typically performed. GSA helps researchers identify gene sets that are enriched or depleted in specific cell types or conditions and are involved in particular biological functions or pathways. By focusing on biologically important gene sets, GSA makes scRNA-seq data easier to understand and confirms the biological relevance of clustering results, making sure they are not just based on technical variation (83). Various GSA methods are available, such as Gene Set Enrichment Analysis (GSEA), which determines whether gene sets tend to occur towards the top or bottom of a list, indicating a correlation with the phenotypic class distinction. GSEA has been used successfully to analyze molecular profiling data, and it has been changed and made more general so that it can be used in more situations. A software package and an initial inventory of gene sets called MSigDB are available for use (84). Another method is GSVA (Gene Set Variation Analysis), which is a GSE method that estimates pathway activity variation over a sample population in an unsupervised manner. GSVA demonstrates robustness in comparison to current state-of-the-art sample-wise enrichment methods and provides increased power to detect subtle pathway activity changes (85). g:GOSt is a web server that performs functional enrichment analysis, also known as over-representation analysis, on an input gene list. It connects genes to sources of functional information and finds statistically significant enriched terms. It works with almost 500 organisms and hundreds of types of identifiers (86). scGSEA and scMAP, which are tailored to scRNA-seq data and can be used to automate cell functional

annotation, scGSEA is a statistical framework that finds the active pathways in a cell. scMAP is a transfer learning algorithm that maps a query set of cell transcriptional profiles onto an existing reference atlas. Both methods are based on non-negative matrix factorization and were validated on simulated and real datasets (87). Lastly, the ClusterProfiler library is a tool for performing over-representation analysis and gene set enrichment analysis using GO and KEGG for several model organisms, supporting thousands of species with up-to-date gene annotation and user annotation data for novel species. The library has some features that make it stand out, such as a clean interface, the ability to see enrichment results with ggplot2, and packages like ChIPseeker, GOSemSim, and enrichplot that go well with it. This makes it one of the most popular Bioconductor packages, used in many CRAN and Bioconductor packages, pipelines, and online platforms (88).

1.3.8 Cell-cell communication networks

ScRNA-seq data allows for the analysis of cell-cell communication networks by examining the expression patterns of genes involved in intercellular signaling (89). These networks are crucial in understanding mechanisms that regulate biological processes, such as development, immune response, and cancer progression. By figuring out how different types of cells interact with each other through ligands and receptors, researchers can learn about the signaling pathways that control cell behavior and find possible therapeutic targets. Cell-cell communication networks can also be used to find new types of cells based on how

they interact with other cells and to find potential biomarkers for diagnosing diseases or measuring how well drugs work.

Various tools have been developed for analyzing cell-cell communication networks in scRNA-seq data. CellChat, an R package, uses a manually curated signaling molecule interaction database to infer, visualize, and analyze intercellular communications from scRNA-seq data, along with several visualization outputs to facilitate user-guided data interpretation, pattern recognition methods, and manifold learning approaches for characterizing and comparing the inferred intercellular communications within complex tissues (90). CellPhoneDB is a public database of ligands, receptors, and how they interact with each other. It uses a computational method to find biologically important interactions between ligands and receptors from scRNA-seq data. This predicts molecular interactions between cell populations through specific protein complexes and creates possible networks for cell-cell communication (91). iTALK is an R package for profiling and visualizing ligand-receptor-It uses a large database of known ligand-receptor interactions (LRdb) and a new regularized product score to adjust to different levels of depth in single-cell datasets.mediated intercellular cross-talk signals from scRNA-seq data, which allows for the capture of abundant ligand-receptor gene pairs, identification of gains/losses in cellular interactions, and tracking of dynamic changes in intercellular communication signals with an efficient data visualization tool and a built-in iTALK ligand-receptor database that is periodically updated and curated (92). Then another available R package for ligand-receptor interactions is SingleCellSignalR, it uses a comprehensive database of known ligand-receptor interactions (LRdb) and a new regularized product score to adapt to

variable levels of depth in single-cell datasets, and provides a range of visualization and complementary analysis tools to infer ligand-receptor interactions between cell populations and put them in context (93). Lastly, NicheNet, it is a tool that integrates input gene expression data with a prior model of ligand-to-target signaling paths, going beyond ligand-receptor interactions to predict which ligands influence the expression in another cell, which target genes are affected by each ligand and which signaling mediators may be involved, using a regulatory potential score calculated between all pairs of ligands and target genes based on network propagation methods (94).

1.3.9 Inferring time dynamics from single-cell expression data

Reconstructing the temporal progression of cellular processes at the single-cell level through inferring time dynamics from single-cell expression data is a valuable tool for identifying subpopulations of cells that share similar temporal patterns and trajectories, revealing the heterogeneity present in single-cell expression data and facilitating the identification of underlying patterns and dynamics of gene expression (95). This approach can shed light on differentiation, cell cycle, response to environmental stimuli, and the cell states and transitions that drive heterogeneity (96). Modeling the temporal progression of cellular processes through inferring time dynamics not only reveals the heterogeneity present in single-cell expression data, but also enables researchers to predict the future fate and response of cells based on their current gene expression patterns. This approach can be used to predict differentiation trajectories of stem cells, cell cycle phases of proliferating cells, and the response of cells to external stimuli or treatments (97). Finally, inferring time

dynamics can enable comparisons across conditions. By identifying the temporal patterns of gene expression across different conditions, such as healthy vs. diseased cells, or drug-treated vs. untreated cells, researchers can gain insight into the pathogenesis of diseases and the mechanisms of drug action (56). This can have important implications for developing personalized medicine approaches and identifying new therapeutic targets. However, one of the challenges from analyzing single cell transcriptomics data is extracting temporal information from a single time snapshot. Presently, researchers employ time-course experiments to gather samples at various time points throughout a biological process, enabling them to track the temporal alterations in gene expression (98-100). Another approach is to use computational methods to infer the temporal dynamics of gene expression from static snapshots of single cell transcriptomics data. These methods use mathematical models to predict how gene expression patterns change over time based on the static data, enabling researchers to identify key regulatory pathways and predict the effects of perturbations on gene expression.

There are several popular tools available to researchers. One widely used tool is Monocle, which harnesses single-cell variation to sort cells in "pseudo-time". The idea behind pseudotime is to order cells along a trajectory that represents a biological process, such as differentiation or response to stimuli, based on their gene expression profiles. In Monocle 2, pseudotime reconstruction was based on a method called reversed graph embedding (RGE), which utilizes a minimum spanning tree (MST) algorithm. Monocle 3, uses an improved method called principal graph learning for trajectory inference (95, 101). Another popular algorithm is SCUBA (Single-Cell Universal Bayesian Analysis), which

infers the temporal progression of gene expression in single cells using a nonlinear dynamics and stochastic differential equation theories (102). Wanderlust is another method, it takes single-cell measurements and a user-defined early cell as input and calculates a trajectory by iteratively refining the position of each cell along the trajectory using a weighted average based on the shortest-path distance from a set of randomly chosen waypoints, repeating the refinement step until the positions of all cells converge, and outputting the average trajectory over all graph trajectories (96). PAGA (Partition-based Graph Abstraction) is a computational method for combining clustering and trajectory inference in single-cell RNA sequencing data. It involves constructing a graph that represents the relationships between individual cells based on their gene expression profiles (103). scVelo is a likelihood-based dynamical model that estimates RNA velocity to infer dynamic information from RNA sequencing data. It solves the full gene-wise transcriptional dynamics, inferring gene-specific reaction rates of transcription, splicing, and degradation. scVelo generalizes RNA velocity estimation to transient systems and systems with heterogeneous subpopulation kinetics (104).

1.3.10 Gene regulatory networks

Gene regulatory networks (GRNs) play a crucial role in comprehending cellular processes such as development, differentiation, and response to stimuli (105). GRNs control the development and physiological functions of animals. Comparative analyses of GRNs across species can reveal evolutionary patterns and help us understand how changes in regulatory networks can lead to novel phenotypic traits (106). The use of GRN tools is

crucial for scRNA-seq analysis, as they enable identification of important regulators and targets, characterization of gene modules, prediction of gene function, and validation of computational models, thereby aiding in the recognition of cellular pathways, the development of therapeutic targets, and improved comprehension of diseases.

Numerous well-known algorithms and tools have been developed for inferring GRNs from scRNA-seq data. Examples of such tools include SCENIC (107), GENIE3 (108), Inferelator (109), GRNBoost2 (110), PIDC (111), and SCODE (112). SCENIC was developed to link cis-regulatory sequences to single-cell gene expression, allowing for the optimization of cell state discovery and characterization. The method consists of three steps: identification of coexpressed genes with transcription factors, cis-regulatory motif analysis, and pruning of modules to remove indirect targets, resulting in processed modules called regulons. GENIE3 is a powerful and flexible GRN inference method. It uses ensembles of regression trees to capture high-order dependencies and non-linear relationships between variables while making a directed graph of regulatory interactions. This makes it a useful tool for figuring out GRNs in different biological systems (108). Inferelator is another tool for single-cell gene regulatory network (GRN) inference based on regularized regression, that provides scalability, integrated support for multi-task GRN inference, and the ability to learn cell-type-specific GRNs (109). GRNBoost2 leverages gradient boosting to pinpoint the most relevant TFs (110), while PIDC uses a probabilistic graphical model to detect relevant TFs (111), and SCODE relies on differential expression analysis to identify TFs that regulate gene expression (112).

1.4 Deep Learning Applications in scRNA-seq

1.4.1 Dimension reduction and cluster

As deep learning (DL) algorithms continue to improve, more and more applications are using them for dimensionality reduction and clustering to manage the large and complicated scRNA-seq datasets and get valuable insights from them.

Analysis of scRNA-seq data has been done successfully using techniques like deep autoencoders and variational autoencoders (VAEs) to reduce the number of dimensions. Deep autoencoders are artificial neural networks that develop efficient, low-dimensional representations of high-dimensional data unsupervised. An encoder maps input data to a lower-dimensional latent space, and a decoder attempts to reconstruct the original data from the lower-dimensional representation (113). Adding a probabilistic layer to the latent space creates a variational autoencoder (VAE) generative deep learning model. It uses deep learning and variational inference to learn complex data distributions unsupervised. In VAE, the encoder network maps input data to a probability distribution over the latent space, and the decoder network reconstructs the data using samples from this distribution. The VAE is trained by maximizing reconstruction loss and a regularization term based on the Kullback-Leibler (KL) divergence to make the learnt latent distribution match a predetermined prior distribution (114).

There have been several studies on using deep autoencoders for dimensionality reduction of scRNA-seq data. One such study proposes a deep count autoencoder (DCA) for denoising scRNA-seq data, The DCA model architecture consists of an encoder and a

decoder with multiple layers, with the goal of learning a low-dimensional representation of the input data. DCA is specifically designed to handle count data, which is common in scRNA-seq datasets. DCA can impute missing values (dropouts) and reduce technical noise, improving the quality of the scRNA-seq data (115). Another such study proposes a scRNA-seq data dimensionality reduction algorithm based on a hierarchical autoencoder, termed SCDRHA. The proposed SCDRHA consists of two core modules: a DCA that is used to denoise data and a graph autoencoder that projects the data into a low-dimensional space (116). scvi-tools (single-cell variational inference tools), a VAE-based method, is a proposed method that combines VAEs with factor analysis to model the underlying structure of scRNA-seq data, The model is trained using a combination of reconstruction loss (negative log-likelihood) and a regularization term based on the Kullback-Leibler (KL) divergence, which encourages the learned latent distribution to be close to the structured prior (117).

The application of DL methods has extended to clustering analysis as well. Clustering techniques, such as deep embedded clustering (DEC) and single-cell deep clustering (SCDC), have been utilized to group cells with similar gene expression patterns, uncovering distinct cell types and states (118, 119). For example, SCDC's proposed method combines deep learning techniques with model-based clustering to handle the unique challenges of scRNA-seq data, such as high dimensionality, noise, and dropout events. The model architecture consists of an encoder network that learns a low-dimensional representation of the input data and a clustering model that groups cells in the latent space (119).

Another notable DL-based clustering approach is graph neural networks (GNNs), which have been applied to single-cell data analysis. GNNs are a class of deep learning models specifically designed to handle data represented as graphs. The GNN model is built on recursive neural networks and operates by iteratively passing messages between neighboring nodes in the graph. The GNN architecture includes a state transition function, which updates a node's hidden state based on information from its neighbors, and an output function that generates the final output for each node (120). For instance, scGNN (single-cell graph neural network) is a method that employs a GNN framework specifically designed for scRNA-seq analyses. It leverages the power of GNNs to model the complex relationships between cells and genes in scRNA-seq data by representing them as a graph. The scGNN architecture consists of an encoder that learns a low-dimensional representation of the input data and a decoder that reconstructs the original data (121).

1.4.2 Cell type prediction and annotation

Deep learning techniques, especially neural networks, have demonstrated exceptional performance in scRNA-seq cell type annotation tasks. For example, scANVI (single-cell ANnotation using Variational Inference), a deep generative model developed for probabilistic representation of scRNA-seq data. The inference procedure for scANVI relies on neural networks, stochastic optimization, and variational inference, allowing scalability to large numbers of cells and datasets. scANVI also achieves transferring labels from one dataset to another (122). scDeepSort is another tool, which is an annotation tool for scRNA-seq data that uses a deep learning model with a weighted GNN and pre-trained cell-type

annotation on human cell landscape (HCL) and a mouse cell atlas (123). Another DL-framework is scDLC (scRNA- seq deep learning classifier). scDLC is a deep learning-based framework specifically designed for the classification of large-scale scRNA-seq data. The framework employs a combination of autoencoders and convolutional neural networks (CNNs) to learn meaningful features and accurately classify cells into different cell types (124).

Large-scale pretrained deep language model also has been applied in cell type prediction and annotation. For example, scBERT (single-cell Bidirectional Encoder Representations from Transformers), which is based on the popular BERT (Bidirectional Encoder Representations from Transformers) model, originally developed for natural language processing tasks. The authors adapt BERT to the scRNA-seq domain by pretraining the model on a massive dataset of gene expression profiles, enabling it to learn biologically relevant features. After pretraining, scBERT is fine-tuned on smaller, labeled scRNA-seq datasets for cell type annotation tasks, making it a transfer learning approach (125).

Transfer learning, a deep learning technique where a model trained for one task is repurposed as the initial stage for a model on a different task, it offers significant benefits in terms of computational and time resources needed to develop neural network models for these complex problems while providing substantial improvements in performance on related tasks (126), has also been applied to cell type annotation in scRNA-seq data analysis. For example, scArches (single-cell architectural surgery) is a transfer learning-based method that uses architectural surgery, which involves reusing neural network models by

adding input nodes and weights (adaptors) for new studies and fine-tuning only those parameters. The approach is demonstrated in applications such as transferring cell type annotations from a reference to a query atlas and mapping COVID-19 query data onto a healthy reference (127).

1.4.3 Treatment response prediction

Deep Learning (DL) has emerged as a powerful tool in scRNA-seq data analysis, enabling the prediction of molecular progress by extracting complex patterns from large-scale data. One such model, deep learning framework, scDEAL (single-cell Drug rEsponse AnaLysis), uses Domain-adaptive Neural Network (DaNN) adapted to predict drug responses using both bulk and scRNA-seq data. scDEAL effectively predicts single-cell drug sensitivity by connecting drug sensitivity, gene features in single cells, and gene features in bulk samples (128). Another DL-framework, VEGA (VAE Enhanced by Gene Annotations), a novel VAE architecture for analyzing the activity of various biological modules in different contexts using single-cell transcriptomics data. The method can simultaneously investigate cell type and cell state in both control and experimentally perturbed conditions. It also can help understand the response of specific cell types to different perturbations and provide interpretable insights on biological module activity (129). Lastly, Precily, a deep neural network (DNN) framework designed to predict cancer therapy responses based on single cell gene expression profiles and drug descriptors. The framework infers cellular fates upon treatment from single-cell expression data, enabling drug response prediction at sub-clonal resolution using tumor scRNA-seq data (130).

In conclusion, recent advances in deep learning have enabled major progress in the ability of artificial intelligence techniques to integrate big data, incorporate existing knowledge, and learn arbitrarily complex relationships. In the context of scRNA-seq, these advances have significantly impacted our understanding of cellular heterogeneity, cell type identification, and treatment response prediction. As deep learning models continue to improve, they will play an increasingly crucial role in the analysis and interpretation of scRNA-seq data, enabling more personalized and targeted therapeutic approaches in the future.

1.5 Complementary Approaches for Studying Immune Cell Heterogeneity: Beyond Single-Cell RNA Sequencing

In addition to single-cell RNA sequencing (scRNA-seq), researchers have come up with a number of other approaches to studying the heterogeneity of immune cells. One such approach is bulk transcriptomics on purified and sorted cell samples, which serves as the closest competitor to scRNA-seq (131). By isolating specific cell types or subpopulations from a heterogeneous sample using cell sorting techniques before performing bulk RNA sequencing, researchers can obtain more focused gene expression profiles for each sorted population. However, analyzing sorted cell populations reduces the averaging effect, it still does not provide single-cell resolution. Also, cell sorting techniques rely on the expression of specific surface markers (132), which may not always accurately represent the functional or transcriptional state of a cell.

Other methods for studying cellular heterogeneity include single-cell proteomics (133), single-cell epigenomics (134), spatial transcriptomics (135), functional imaging (136), and CRISPR-based genetic screens (137) are some of these methods.

Single-cell proteomics is an approach that involves the analysis of protein expression at the single-cell level, allowing researchers to investigate cellular heterogeneity from a protein perspective (138). Single-cell proteomics techniques include single-cell proteomics by mass spectrometry (scp-MS) (139), mass cytometry (CyTOF) (140), single-cell Western blots (141), and Single Cell ProtEomics by mass spectrometry (SCoPE-MS) (142). Single-cell epigenomics is the technique that studies the epigenetic landscape of individual cells, providing insights into the regulatory mechanisms that contribute to immune cell heterogeneity (143). Methods such as single-cell ATAC-seq (Assay for Transposase-Accessible Chromatin with high-throughput sequencing) (144), single-cell ChIP-seq (Chromatin immunoprecipitation followed by sequencing) (145), and single-cell Hi-C (146) allow for the exploration of chromatin accessibility and histone modifications, respectively, at the single-cell level. Spatial transcriptomics is a method that enables the simultaneous measurement of gene expression and spatial information within a tissue, allowing researchers to study the cellular heterogeneity of immune cells within their native microenvironment. Techniques such as MERFISH (Multiplexed Error-Robust Fluorescence In Situ Hybridization) (147), Visium and Xenium platforms from 10X Genomics (148), and nanoString GeoMx ® Digital Spatial Profiler (DSP) platform (149) provide spatially resolved transcriptomic data. Functional imaging techniques like confocal microscopy, multiphoton microscopy, and super-resolution microscopy can visualize the

morphology, distribution, and interactions of immune cells within their native tissue context (150). Lastly, CRISPR-based genetic screens. The CRISPR/Cas9 system can be used to perturb gene function in immune cells, helping researchers to identify key genes and regulatory elements that contribute to immune cell heterogeneity. Pooled CRISPR screens can be combined with single-cell readouts to investigate the impact of genetic perturbations on cellular phenotypes at a high-throughput scale (151).

These parallel approaches, when used in conjunction with scRNA-seq, can provide a more comprehensive understanding of immune cell heterogeneity and offer valuable insights into the development, activation, and function of immune cells in various physiological and pathological contexts.

1.6 Discussion

scRNA-seq has indeed become a widely used experimental method for transcriptomic profiling. This is because scRNA-seq allows researchers to study gene expression at the level of individual cells, which provides greater resolution and sensitivity compared to bulk RNA-seq methods that measure gene expression in populations of cells. scRNA-seq involves isolating individual cells, reverse transcribing the RNA molecules within each cell into cDNA, and then sequencing the cDNA to determine the identity and abundance of each transcript in each cell (152). The resulting data can be used to identify subpopulations of cells, study gene expression heterogeneity within cell populations, and investigate cellular responses to different stimuli or conditions.

The development of new scRNA-seq technologies and analysis methods has also made it easier and more accessible for researchers to perform transcriptomic profiling at the single-cell level. As a result, scRNA-seq has become a powerful tool for understanding biological processes and diseases at the cellular level. However, scRNA-seq data analysis poses several challenges, including high levels of technical noise, sparsity, and batch effects, among others (153). We have summarized a variety of the statistical and computational methods that can address these challenges and have been developed for processing and analyzing scRNA-seq data. The key methods involved in data processing include quality control, normalization, and imputation. Quality control involves filtering out low-quality cells and genes that do not meet certain criteria, such as low expression levels or high mitochondrial gene content. Normalization tries to fix the expression levels of different genes and cells to make up for technical differences and make sure that samples are comparable. Imputation methods are used to fill in values for missing expressions and make the data less sparse. Methods for combining scRNA-seq data from multiple experiments and correcting for technical differences between batches or samples are also important. Dimensional reduction and clustering algorithms are used to see and find groups of cells with similar profiles of how their genes are expressed. Cell type prediction and annotation methods can be used to assign cell types to these populations based on known cell markers or transcriptional signatures. Differential analysis and gene set analysis are used to find genes or pathways that are expressed differently in different cell populations or when the experiment is done in different ways. Inferring time dynamics methods can be used to model the temporal relationships between different cell populations during

development or in response to different stimuli. Gene regulatory network methods can be used to identify regulatory interactions between genes and predict transcription factor activities. Machine learning approaches can be used to build predictive models for cell classification, trajectory inference, and other applications.

The advent of scRNA-seq technology has revolutionized our ability to study the immune system in health and disease. In particular, scRNA-seq has enabled researchers to investigate the heterogeneity of immune cell populations at the single-cell level and to uncover the molecular mechanisms underlying immune responses to infectious diseases and cancer. In Chapter 2, we will explore the various applications of scRNA-seq technology in the immune system. The immune system is a complex and dynamic system that plays a critical role in protecting the body against infectious diseases and cancer. The development, activation, and function of immune cells are tightly regulated at the transcriptional level, making scRNA-seq an ideal tool for investigating the heterogeneity and plasticity of immune cells.

In Chapters 3, 4, and 5, we will conduct our analysis by employing various tools discussed in this chapter, demonstrating their application and the insights they provide into the effects of sepsis, COVID-19, and smoking on the tumor microenvironment.

CHAPTER 2 Unraveling Immune System Complexity through Single-Cell RNA Sequencing

2.1 Introduction

The immune system exhibits remarkable complexity, comprising a diverse array of cell types, each characterized by unique gene expression patterns that consequently give rise to distinct functions. Immune cells also exhibit rapid and dynamic changes in gene expression in response to stimuli such as pathogens, antigens, or cytokines. They also exhibit intricate cell-cell interactions between various cell types that can also lead to rapid phenotype changes. Overall, single-cell RNA sequencing (scRNA-seq) has emerged as a powerful technology for dissecting the complexity and diversity of the immune system at unprecedented resolution (154, 155). By letting us figure out how genes are expressed in individual cells, scRNA-seq has helped us learn a lot about the diversity and functional states of immune cells. This has led to new insights into immune responses, homeostasis, and diseases related to the immune system (156, 157). The immune system is made up of many different types of cells, each of which has its own role in coordinating immune responses and maintaining immune tolerance (158). Traditional bulk RNA-seq methods, which measure the gene expression profiles of large groups of cells, don't capture the full range of cellular diversity and the complex relationships between different types and states of immune cells. scRNA-seq, on the other hand, has the potential to show the transcriptional landscape of each immune cell. This would give a completer and more detailed picture of the immune system and how it works.

2.2 Balancing Immune Homeostasis through Immune Cell Diversity

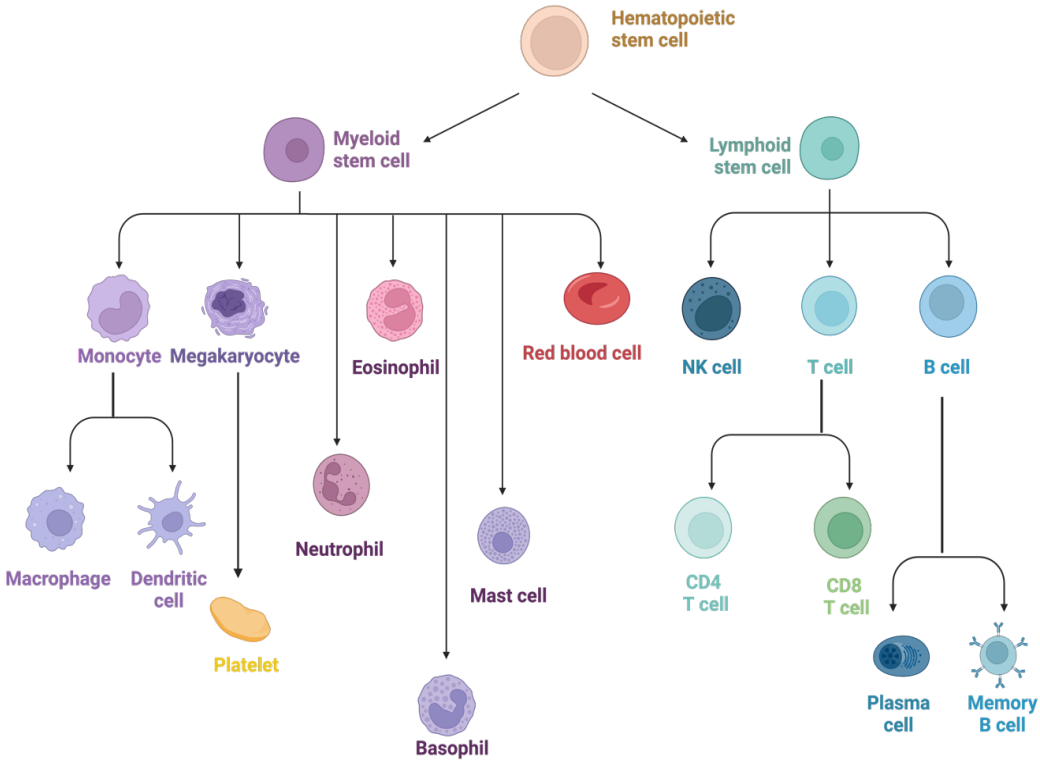

Figure 2.1 Overview of the immune system.

A schematic diagram illustrating the major components of the immune system. Figure generated by using BioRender.

Immune cell diversity is an essential part of the immune system, playing a crucial role in maintaining health and defending against a variety of pathogens and diseases. This diversity enables the immune system to recognize and respond to a broad spectrum of antigens, ensuring effective protection against infections while also promoting self-tolerance and preventing autoimmune disorders. The diverse immune cell populations can be broadly divided into two categories: innate immunity and adaptive immunity. Each variety of immune cell has a distinct pattern of gene expression, resulting in specialized

functions and capabilities. T cells, for instance, can differentiate into subsets such as helper T cells, cytotoxic T cells, and regulatory T cells, each of which plays a distinct function in the immune response. Similarly, B cells can differentiate into plasma cells that produce antibodies or memory B cells that provide long-lasting immunity. (Figure 2.1).

The innate immune system is the first line of defense against pathogens and is present since birth, providing nonspecific immunity that does not require prior exposure to pathogens. It responds rapidly to infections and has a broad spectrum of resistance to different types of pathogens (159). The innate immune system includes various types of cells, such as phagocytes, natural killer cells, and mast cells, as well as pattern recognition molecules/receptors, antimicrobial peptides, the complement system, inflammatory mediators, and cytokines produced by immune cells (160). Toll-like receptors (TLRs) are pattern-recognition receptors (PRRs) that recognize conserved structures in pathogens, known as pathogen-associated molecular patterns (PAMPs). TLRs trigger innate immune responses by activating signaling pathways that lead to the production of pro-inflammatory cytokines, chemokines, and type I interferons. This helps to eliminate the invading pathogen and recruit other immune cells to the site of infection (161).

The adaptive immune system, on the other hand, consists of T cells and B cells. The major histocompatibility complex (MHC) molecules on the surface of antigen-presenting cells (APCs) allow T cells to recognize and respond to specific antigens, while B cells make antibodies against specific antigens (162). The enormous diversity of T cell receptors (TCRs) and B cell receptors (BCRs) is generated through a process called V(D)J recombination, which shuffles and recombines variable (V), diversity (D), and joining (J)

gene segments during lymphocyte development (163). This process, along with somatic hypermutation in B cells, generates an almost limitless range of antigen specificities (164).

The immune system of humans is continuously active. This immune balance shifts from a tolerogenic state to an immunogenic/inflammatory state in response to microbial infection, tissue injury, or vaccination (165). For example, during lymphocyte development, progenitors expressing high-affinity receptors for self-antigens or those unresponsive to antigens are eliminated by apoptosis. The surviving lymphocytes form the naive T cell and B cell compartments. When naive lymphocytes encounter pathogens, they proliferate and activate to combat the invaders, eventually dying after the pathogens are cleared. A similar situation occurs with neutrophils during inflammation. Also, cytotoxic T lymphocytes and natural killer cells recognize and induce apoptosis in virus-infected, bacteria-infected, and transformed cancer cells (166). After the immunogen has been eliminated, homeostatic regulatory mechanisms return the system to its initial tolerogenic state. However, when diversity and control of immune cells break down, it can lead to autoimmune diseases and long-term inflammation (167).

2.3 Immune Cell Profiling and Heterogeneity

2.3.1. Identifying immune cell types and subtypes

One of the most important things that scRNA-seq has brought to the study of the immune system is the discovery of new cell types and subtypes, as well as the discovery of heterogeneity within known cell populations that were not known before (154). Cellular heterogeneity arises from differences between individual cells within a subtype or

population. These differences can be attributed to factors such as genetic variations, epigenetic modifications, and environmental influences, leading to variations in gene expression and functional characteristics even among cells of the same subtype (168). The hierarchy of subtypes typically refers to the classification of immune cells based on their lineage, developmental stages, and distinct functions (169). For example, T cells can be further differentiated into CD4+ T helper cells, CD8+ cytotoxic T cells, and regulatory T cells. In turn, CD4+ T helper cells can be classified into Th1, Th2, Th17, and T follicular helper (Tfh) cells, each with specific functions and cytokine profiles (170).

For example, scRNA-seq has been used to find rare and previously unknown immune cell populations, such as distinct populations of innate lymphoid cells (ILCs) (171) and proinflammatory microglia in developing glioblastomas and anti-inflammatory macrophages and myeloid-derived suppressor cells in end-stage tumors (172). These findings have expanded our understanding of immune cell diversity and have broad implications for our knowledge of immune regulation and function.

In addition to finding new types of cells, scRNA-seq has been used to find subpopulations of known immune cell types like T cells, B cells, and myeloid cells that were not known before. For example, scRNA-seq has been utilized to reveal subpopulations of Treg cells with distinct degrees of non-lymphoid tissue phenotype (173). Oh et al. revealed heterogeneity and clonal expansion of cytotoxic CD4+ T cells in tumors, with a gene signature of these cells predicting response to PD-1 blockade in metastatic bladder cancer patients treated with anti-PD-L1 (174). Similarly, scRNA-seq has been used to uncover novel B cell subsets. For example, researchers used scRNA-seq to uncover and

characterize heterogeneities in naïve B cells, classical memory B cells, and atypical B cell (ABC) subsets in chronic infectious diseases (175). Furthermore, scRNA-seq has provided insights into the heterogeneity of myeloid cells (176), the identification of different dendritic cell (DC) subsets (177) and monocyte-derived cells with unique functions (156).

Beyond the identification of immune cell types and subtypes, scRNA-seq has also been instrumental in characterizing the transcriptional and functional states of immune cells during various physiological and pathological conditions, such as infection, inflammation, and cancer (178-181). scRNA-seq has helped researchers learn more about how immune cells become active, change, and react to different stimuli by recording the transcriptional dynamics of each immune cell.

2.3.2. Characterizing immune cell states and activation

The characterization of immune cell states and activation has been facilitated by the application of scRNA-seq to various immune cell populations. For example, scRNA-seq has been used to study T cell activation and differentiation in response to stimuli. For example, Fernández-García, et al. identified a differential time-dependent reliance of activating T cells on the synthesis versus uptake of various non-essential amino acids as well as metabolic genes such as asparagine synthetase (Asns), whose expression dynamics modulated CD8+ T cell differentiation and anti-tumor response in a mouse melanoma model (182). Similarly, scRNA-seq has been utilized to characterize the activation and differentiation of B cells, revealing distinct transcriptional signatures associated with the development of B cells. Scharer, et al (183) explored the molecular reprogramming that

leads to heterogenous cell fate outcomes during B-cell differentiation to T cell-independent antigens in vivo, defining a path to antibody-secreting cells that includes an early decision point and two branches with distinct gene expression profiles and downstream outcomes. Also, scRNA-seq has been used to study how myeloid cells like macrophages, dendritic cells (DCs), and monocytes become activated and polarized. O'Neill, et al. used scRNA-seq to investigate inter-individual and inter-population variability in monocyte responses to influenza A virus (IAV), revealing widespread variability in the percentage of IAV-infected monocytes, a lower activation at basal state of an IRF/STAT-induced transcriptional network in individuals with high cellular susceptibility to IAV, a stronger resistance of CD16+ monocytes to IAV infection, CD16+ specific mRNA expression of IL6 and TNF in response to IAV, and a higher number of CD16+ monocytes and lower susceptibility to IAV infection among monocytes from individuals of African-descent (184).

Using scRNA-seq has also helped us learn more about the states of immune cells and how they become activated during diseases like infection, autoimmunity, and cancer. For example, scRNA-seq has been used to describe the immune cell landscape during viral infections (185-187). This showed that antiviral immunity and immune exhaustion are linked to different transcriptional signatures. In the context of autoimmunity, scRNA-seq has been used to study the immune cell states and activation dynamics in autoimmune diseases like systemic lupus erythematosus (SLE) (188), multiple sclerosis (MS) (189), and rheumatoid arthritis (RA) (190), giving new insights into disease pathogenesis and possible therapeutic targets.

Also, scRNA-seq has helped figure out the immune cell states and activation within the tumor microenvironment (TME). This has shown the complex relationship between cancer cells and immune cells like tumor-infiltrating lymphocytes (TILs) (191, 192) and tumor-associated macrophages (TAMs) (193). These studies have shed light on how different and changing the immune cell landscape is in the TME. They have also shown how the activation and function of immune cells play a role in how tumors grow, how they hide from the immune system, and how they respond to immunotherapy. Using scRNA-seq to describe the state of immune cells and how they are activated has also opened up new ways to develop therapies that target the immune system. For instance, insights gained from scRNA-seq studies have informed the design of novel immunotherapies, such as chimeric antigen receptor (CAR) T cell therapy (194) and immune checkpoint blockade (195), which aim to harness the power of the immune system to combat cancer and other diseases. Furthermore, scRNA-seq has been instrumental in identifying potential biomarkers for patient stratification (196) and predicting response to immunotherapy, providing a more personalized approach to treatment (197-199).

In conclusion, scRNA-seq has helped us learn a lot more about immune cell states and how they become active in response to different things, such as infections. scRNA-seq has made it possible to find out about new cell subsets and their roles in immune responses by giving a high-resolution description of how genes are expressed in a single cell. However, the distinction between subtypes and heterogeneity in immune cells can be challenging to pinpoint, as the classification of cell subtypes and the degree of heterogeneity are context-dependent and can be influenced by various factors, such as the

method of analysis and the biological conditions being studied. Also, how many clusters or subtypes exist can be fundamentally flawed due to its dependence on the threshold used for classification. This is particularly relevant in the context of single-cell analyses, where the resolution of data allows for the identification of finer distinctions between cells. The choice of clustering algorithms, distance metrics, and threshold parameters can greatly influence the resulting classification of cell populations. To thoroughly examine the heterogeneity in immune cells, researchers must be well-versed in the methodologies employed and have a deep understanding of the biological context under investigation.

2.4 Inflammatory Responses in Infectious and Chronic Diseases

Inflammatory responses play a crucial role in both infectious diseases and chronic diseases, but the nature and outcome of these responses can differ significantly between the two contexts. In the context of infectious diseases, the inflammatory response is an essential part of the immune system's defense against invading pathogens, such as bacteria, viruses, fungi, or parasites. When pathogens enter the body, the immune system mounts an inflammatory response to contain, control, and ultimately eliminate the infection (169). In chronic diseases, inflammation often plays a central role in disease pathogenesis and progression. These diseases are not caused by an acute infection but may be triggered or exacerbated by other factors, such as genetic predisposition, environmental factors, or dysregulation of the immune system. In these cases, the inflammatory response can become chronic and contribute to ongoing tissue damage and disease progression (200).

2.4.1. Dynamics of immune cell populations during infectious diseases

During infection, the immune system mounts a complex and dynamic response to eliminate pathogens while minimizing host tissue damage. Immune cells must adapt to the ever-changing landscape of the infection site and coordinate their activities to effectively combat the invading pathogens. By applying scRNA-seq to study immune cells in the context of infection, researchers can gain a detailed understanding of the diverse cellular responses and the underlying molecular mechanisms that drive their plasticity and specialization.

One key application of scRNA-seq in the context of infectious diseases is the study of immune cell dynamics in response to viral infections. For example, during the COVID-19 pandemic, scRNA-seq has been extensively utilized to characterize the immune response to SARS-CoV-2 infection. For instance, by using scRNA-seq researchers found that proinflammatory monocyte-derived macrophages were abundant in the bronchoalveolar lavage fluid from severe COVID-19 patients (201). Another single cell analysis revealed immune cells in COVID-19 patients have expansion of complement-expressing monocytes, increased megakaryopoiesis, clonally expanded CD8+ T cells, and altered B cell responses, providing insights into the coordinated immune response and potential therapeutic targets (202). Using scRNA-seq researchers also identified changes in peripheral immune cell phenotype, including a heterogeneous interferon-stimulated gene signature, HLA class II downregulation, and a developing neutrophil population related to plasmablasts, in COVID-19 patients with acute respiratory distress syndrome (203). Another example is from an HIV study, researchers found an increase in plasmacytoid dendritic cells was observed during the acute phase of HIV-1 infection, monocytes showed

enhanced inflammatory and antigen presentation functions, with the upregulation of MHC II, T cells exhibited a pro-inflammatory phenotype with elevated expression of inflammatory cytokines (204).

In addition to viral infections, scRNA-seq has been instrumental in elucidating the immune response to bacterial pathogens. For instance, Aldrich, et al. found Staphylococcus aureus craniotomy-associated biofilm infection elicits a transcriptionally diverse leukocyte response, chemokine receptors CX3CR1 and CCR2 are differentially expressed in infiltrating leukocyte populations within the biofilm infection. Infected animals demonstrated increased numbers of monocytes/macrophages with a distinct transcriptional profile (205).

ScRNA-seq has also been employed to investigate the dynamics of immune cell populations during parasitic infections, such as Toxoplasma gondii (T. gondii) infection. Patir, et al. found scRNA-seq revealed that CD16- monocytes are the primary cell type responsible for the transcriptional response to T. gondii. CD16- monocytes also showed increased phagocytic activity and lysosomal activity, suggesting a role in T. gondii clearance (206).

Finally, scRNA-seq has been employed to investigate sepsis. Sepsis is a life-threatening condition caused by the body's response to an infection. However, the activation of these immune cells is not uniform across all patients, leading to heterogeneity in the immune response and subsequent patient outcomes. ScRNA-seq can provide insights into the transcriptional diversity of immune cells during sepsis, allowing for the

identification of key genes and signaling pathways that contribute to disease progression. We will present more details in Chapters 3 and 4.

Besides its use in characterizing immune cell dynamics during an ongoing infection, researchers have also utilized scRNA-seq to investigate immune cell memory and development, including the study of the development of protective immunity after vaccination or infection. For example, Wen, et al. used scRNA-seq to analyze the transcriptomes of peripheral blood mononuclear cells (PBMCs) from COVID-19 patients in the recovery stage. They found an inflammatory immune signature in the early recovery stage of COVID-19, characterized by decreased T cells, increased inflammatory CD14++ monocytes, high levels of inflammatory gene expression in T cells, and novel B cell-receptor changes, highlighting potential therapeutic targets for the treatment of COVID-19 (207).

In conclusion, scRNA-seq has shown tremendous potential for advancing our understanding of infectious diseases. scRNA-seq has made it possible to find new cell types and functional states by giving researchers new insights into the transcriptional landscapes of individual immune cells during an infection. This has revealed complex interactions between the host and the pathogen. These studies have led to the discovery of new immune response mechanisms and improved our understanding of the pathogenesis of infectious diseases. Furthermore, scRNA-seq has helped identify biomarkers for early detection and diagnosis of infectious diseases, and has facilitated the development of more effective vaccines and treatments. Overall, scRNA-seq has made a big difference in how we

understand infectious diseases and has a lot of potential to help advance the field in the future.

2.4.2. Characterizing inflammatory responses in chronic diseases

In chronic diseases, inflammation often plays a central role in disease pathogenesis and progression. These diseases are not caused by an acute infection but may be triggered or exacerbated by other factors, such as genetic predisposition, environmental factors, or dysregulation of the immune system. In these cases, the inflammatory response can become chronic and contribute to ongoing tissue damage and disease progression. Examples of chronic diseases with an inflammatory component include autoimmune disorders, neurodegenerative diseases, cardiovascular diseases, and chronic respiratory diseases, etc. The emergence of scRNA-seq has enabled researchers to examine the individual cell types involved in chronic disease development, paving the way for the identification of therapeutic strategies to manage and potentially cure inflammatory responses in these conditions.

2.4.2.1 Autoimmune disorders

Autoimmune disorders, these are diseases in which the immune system mistakenly attacks healthy tissues in the body. Examples of autoimmune disorders include rheumatoid arthritis (RA), a chronic inflammatory disease affecting the joints, leading to pain, swelling, and stiffness. Systemic lupus erythematosus (SLE), a multisystem autoimmune disease that can affect various organs, such as the skin, kidneys, and heart. Multiple sclerosis (MS), a neurological disorder characterized by inflammation and damage to the myelin sheath

surrounding nerve fibers, resulting in various neurological symptoms. Inflammatory bowel disease (IBD), a group of chronic inflammatory conditions affecting the gastrointestinal tract, primarily including Crohn's disease and ulcerative colitis, characterized by persistent inflammation, abdominal pain, and gastrointestinal symptoms.

One of the ways scRNA-seq is used in the studies of autoimmune disorders is to find molecular changes in specific cell types. For example, scRNA-seq has revealed upregulated CD52 expression in B cells in systemic lupus erythematosus (SLE) patients (208). In rheumatoid arthritis (RA), researchers identified immune cell abnormalities and molecular pathway differences in anticitrullinated-peptide antibodies (ACPA) subtypes of rheumatoid arthritis in myeloid cells. The study also identifies the HLA-DR15 haplotype as a risk factor for developing the active disease in ACPA+ RA and highlights the potential for different therapeutic strategies based on ACPA status (190). In multiple sclerosis (MS), a study using scRNA-seq found B cells in the central nervous system were found to have various potential roles, such as secreting proinflammatory cytokines and chemokines, presenting autoantigens to T cells, producing pathogenic antibodies, and serving as reservoirs for viruses that trigger demyelination (209).

Another important way of scRNA-seq is used in autoimmune diseases is to find new immune cell populations and subpopulations that play a role in how the disease develops. In RA, scRNA-seq has revealed a previously unrecognized subset of fibroblast subpopulations that display distinct transcriptional profiles and functional properties, suggesting a potential role in joint destruction and tissue damage (210). Another study using scRNA-seq found two subsets of macrophages, CD163+ and CD206+ macrophages,

were associated with remission in RA patients, while another subset, CD163- CD206- macrophages, was associated with active disease. They also found that these macrophage subsets had distinct gene expression profiles and different abilities to produce inflammatory cytokines (211). In MS, scRNA-seq has identified sub-clusters of oligodendrocytes (OLs) in white matter (WM) areas in MS tissue that are under-represented, while some other are more prevalent, indicating different functional states of OLs in MS lesions. These findings suggest that OL heterogeneity may contribute to the variability of MS pathology and could be important for developing therapeutic approaches (212). In inflammatory bowel disease (IBD), scRNA-seq were used to profile T cells purified from the intestinal epithelium and lamina propria (LP) of Crohn's disease (CD) patients. The researchers found that intraepithelial lymphocytes (IEL) contain several unique T cell subsets, including NKp30+γδT cells expressing RORγt and producing IL-26 upon NKp30 engagement (213).

scRNA-seq has also given researchers important information about the genetic factors that increase the risk of autoimmune diseases. A study presents scRNA-seq data from 982 healthy human subjects' peripheral blood mononuclear cells (PBMCs), identifying 26,597 independent cis-eQTLs and 990 trans-eQTLs with cell type-specific effects on gene expression. The researchers show dynamic allelic effects in B cells during the transition from naïve to memory states and identify the causal route by which 305 risk loci contribute to autoimmune disease at the cellular level using Mendelian randomization (214).

In conclusion, scRNA-seq has had a big impact on the study of autoimmune diseases because it has helped researchers find cell types and subtypes that were not known to be involved in disease pathology before. scRNA-seq can show the diversity of immune cell populations by looking at gene expression on a single cell level. This is especially important for understanding the complexity of autoimmune diseases. Researchers can use scRNA-seq to find out which cell types are involved in autoimmune diseases, as well as their gene expression profiles, signaling pathways, and possible therapeutic targets. scRNA-seq can also help find the key driver genes, regulatory mechanisms, and signaling pathways that cause disease. By combining scRNA-seq data with genetic and clinical data, researchers can learn more about how autoimmune diseases work and make treatments that are more targeted and effective.

2.4.2.2 Neurodegenerative diseases

Neurodegenerative diseases, like Alzheimer's and Parkinson's disease, where chronic inflammation contributes to the progressive loss of neurons and their functions, leading to cognitive decline, motor impairment, and various other debilitating symptoms.

The researchers identified several novel cell types and subtypes in the single-cell transcriptomic analysis of Alzheimer's disease. These included new subtypes of inhibitory neurons that were selectively vulnerable to Alzheimer's disease pathology, as well as previously uncharacterized subpopulations of astrocytes and oligodendrocytes. The researchers also identified 40 transcriptionally distinct subpopulations of cells, some of which were preferentially overrepresented in Alzheimer's disease pathology and differentially represented between sexes. These findings provide new insights into the

cellular heterogeneity of the brain and may have important implications for understanding the molecular mechanisms underlying Alzheimer's disease (215). In Parkinson's disease (PD), researchers utilized scRNA-seq combined with comprehensive histological analyses to characterize intracerebral grafts from human embryonic stem cells (hESCs) and fetal tissue after functional maturation in a pre-clinical rat PD model. The study finds that neurons and astrocytes are major components in both fetal and stem cell-derived grafts, and identifies a previously unknown cell type resembling a class of perivascular-like cells in stem cell-derived grafts. The information provided from the study can help researchers identify the most effective cell types for transplantation and improve the success of cell replacement therapy for Parkinson's disease (216).

2.4.2.3 Cardiovascular diseases

Cardiovascular diseases, such as atherosclerosis, where inflammation plays a role in plaque formation and the narrowing of arteries. It encompass a range of conditions affecting the heart and blood vessels, including coronary artery disease, heart failure, and stroke, which are leading causes of morbidity and mortality worldwide, often associated with lifestyle factors, genetic predisposition, and underlying chronic inflammation.

In atherosclerosis, scRNA-seq has identified distinct features of T cells and macrophages in carotid artery plaques of patients with clinically symptomatic disease compared to asymptomatic disease, which may enable the design of more precisely tailored cardiovascular immunotherapies (217).

2.4.2.4 Chronic respiratory diseases

Chronic respiratory diseases are a group of long-term conditions affecting the lungs and airways, including chronic obstructive pulmonary disease (COPD), asthma, and pulmonary fibrosis, characterized by persistent breathing difficulties, reduced lung function, and often linked to genetic, environmental, and lifestyle factors.

In chronic obstructive pulmonary disease (COPD), scRNA-seq has been used to identify previously unrecognized changes in gene expression and cellular interactions in distinct epithelial, endothelial, and macrophage cell populations in COPD, highlighting the complexity and diversity of cellular injury and inflammation in COPD. The researchers found Hedgehog-interacting protein (HHIP) is a genetic determinant that regulates chronic obstructive pulmonary disease (COPD). And identified HHIP-expressing alveolar epithelial type II (AT2) cells are a subpopulation of epithelial cells that mediate COPD heritability and have aberrant expression of metabolic, antioxidant, and cellular stress response genes in COPD, suggesting their potential as therapeutic targets (218).

2.5 Cancer Immunology and Immuno-oncology

2.5.1. Tumor microenvironment and immune cell infiltration

ScRNA-seq has become an indispensable tool for studying the tumor microenvironment (TME) and immune cell infiltration, providing unparalleled insights into the cellular and molecular interactions that drive tumor progression, immune evasion, and response to therapy (219-223). By profiling the transcriptomes of individual cells within the TME,

scRNA-seq has uncovered the cellular heterogeneity, immune cell activation states, and functional alterations that underlie tumor development and resistance to therapy.

A key application of scRNA-seq in the study of the TME is the identification of novel immune cell populations and subpopulations that contribute to tumor progression and immune evasion. For example, in colon cancer, researchers generated an atlas of immune and non-immune cells from human colorectal cancer (CRC) patients using scRNA-seq approaches and identified two distinct tumor-associated macrophages (TAM) populations, consisting of C1QC+ and SPP1+ TAMs, both of which may arise from an intermediate state of FCN1+ monocyte-like cells in the tumor , the study also found the expansion of Bhlhe40+ Th1-like CD4+ T cells may be downstream of Ccl22+ cDC1 activation, and may potentiate cDC1-mediated CD8+ T cell infiltration, expansion, and anti-tumor function, which provided mechanistic insights for immunotherapies (224). In early-stage lung adenocarcinoma (LUAD), with scRNA-seq, researchers found diverse stromal and tumor cell types with complex interactions, showing pro-tumoral functions of tumor-associated macrophages and exhausted and regulatory features of tumor-infiltrating T cells, and revealing that the upregulation of ELF3 in response to inflammatory cytokines from immune infiltrates triggers the activation of PI3K/Akt/NF-κB pathway and elevated expression of proliferation and anti-apoptosis genes (225). In breast cancer, researchers use scRNA-seq identify both heterogeneity and core gene expression signatures for subtype-specific breast cancer cells and classified non-tumor cells into three immune cell types with activating and suppressive gene expression signatures, suggesting dynamic immune cell interactions and a distinct immune system status in each tumor (226).

ScRNA-seq has also been invaluable for characterizing the dysfunctional immune responses that occur within the TME, providing insights into the activation and effector functions of various immune cell types. For example, in non-small cell lung cancer (NSCLC), scRNA-seq has been used to study T cells in the TME, identified exhausted T cells were found to be enriched in NSCLC tumors and were identified as a distinct subpopulation with unique gene expression patterns (227). In colorectal cancer (CRC), scRNA-seq has shown that the CD8+ effector and "exhausted" T cells exhibited high clonal expansion and were independently connected with tumor-resident CD8+ effector memory cells, they also identified two IFNG+ TH1-liell clusters in tumors that were associated with distinct IFNγ-regulating transcription factors (228).

In conclusion, scRNA-seq has provided insights into the TME and immune cell infiltration, revealing substantial heterogeneity and complex interactions among tumor cells, stromal cells, and immune infiltrates in various types of cancer.

2.5.2. Immune checkpoint blockade and personalized immunotherapy

scRNA-seq has given new insights into studying immune checkpoint blockade and personalized immunotherapy. Shed light on the cellular and molecular mechanisms that control response to therapy and resistance (199, 229-231). By looking at the transcriptomes of individual immune cells in the tumor microenvironment (TME) and peripheral blood, scRNA-seq has helped researchers find predictive biomarkers, develop strategies for optimizing therapies, and guide personalized treatment approaches (232).

One of the key applications of scRNA-seq in the context of immune checkpoint blockade is to provide the molecular mechanisms that underlie the response to immune checkpoint blockade. This has led to a better understanding of how different types of immune cells are activated and what their effector functions are. For example, in triple-negative breast cancer (TNBC), using scRNA-seq, the researchers identified several immune cell subsets, including CD8+ T cells, CD4+ T cells, B cells, and NK cells, that were associated with responses to PD-L1 blockade. They found that the frequency of CD8+ T cells expressing immune checkpoint molecules such as PD-1 and TIM-3 was higher in patients who responded to PD-L1 blockade. The researchers also identified CXCL13+ T cells were found to be important for effective responses to anti-PD-L1 therapies (231). In NSCLC, researchers investigated the response to anti-PD-1 therapy in lung cancer patients at the single-cell level. They found increased levels of precursor exhausted T cells (Texp) in responsive tumors. In addition, TexP cells were shown to undergo clonal expansion during therapy, indicating a potential mechanism for durable responses to anti-PD-1 treatment (230).

ScRNA-seq has also been very helpful in figuring out the landscape of immune checkpoint inhibitors-resistant cell states. For example, in melanoma, scRNA-seq has revealed a resistance program expressed by malignant cells that promotes T cell exclusion and immune evasion in melanoma patients receiving immune checkpoint inhibitors (ICIs), and CDK4/6-inhibition was shown to repress this anti-tumor progress, induce senescence, and reduce tumor outgrowth in mouse models when given in combination with immunotherapy (199).

In addition to its applications in immune checkpoint blockade, scRNA-seq has facilitated the development of personalized immunotherapy approaches, such as targeting tumor-specific antigens (TSAs) using T-cell receptors (TCRs) from naturally occurring tumor antigen-specific T (Tas) cells in a patient to generate the efficient and safe immune response (232). Similarly, scRNA-seq has been employed to investigate the immune responses elicited by personalized cancer vaccines. Incorporating scRNA-seq to assess immune responses to vaccine candidates offers a comprehensive approach to evaluating host responses, including gene expression, antigen-specificity, clonality, and copy number variants (233).

ScRNA-seq has played a critical role in advancing our understanding of the immune response to immune checkpoint blockade and personalized immunotherapy. By profiling individual immune cells, scRNA-seq has helped identify key cell subsets associated with response and resistance to therapy and enabled the development of more precise and effective personalized immunotherapies targeting tumor-specific antigens. As scRNA-seq technology and tools for analyzing data continue to improve, we should be able to enhance the design and testing of immunotherapies.

2.6 Discussion

Emerging technologies in scRNA-seq immune system research have the potential to transform our understanding of immune cell diversity, function, and interactions. These advancements are expected to accelerate the development of novel therapeutic strategies and personalized medicine approaches. In this chapter, we illustrated the scRNA-seq

applications in identifying immune cell types and subtypes, characterizing immune cell states and activation, studying the dynamics of immune cell populations during infectious diseases, immune cell dysregulation in autoimmune diseases, inflammatory responses in chronic diseases, tumor microenvironment, and immune cell infiltration, and how scRNA-seq helped immune checkpoint blockade and personalized immunotherapy.

Nonetheless, numerous unanswered questions remain, and further exploration of disease-related heterogeneity within the immune system is necessary to uncover the answers. In Chapter 3, we will explore the application of scRNA-seq technology in the study of sepsis, a life-threatening condition characterized by a dysregulated immune response to infection. Sepsis is a major cause of morbidity and mortality worldwide, and despite extensive research efforts, its underlying molecular mechanisms remain poorly understood. In Chapter 4, we will focus on the application of scRNA-seq technology to the study of platelets in fatal cases of sepsis and COVID-19. Platelets are small, anucleate cells that play a critical role in hemostasis and immune regulation (234). In sepsis and COVID-19, platelet dysfunction is a common feature and is associated with poor clinical outcomes (235, 236). By using scRNA-seq to analyze platelet transcriptomes at the single-cell level, researchers can identify platelet subpopulations with distinct gene expression profiles and functional states and investigate how these subpopulations are altered in sepsis and COVID-19. In Chapter 5, we will delve into the challenges and opportunities of using scRNA-seq to study the effects of smoking on the immune environment of lung cancer. By looking at scRNA-seq data from lung cancer patients who have smoked in the past,

researchers can find out what kinds of immune cells are in the tumor microenvironment and how smoking changes these populations.

CHAPTER 3 Dynamic Changes in Human Single Cell Transcriptional Signatures During Fatal Sepsis

3.1 Introduction

Sepsis is an inflammatory syndrome caused by a systemic infection that can lead to multisystem organ failure and death. Sepsis is responsible for a significant percentage of in-hospital healthcare costs both in the United States and worldwide, and it is associated with a high mortality rate (237, 238). Despite many efforts, no targeted therapeutics against sepsis have been developed in the last decades. One acknowledged challenge is the complexity of the disease involving the competing interplay between rampant inflammation (cytokine storm) and, paradoxically, the almost simultaneous shutdown of the immune system (immunoparalysis) (239, 240). Another sepsis challenge is that some patients with nearly identical clinical phenotypes quantified by qSOFA and APACHE scores die at every stage of the disease while others survive (241). This supports the need to understand the molecular level host response to sepsis, which has been studied in blood and peripheral blood mononuclear cell (PBMC) profiling studies by gene expression or proteomics methods (242). These studies identify several prognostic biomarkers, such as lactate, procalcitonin, C-reactive protein (CRP), ferritin, and erythrocyte sedimentation rate (ESR), which along with clinical scores, are standardly utilized to evaluate sepsis patients and determine their care.

However, connecting these high-level observations to accurate clinical outcomes presents an unresolved challenge, likely due to the complexity and heterogeneity of this disease. Many studies have been conducted to identify a potential sepsis molecular

signature to gain molecular insights into this heterogeneity, which could aid in diagnosis or treatment (243). Recently, the first single-cell analysis of the status of immune cells in sepsis was reported, which identified abnormal monocyte states associated with immune dysregulation (244). Here, we apply the same approach to focus on the additional question of immune cell trajectory immediately after diagnosis in sepsis survivor and nonsurvivor outcomes. We performed single-cell transcriptomics analyses in fatal or surviving sepsis using a within-subject study design of PBMC collected from septic patients in the intensive care unit (ICU) at 0 and 6 h post sepsis diagnosis. There is clinical utility in choosing a 6-h time point, as sepsis resuscitation bundles (both in the United States and internationally) have been modeled after landmark studies (245) that demonstrated a significant reduction in mortality with aggressive resuscitation in the first six hours after presentation. Additionally, there is robust data (246) that early administration of intravenous antibiotics in the first 60 min after the recognition of septic shock significantly improves mortality. While subsequent trials (PROMISE, ARISE, PROCESS) showed no difference in clinician-driven versus protocol- driven resuscitation at 6 h, the Centers for Medicare and Medicaid Services (CMS) and the Surviving Sepsis Campaign continue to advocate for hospitals and clinicians to use the 6-h time point from initial Emergency Department presentation as a benchmark for resuscitation (247-249). Thus, this time point was chosen to assess molecular changes in the patients after they had received their initial resuscitation (including intravenous antibiotics, intravenous fluids, and vasopressor support).

Our timed analyses revealed the emergence and continuous changes in abnormal immune cells, including new types of cells unique to sepsis and classical cell-types present

in both sepsis and healthy controls, but with abnormal gene expression profiles and changes in population ratios. Specifically, we observed that fatal sepsis is associated with the expansion of platelets and erythroid precursors and the immunosuppressive trend of monocytes. Additionally, we identified CD52 expression in lymphocytes as a potential biomarker and therapeutic target for sepsis, where it correlated with increased lymphocyte activation and survival outcomes. At the cellular level, we also observed a switch in the metabolic state from oxidative phosphorylation in survivors to glycolysis in non-survivors. Last, we observed that fatal sepsis shared many gene signatures with severe COVID- 19 patients, indicating convergent molecular pathways in severe disease. These included genes associated with increased platelet activity, elevated erythroid precursors, and chemokine expression in monocytes.

Overall, this study, which focused on within-subject analyses of PBMC over time, offers a unique perspective on the dynamic changes in immune cells in fatal sepsis. Specifically, we identify abnormal immune cell subsets, changes in functional pathways, and molecular signatures at the single-cell resolution associated with fatal or surviving outcomes in sepsis. This study provides foundation data and identifies specific cell subsets and molecular pathways that can be further explored to better predict and possibly modify sepsis outcomes.

3.2 Materials and Methods

3.2.1 Human blood collection and harvest of PBMCs.

Peripheral blood was collected from non-sepsis donors from the Riverside Free Clinic and septic patients with signed informed consent and approval of the University of California, Riverside (UCR, #HS-17-707), and Riverside University Health System (RUHS, #1024190-3) Institutional Review Board. Sepsis patient enrollment was performed according to the following inclusion criteria: (1) Admission to Intensive Care Unit; (2) Age greater than or equal to 18 years old; (3) Suspected or confirmed infection; (4) qSOFA score ≥ 2 (qSOFA variables: altered mentation [GCS ≤ 13], systolic blood pressure < 100 mm Hg and respiratory rate > 22 breaths/min) and/or; (5) Lactate greater than or equal to 2.0 mmol/L and on vasopressor therapy to maintain MAP > 65 mm Hg after 30 mL/kg intravenous fluid bolus.

3.2.2 PBMC analysis

Blood was recovered in Vacutainer glass collection tubes with heparin (BD Biosciences). PBMC were isolated by gradient centrifugation with Histopaque-1077. Plasma was recovered for cytokine quantification by cytokine bead array (BD Biosciences) and resistin ELISA (Peprotech). Cell aliquots were frozen in liquid nitrogen. Following blood draw, PBMC isolation was performed within 24 h through density gradient centrifugation, and cells were stored immediately in liquid nitrogen. Flow cytometry characterization of PBMC involved incubation with Human TruStain FcXTM (Biolegend) and staining with primary Abs: CD14 (HCD14, Biolegend), CD16 (3G8, Biolegend), CD66b (G10F5,

eBioscience), CD3 (OKT3, eBioScience). Samples were acquired on a BD LSRII and analyzed on FlowJo (v10).

3.2.3 10X Genomics

For single-cell sequencing, thawed PBMC live cells were recovered by column-based dead cell removal kit (Miltenyi), and viable cells were confirmed by hemocytometer counting (>85% viable). A total of 15,000 cells per sample were loaded onto the 10x genomics platform, and cDNA libraries were prepared according to the manufacturer's instructions (Chromium Next GEM Single Cell V3.1). Samples were sequenced at the UCSD Genomics center on the NovaSeq platform at 250 M reads/sample.

3.2.4 Process and quality control of the single-cell RNA-seq data

The Cell Ranger Software Suite (v.3.1.0) was used to perform sample de-multiplexing, barcode processing, and single-cell 5′ unique molecular identifier (UMI) counting. Specifically, splicing-aware aligner STAR was used in FASTQs alignment. Cell barcodes were then determined based on the distribution of UMI counts automatically. The following criteria were applied to each cell of four sepsis samples and two healthy controls: gene number between 200 and 6000, UMI count > 1000, and mitochondrial gene percentage < 0.2. After filtering, a total of 57,133 cells were left for the following analysis. Finally, all samples' filtered gene-barcode matrix was integrated with Seurat v.3 (64) to remove batch effects across different samples.

3.2.5 Dimensionality reduction, clustering, and consensus-based cell-type annotation

We first analyzed scRNA-seq data from 57,133 cells with 4761 cells on the average per sample. Two-time points were analyzed per sepsis patient. Uniform Manifold Approximation And Projection was used to visualize the cell populations (Figure 3.2A and B). The filtered gene barcode matrix was normalized using "LogNormalize" method from Seurat package v.3 with default parameters. In the next step, the vst method implemented in the FindVariableFeatures function of the Seurat package was applied to find the top 2000 most variable genes. It was followed by the principal component analysis (PCA), and the application of the uniform manifold approximation and projection algorithm for cell data visualization performed based on the top 50 principal components. Then the graph-based clustering was performed by applying the FindClusters function of the Seurat package on the PCA-reduced data. With the resolution set to 1.0, 57,133 cells were grouped into 34 clusters. The first method of assignment of cell types to cell clusters was based on their canonical markers: B cells (MS4A1), $CD14^+$ monocytes (CD14 and LYZ), $CD4^+$ T cells (IL7R, CCR7, and CD27), $CD8^+$ T cells (CD8A), DCs (FCER1A, CST3, CD123, and GZMB), erythroid precursors (GYPB and AHSP), $FCGR3A^+$ monocytes (FCGR3A and MS4A7), neutrophils (JAML and SERPINB), NK cells (GNLY and NKG7), and platelets (PPBP). Independently from this initial marker-based cell type assignment, we applied cell-type annotation tools SingleR (75) and scCATCH (76). The SingleR program first identifies genes with significant variation between cell types in the reference data set, compares each cell's scRNA-seq data with each sample from the reference data set, and performs iterative fine-tuning to select the most likely cell type of each cell. The microarray dataset from

Human Primary Cell Atlas Data with assigned labels was used as the reference. Finally, each cluster was assigned a cell type with the highest percentage of cells assigned to that type by SingleR. The third applied method of cell type assignment was scCATCH, where cell types are assigned using the tissue-specific cellular taxonomy reference databases (250-252) and the evidence-based scoring protocol. Our final assignment of cell types to clusters was based on the consensus of the three methods mentioned above as follows: first, each cluster was assigned a cell type selectedmarkert methods if possible. If each method gave a different result, then the priority was given to the assignment based on canonical markers. If the markers-based assignment was inconclusive, the consensus assignment was based on the results from SingleR method.

3.2.6 Differential gene expression analysis and functional annotation of genes

The MAST method (253) from the Seurat v.3 package (implemented in FindAllMarkers function) was used with default parameters to perform differential gene expression analysis. A difference in gene expression was considered significant if an adjusted P-value was below 0.05. The false discovery rate (FDR) adjustment was performed by MAST. Only genes with FDR-adjusted P-values < 0.05 were considered in the second step of DEG analysis, where we analyzed differences between the results of the comparisons listed earlier. Pathway enrichment analysis was performed by clusterProfiler (254) using database Gene Ontology biological process terms (GO-BP) and Kyoto Encyclopedia of Genes and Genomes pathways. The clusterProfiler program was used for statistical

analysis and visualization of functional profiles for DEGs with FDR- adjusted P-value < 0.05.

3.2.7 Comparison of module scores

We used cell module scores to measure the degree to which individual cells expressed certain predefined expression gene sets. The AddModuleScore function from the Seurat v.3 package with default settings was used to perform all calculations and comparisons of module scores. We compared the expression of modules such as T cell activation (GO:0042110), B cell activation (GO:0042113), coagulation (GO:0050817), platelet activation (GO:0030168), OXPHOS, glycolysis, MHC class I, MHC class II, translation initiation (GO:0006413), response to type I IFN (GO:0034340), response to IFN-γ (GO:0034341), response to IFN-β (GO:0035456), Coronavirus disease COVID-19 (hsa05171), and HLA-DR related genes. The lists of genes defining these modules were prepared based on Gene Ontology and literature. Genes without detectable expression in our data were ignored. The sets of genes defining the modules used in our analysis are listed in Table B.2.

3.2.8 Statistics

The statistical tools, methods, and significance thresholds for each analysis are described in the Results or Materials and Methods section or in the figure legends.

3.3 Results

3.3.1 Subject characteristics

To gain a molecular understanding of the immune state in surviving or non-surviving sepsis outcomes, we performed retrospectively single-cell RNA sequencing on PBMCs from 5 hospitalized patients with Gram-negative bacterial sepsis at 0 and 6 h post diagnosis. Three patients survived (Survivor, S) and were discharged from the ICU; two patients had fatal disease courses (Nonsurvivor, NS). Clinical parameters (qSOFA and APACHE scores) were high and could not distinguish between sepsis survivors and nonsurvivors, and all sepsis patients had plasma cytokine levels that were dramatically elevated compared to baseline nonsepsis volunteers (Table 3.1). These results are consistent with the known phenomenon of the sepsis-induced cytokine storm.[4] In contrast, re-stimulation of PBMC from the same sepsis patients with LPS led to reduced TNF-α secretion as compared to PBMC from nonsepsis controls (Table 3.1), suggesting monocytic deactivation that has been reported in sepsis immunoparalysis (255). Flow cytometric analysis of PBMC was performed according to previously published gating strategies (256-258), and revealed different immune subset distribution with sepsis patients, including increased neutrophils but reduced T cell subsets, especially in the nonsurvivors (Figure 3.1A and B). We also observed the emergence of cell subsets that we were unable to define with common PBMC surface antibodies (Figure 3.1C, “other”). Together, these data characterize clinical and peripheral immune profiles in sepsis. However, more detailed sub setting of specific immune cells and insights into how temporal changes in their gene expression relate to

sepsis outcome were lacking, which we addressed by single-cell RNA sequencing.

Clinical parameters, cytokine levels in the plasma, and supernatant following LPS stimulation (10 ng/mL) of PBMCs. APACHE II: Acute physiology and chronic health evaluation II; SOFA: Sequential organ failure assessment; N.D.: not detected, n/a: not applicable.

Table 3.1 Characteristics of enrolled non-sepsis volunteers and sepsis patients at sepsis recognition (T0).

	Non-sepsis control (n=2)		**Sepsis non-survivor (n=2)**		**Sepsis survivor (n=3)**		
Gender	Male	Female	Male	Female	Male	Female	Female
Age range	35-40	45-50	90-95	65-70	45-50	65-70	70-75
Sepsis etiology	n/a	n/a	*E.coli* bacteremia		*E.coli* bacteremia		
APACHEII	n/a	n/a	18	38	31	41	19
SOFA	n/a	n/a	11	16	11	15	7
Time of death (days post enrollment)	n/a	n/a	<30	1	n/a	n/a	n/a
Plasma cytokines (ng/mL)							
Resistin	22.5	36.7	202	135.9	147	281	92
IL-6	N.D	0.002	30.2	142.3	133	2.48	0.31
IL-8	0.03	0.026	6.65	27.2	41.7	0.61	0.4
IL-10	N.D.	N.D.	0.13	9.71	0.52	0.15	0.39
LPS-induced TNFα (ng/mL)							
TNFa	0.656	0.979	0.45	0.005	n/a	0.047	0.1

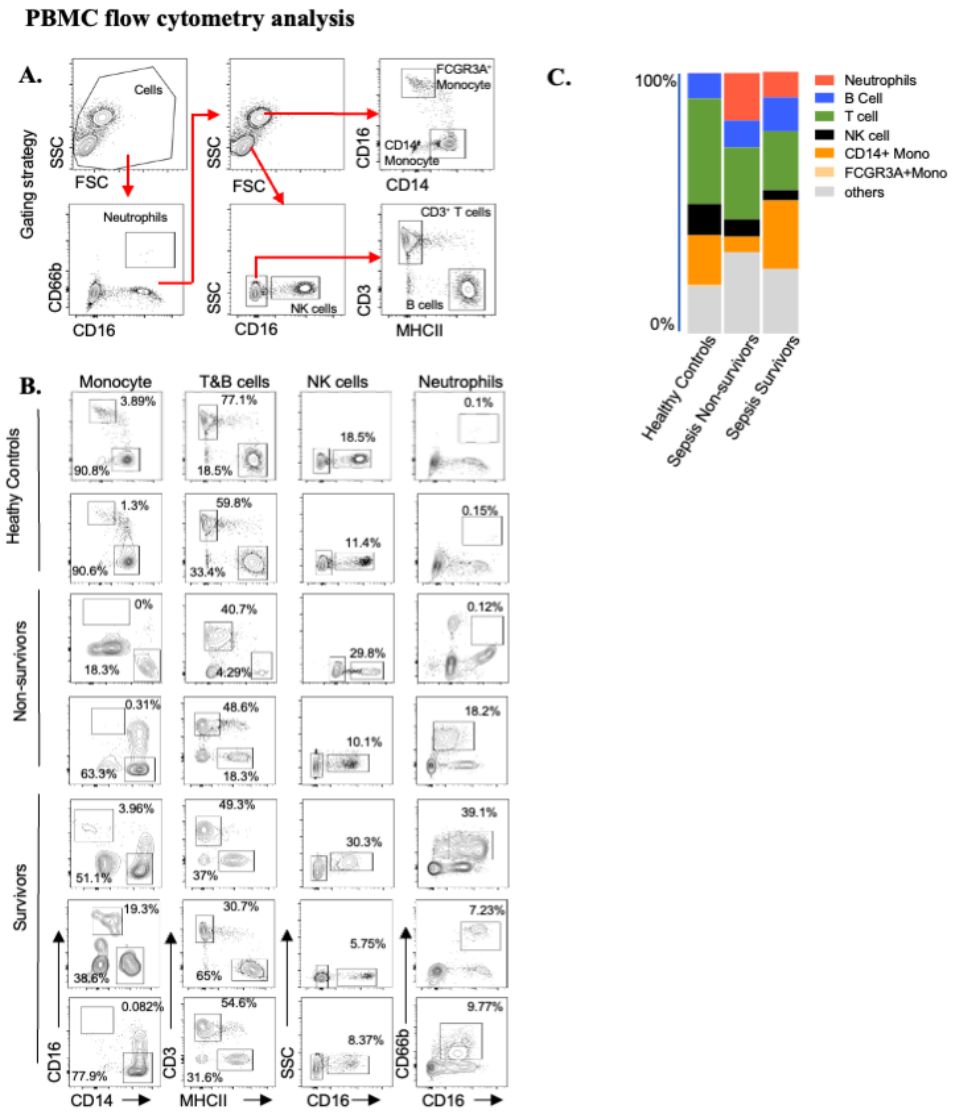

Figure 3.1 Flow cytometric analysis of PBMC from healthy control (HC), non-survivor (NS), and survivor (S) sepsis patients at first blood collection (T0).

(A) Gating strategy. Neutrophils (CD66b+CD16+), FCGR3A+Monocytes (CD66b−SSCHiCD16+CD14low), CD14+Monocytes (CD66b−SSCHiCD16−CD14hi), NK cells (CD66b−SSClowCD16+), T cells (SSClowCD3+), and B cells (SSClowMHCII+). (B) Frequency of immune cell subsets in PBMC. (C) Immune cell proportions in PBMC.

3.3.2 Single-cell transcriptomics identify immune cell subsets associated with sepsis severity

Single-cell RNA-seq was performed on a 10× Genomics platform. Consensus-based assignment of cell types from all subjects revealed 11 cell types: $CD4^+$ T cells (24%), $CD8^+$ T cells (8%), B cells (20%), Natural killer (NK) cells (9%), $CD14^+$ monocytes (16%), $FCGR3A^+$ monocytes (11%), dendritic cells (DC) (3%), erythroid precursor cells (2%), platelets (6%), neutrophils (1%), and common myeloid progenitor cells (CMP) (< 1%) (Figure 3.2A; Table B.1). Comparison between the samples indicated variability between the individuals and no striking changes between cell subsets within 6 h (Figure 3.2B and C; Table B.1). Among the sepsis patients, the female nonsurvivor (P50) showed an immune profile that was distinct from the ones in the male nonsurvivor (P34) and survivor samples. This individual showed advanced sepsis disease with fatality within 24 h, while the other nonsurvivor passed away within 30 days and had cell distributions more similar to that of the survivors (Figure 3.2C; Table B.1). Therefore, for subsequent analysis, P50 was designated nonsurvivor, late-stage sepsis (NS LS) while P34 was labeled nonsurvivor, early-stage sepsis (NS ES). Sepsis fatality, even within 24 h, was not unexpected in these studies, as these were patients admitted to the ICU with sepsis/septic shock, multisystem organ failure, and high qSOFA scores.

Analysis of immune subsets in all the sepsis patients indicated B cell depletion that followed disease severity: 25% in healthy controls (HC), 21% in S, 16% in NS ES, and 6% in NS LS (Figure 3.2C; Table B.1). $CD4^+$ T cell lymphopenia was observed in NS LS (8%

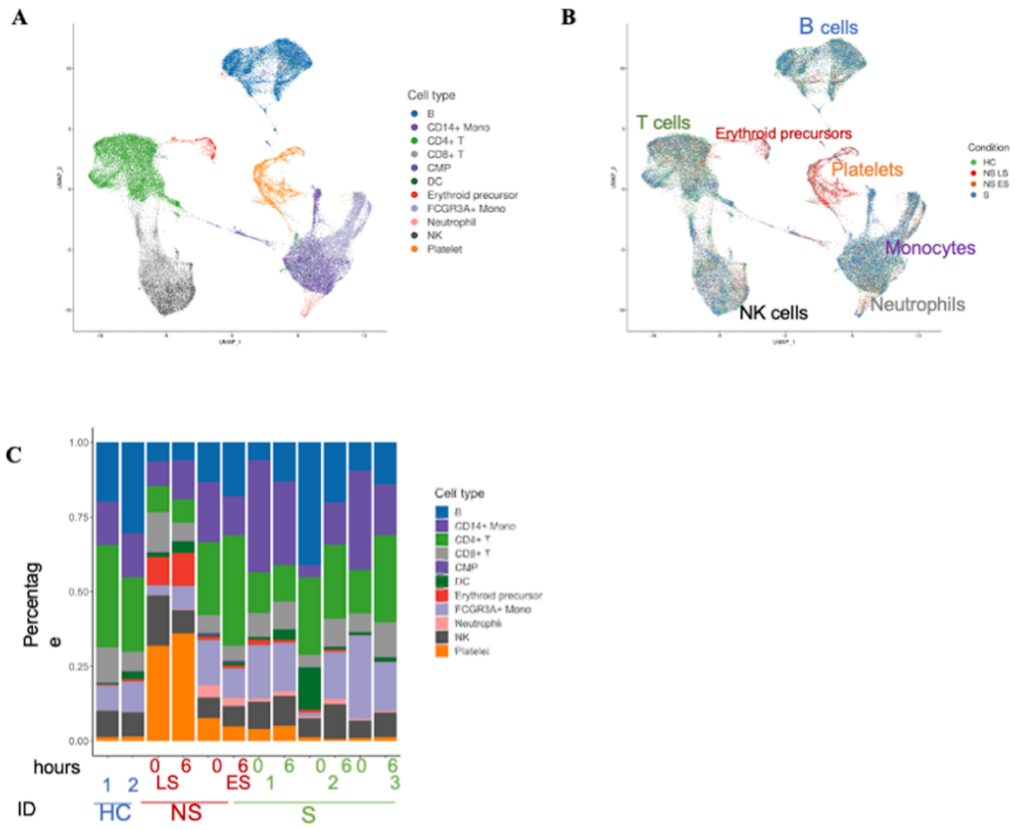

Figure 3.2 Single-cell transcriptional profiling of PBMC from healthy controls and gram-negative sepsis patients.

(A) Cell type UMAP representation of all merged samples. A total of 11 cell types were identified by the consensus method. In total, 57,133 cells are depicted. (B) Sample of origin UMAP representation of all merged samples. Cells were colored by the condition. (C) Bar plots showing the fraction of each sample.

compared to 20% in HC, S, and NS ES). We noted an increased proportion of platelets in NS samples, especially in NS LS (34%). The proportion of erythroid precursor cells was also increased with sepsis severity, from 1% in HC, 2% in S, to 6% in NSES, and 10% in NSLS (Table B.1). We investigated potential gender effects among single-cell transcriptomics and found that there were no batch effects related to gender differences

with this small sample size (Figure A.1). Together, this immune subsetting data by scRNA-seq indicate that lymphocyte subsets are reduced in sepsis, especially in fatal outcomes, and identify the emergence of platelet and erythroid precursors in late-stage fatal sepsis. However, no striking changes in immune subsets were observed within 6 h, prompting us to investigate transcriptional changes within the individual cell types.

3.3.3 Platelet responses are a hallmark of fatal sepsis with similar transcriptional pathways to severe COVID-19 disease

The role of platelets in the development of sepsis pathophysiology is increasingly recognized. Recent studies show that platelets are altered in sepsis and that transcriptional and translational changes in platelets are related to mortality (259). However, timed analysis during the critical early timepoints post sepsis diagnosis has not been performed. Analysis of pathway module scores revealed that platelets in sepsis patients presented with coagulation abnormalities (GO term coagulation, GO:0050817) that was exacerbated over time, especially in fatal disease (Figure 3.3A). We also found increased platelet activation (GO:0030168; Figure 3.3B) and ATP production modules, which included oxidative phosphorylation (OXPHOS) genes and glycolysis genes (260) (Figure 3.3C and D). These results are consistent with previous studies that have shown that platelet aggregation is fueled primarily by glycolysis, and that reticulated platelets are more prothrombotic and hyperreactive than mature platelets (261). MHC class I-related genes such as HLA-A, HLA-B, HLA-C, HLA-E, and HLA-F also followed the same trend according to sepsis disease severity (Figure 3.3E), indicating CD8 T cell dysregulation by platelets (234, 262).

Translation initiation modules (GO:0006413) were also changed in sepsis with the lowest score in nonsurvivors, especially in the NS LS, indicating a halt in protein translation as a result of disease (Figure 3.3F). Conversely, interferon response modules, including response to type I IFN (GO:0034340), IFN-γ (GO:0034341), and IFN- β (GO:0035456) exhibited the opposite trend with increased scores in sepsis, especially in the nonsurvivors (Figure 3.3G–I). This phenomenon might also reflect a general suppression of the protein synthetic apparatus by type I IFN (263).

We next focused on dynamic changes in the platelets within 6 h. Interestingly, analysis of 6-h trajectory transcriptional changes in individual sepsis patients' platelets shared pathways induced in platelets in severe COVID-19 infections, suggesting that platelet transcriptional changes are predictors of severe disease regardless of infection etiology. The most distinguishing pathway, as identified by VENN diagram, which was down-regulated in NS T0→T6 while up-regulated in S T0→T6, was hemostasis (Figure A.2). We analyzed the GO term "positive regulation of hemostasis" (GO:1900048; Figure 3.3J) and confirmed that only sepsis survivors exhibited an upward trend, suggesting improved platelet function. Conversely, the most shared pathways that were up-regulated in NS T0→T6 but down-regulated in S T0→T6 included translation initiation, ribosome, and COVID-19 (Figure A.2). To further investigate shared pathways in COVID-19 and sepsis, we investigated genes from Kyoto Encyclopedia of Genes and Genomes Coronavirus disease COVID-19 (hsa05171) and found that the module scores were significantly increased in NS T6 platelets and decreased in S T6 platelets (Figure 3.3K).

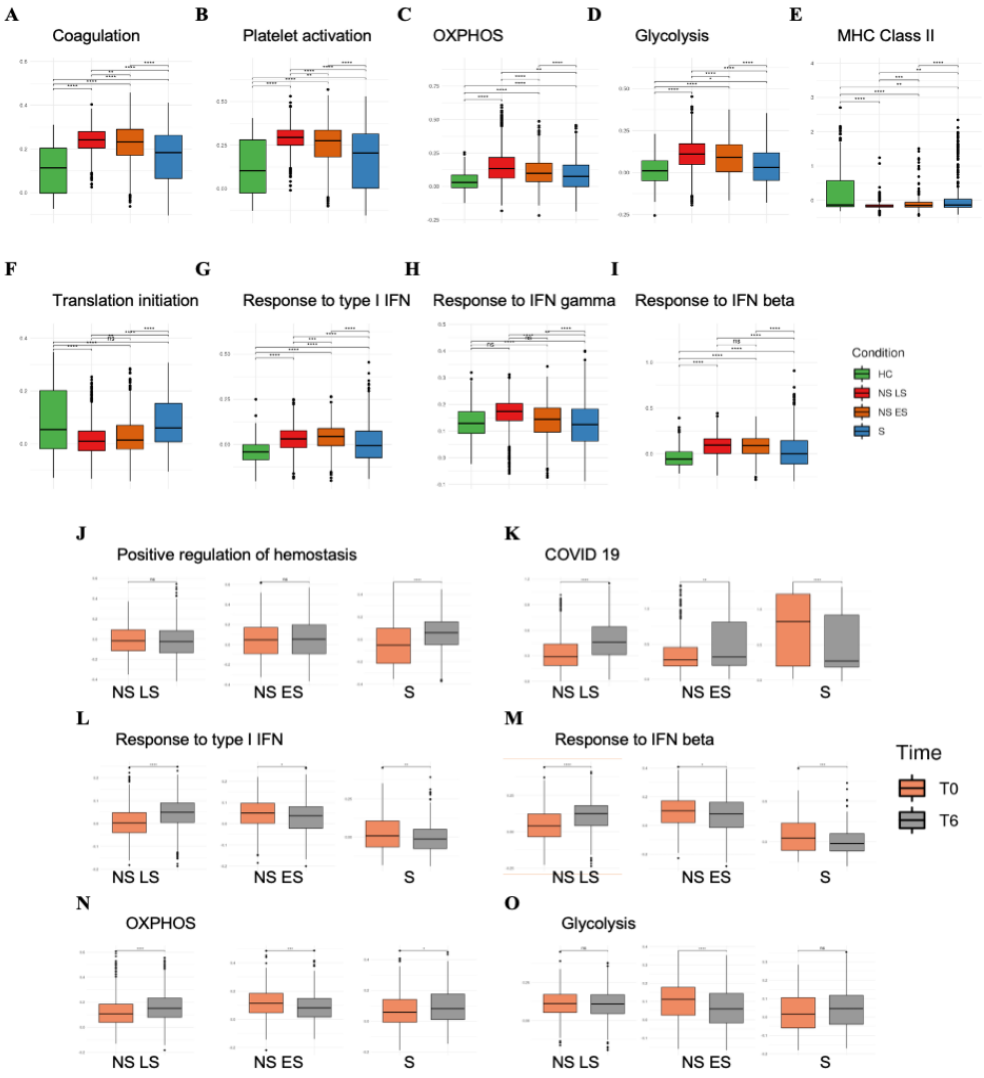

Figure 3.3 Platelet transcriptional changes over 6 h are associated with sepsis severity.

(A– I) Comparisons of pathway module scores across four conditions in platelets. The included modules contain genes related to (A) Coagulation, (B) Platelet activation, (C) OXPHOS, (D) Glycolysis, (E) MHC Class II, (F) Translation initiation, (G) Response to type I IFN, (H) Response to IFN gamma, (I) Response to IFN beta. (J–O) Pathway module scores comparison between T0 vs. T6 in platelets. The included modules contain genes related to (J) positive regulation of hemostasis, (K) COVID-19, (L) response to type I IFN, (M) response to IFN-β, (N) OXPHOS, and (O) glycolysis. The differences in scores associated with adjusted P-values below 0.05, 0.01, 0.001, and 0.0001 are indicated as *, **, ***, and ****, respectively and "ns" – not significant. The significance analysis was performed using Wilcoxon tests.

The COVID- 19 megakaryocyte (MK) cell trajectory study reported dysregulated IFN responses in MK cells from patients with severe COVID-19 severe patients, including increased metabolic activity of MKs along the disease trajectory (264). We investigated if IFN response modules changed within the 6-h timeframe and observed that the IFN responses, including type I IFN and IFN-β (Figure 3.3L and M), were significantly increased at T6 in NS LS but decreased at T6 in NS ES and S patients. Changes in metabolic activity within 6 h included significantly increased OXPHOS scores in NS LS at T6, while NS ES and S patients had significantly decreased OXPHOS scores (Figure 3.3N). The glycolysis score in NS ES was also significantly decreased at T6 (Figure 3.3O). Another COVID-19 study reported that circulating platelet-neutrophil, -monocyte, and T-cell aggregates were elevated in COVID-19 patients compared to healthy donors (265). We used the ligand and receptor database from iTalk (92) to score these interactions by calculating the product of average receptor expression and average ligand expression in the respective cell types (see Materials and Methods). Platelet-monocyte interaction scores were significantly elevated in NS LS (Figure A.2). The increased aggregation score to monocytes may explain the unexpected appearance of the platelets in the PBMC fraction. This aggregation was specific to monocytes, as platelet-neutrophil and -T-cell scores were decreased (Figure A.2).

Together, our study confirms the theory that platelet coagulation, activation, and energy consumption are functionally linked to sepsis disease severity and identifies shared pathways with COVID-19 disease progression. Further, our timed analysis reveals that these platelet responses are dynamic, changing within a 6-h window, especially in the late

stages of fatal sepsis. These data implicate platelet dysfunction as prognostic for disease progression in many infections and suggest that targeting these cell types may be important to prevent fatal outcomes.

3.3.4 Hypoxic stress is a driving factor for erythropoiesis in sepsis with shared pathways in COVID-19 infection

Based on the immune profiling results, which revealed the emergence of erythroid precursors especially in NS LS (Figure 3.2B and C), we investigated transcriptional changes in these cells. Only traces of these cell types are typically present following PBMC isolation by gradient centrifugation, therefore their high levels may suggest abnormal expansion and activation in the oxygen-limiting sepsis environment. Indeed, erythrocyte precursors can be generated through stress erythropoiesis (266) as a response to hypoxic conditions (267, 268). To test whether these cells were responding to hypoxia, we analyzed the average HIF1A expression in each cell. To avoid the drop-out events known as artifacts in single-cell studies, we removed cells with HIF1A "zero" expression, which left us with 32,185 cells. The HIF1A had the highest expression in NS LS, followed by NS ES, HC then S (Figure 3.4A). We investigated dynamic transcriptional changes in erythroid precursors by identifying differentially expressed genes (DEGs) in T0 versus T6. Only the

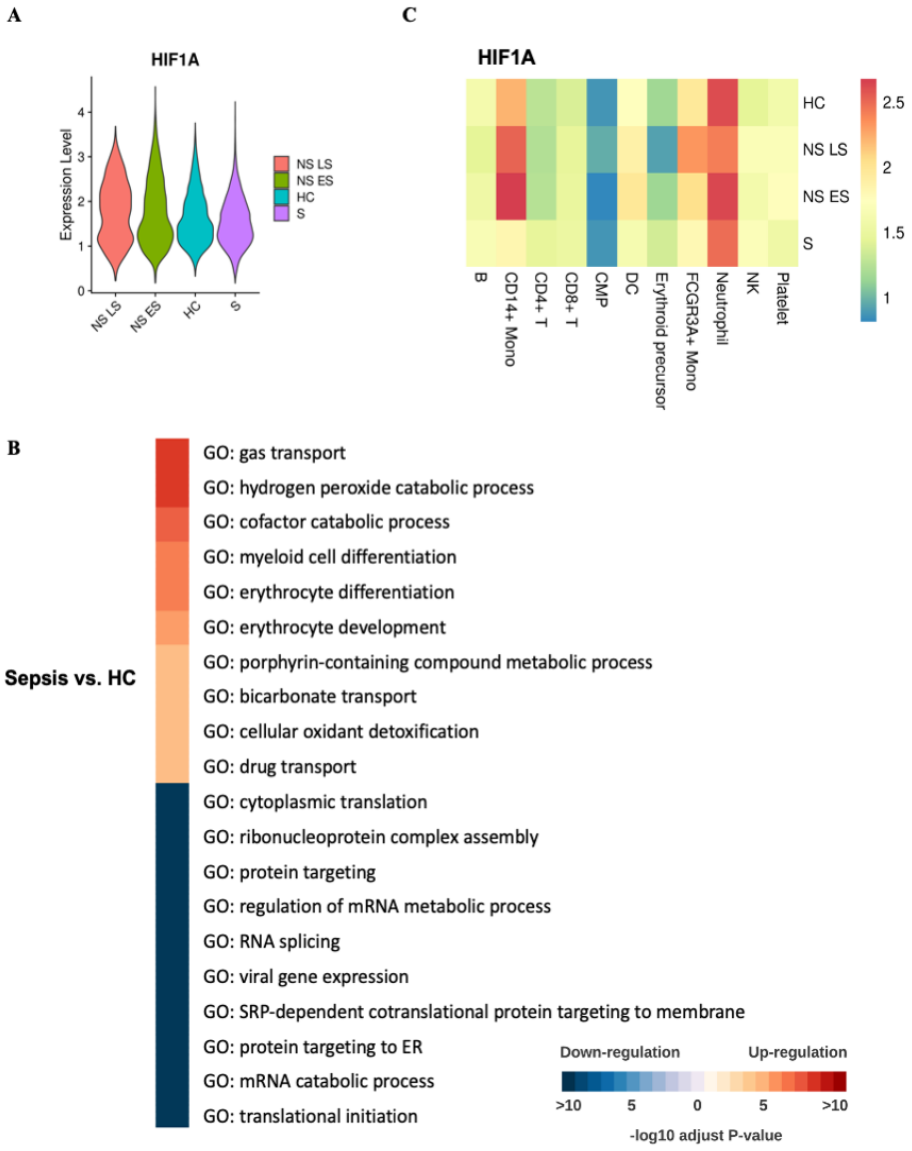

Figure 3.4 Elevated erythroid precursor cells are associated with hypoxic stress.

(A) The expression of the HIF1A gene in erythroid precursors across four conditions. Violin plots are ordered according to the decreasing average value of HIF1A expression. (B) Pathway enrichment when comparing erythroid precursors in sepsis vs. HC. All the GO terms are aligned to representative ones by Revigo (300) with a similarity of 0.4. The top 10 -log10 adjust P-values were selected shown in the heatmap. Color red are up-regulated pathways in sepsis patients. The color blue is downregulated pathways in sepsis patients. (C) The comparison of expression of HIF1A in the four conditions. Heatmap coloring represents log-normalized mean gene expression counts averaged across all cells.

NS LS patient exhibited statistically significant DEG, with the inflammatory protein S100A9 being significantly increased at T6 (adj. P-value < 0.001, log2 FC = 0.34).

We next examined gene expression differences in the erythroid precursor cells of sepsis patients and healthy controls (Figure 3.4B). Erythroid precursors in sepsis expressed genes related to hypoxic stress (hydrogen peroxide catabolic process, erythrocyte differentiation, cofactor catabolic process, and cellular oxidant detoxification). The down-regulated pathways in sepsis versus HC included cytoplasmic translation, ribonucleoprotein complex assembly, and RNA splicing, suggesting that erythroid precursors in sepsis underwent a halt in protein translation. Our results suggest a strong association between erythropoiesis and fatal sepsis outcomes. This association was also found in a study on COVID-19 infections (264), which proposed that erythroid cells are pivotal components of an unfavorable course of COVID-19. We also investigated if other immune cells were responding to hypoxia and found that monocytes from the sepsis nonsurvivors had the highest HIF1A expression compared to the other patients and other cell types (Figure 3.4C). Combined, our findings indicate that sepsis drives hypoxic stress that is associated with disease severity as well as dysfunctional erythropoiesis, which is a shared mechanism in many disease etiologies, including COVID-19 infection.

3.3.5 Monocyte transcriptional changes occur within hours of sepsis recognition and reflect immunosuppression

Monocytes are innate immune cells that sense and respond to pathogen invasion by producing inflammatory cytokines and mediating pathogen killing. However, a

dysregulated monocyte response can be damaging and fatal. Studies have found that in sepsis, monocytes may produce a flood of inflammatory cytokines triggering a "cytokine storm", causing widespread inflammation that can lead to a collapse in blood pressure, coagulation abnormalities, and ultimately organ failure and death. In the later stages, patients who survived the cytokine storm may die from immunosuppression, called "immune paralysis" in its extreme form (269). Moreover, the proinflammatory and immunosuppression stages might overlap (270). The stages at which the immune system transits from proinflammatory to suppressive at the cellular and molecular level have not been well studied, mostly because analyses are only performed at a single snapshot in time. To address this question, we analyzed samples from the same patients within 6 h of sepsis diagnosis, providing a picture of the timed trajectory of monocytes during this critical time window. Investigation of the GO Term cytokine activity (GO:0005125) indicated that most cytokines were higher in sepsis nonsurvivors (NS) compared to survivors (S) (Figure 3.5A; Figure A.3). However, when investigating changes between T0 and T6, we found that the cytokines were down-regulated in the NS patients from T0 →T6, especially in NS LS (Figure 3.5B and C; Figure A.3). The observation may indicate that the NS patients had already passed the proinflammatory stage and had instead begun immune shutdown. The proinflammatory cytokines up- regulated in NS compared to S monocytes, but down-regulated at time T6 included CCL2, CCL7, and NAMPT. The chemokines CCL2, CCL7 are vital for the recruitment of CC-chemokine receptor 2-positive ($CCR2^{+}$) monocytes (271). CCL2 and CCL7 expression were also enriched in the bronchoalveolar fluid from patients with severe COVID-19 (272). The only cytokine that was consistently up-

regulated in NS T0 versus T6 but down-regulated in S T0 versus T6 was NAMPT (Figure 3.5B-D; Figure A.3). NAMPT has been reported as a biomarker in sepsis and sepsis-induced acute respiratory distress syndrome (ARDS) in multiple studies (273-276). The NAMPT/TLR4 inflammatory pathway has also been studied as the COVID-19-induced ARDS drug target (277). The cytokines down-regulated over time in NS monocytes, but up-regulated in S monocytes at T6, included TNFSF10/TRAIL, an immunoregulator that mediates leukocyte apoptosis which had been reported to enhance survival in murine polymicrobial sepsis (278, 279), and TNFSF13B/BAFF, a stimulatory factor for B cells (280, 281) (Figure 3.5D; Figure A.3). These data suggest that fatal sepsis is associated with mixed effects on lymphocyte responses, which may be mediated by monocytes.

To profile metabolic changes in monocytes, we investigated the genes that belong to OXPHOS and glycolysis modules, and examined correlation with HIF1A, which was the most highly expressed in monocytes (Figure 3.4C). The OXPHOS modules from all conditions were negatively correlated with HIF1A (Figure 3.5E; Figure A.3). On the other hand, the glycolysis modules were positively correlated with HIF1A (Figure 3.5F; Figure A.3). Within the group of sepsis survivors, the metabolic activity in $CD14^+$ and $FCGR3A^+$ monocytes was dominated by OXPHOS at both time points. Moreover, the $FCGR3A^+$ monocytes in S had more OXPHOS energy consumption at T6. Overall, monocytes from the sepsis nonsurvivors had shifted energy consumption from OXPHOS to aerobic glycolysis, potentially indicating host defense activation such as production of reactive oxygen species (Figure 3.5G). However, within the 6 h timeframe, both $CD14^+$ and $FCGR3A^+$ monocytes from sepsis nonsurvivors exhibited a drop in their glycolysis module

scores (Figure 3.5G; Figure A.3). One group had demonstrated that defects in the energy metabolism of leukocytes underlie immune paralysis in sepsis, and restoring the ability of immunotolerant leukocytes to mount a glycolytic response might represent a promising novel therapeutic approach to revert the immunotolerant state of sepsis (282). The energy shift to glycolysis, and the decrease in glycolysis consumption at T6 in NS suggest that these monocytes were undergoing immune suppression at the point of sepsis recognition. In contrast, the glycolysis score was increased at T6 in sepsis survivors. We also observed that the HLA-DR module score, as an indicator of monocyte antigen-presenting function (283), was decreased at T6 in the NS LS, indicating immune suppression. However, in the S and NS ES monocytes, the HLA-DR expression was increased at T6, especially in S (Figure 3.5H; Figure A.3). Together, these data reveal dynamic transcriptional changes in monocytes within 6 h of sepsis diagnosis, which follow opposite trends in surviving and fatal outcomes. Fatal sepsis is associated with heightened inflammatory and metabolic activity that is down-regulated over time, while improved sepsis outcomes are associated with the restoration of monocyte function within 6h.

3.3.6 CD52 is a prognostic biomarker for beneficial outcomes in sepsis and is associated with lymphocyte activation

Our preliminary analysis indicated severe lymphopenia in sepsis, especially in the NS LS (see Figure 3.2). We further investigated the transcriptional profile of the lymphocytes in sepsis patients and whether it changed over 6 h. Evaluation of activation module scores for $CD4^+$, $CD8^+$ T cell, and B cells (GO terms T cell activation, GO:0042110; B cell activation,

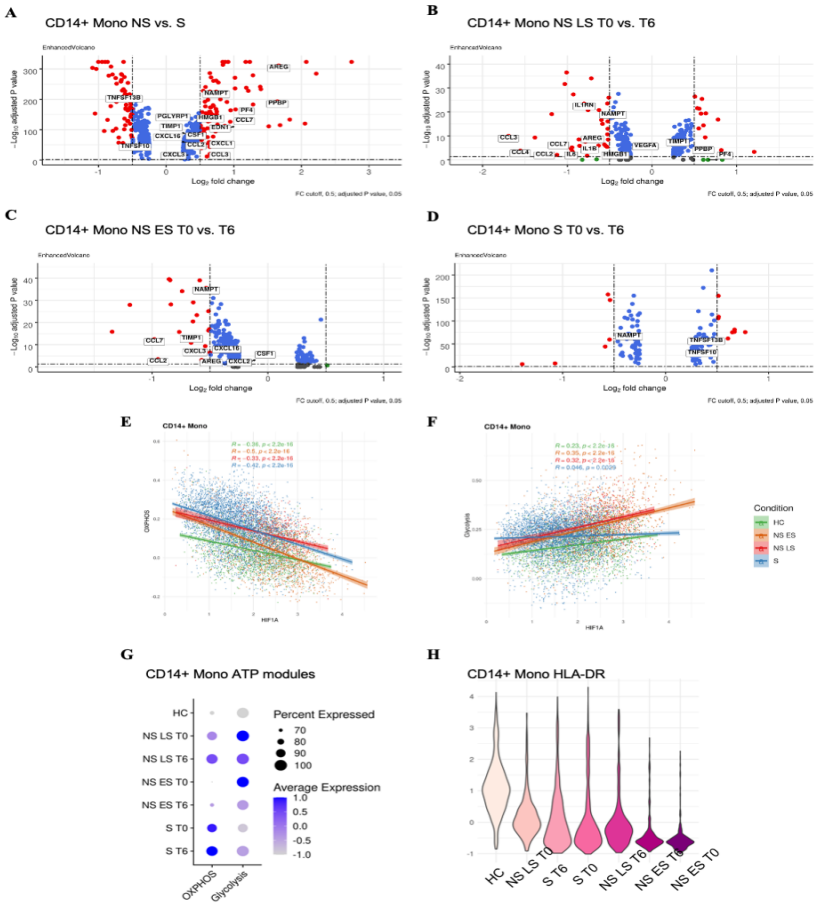

Figure 3.5 Fatal sepsis patients exhibit immunosuppressive pathways in monocytes.

(A-D) Differential expression genes in CD14+ monocytes from (A) NS versus S, (B) NS LS T0 versus T6, (C) NS ES T0 versus T6, (D) S T0 versus T6. Volcano plots were prepared with R package EnhancedVolcano (301). (E and F) The correlations between the HIF1A expression and module score for (E) OXPHOS and (F) glycolysis in CD14+ monocytes across each condition. R-values from Pearson's correlation, exact 2-sided P-values, and the 95% confidence intervals are shown on each graph. Each dot represents a single cell. Only cells with HIF1A expression ≠ 0 were included in the analysis. Green, orange, red and blue points represent cells from HC, NS ES, NS LS, and S samples, respectively. (G) The percentage of cells with ATP-related pathway modules in CD14+ monocytes across healthy controls and sepsis conditions at T0 and T6. The color saturation indicates the average expression level, and the circle's size indicates the percentage of cells expressing a given module. (H) HLA-DR-related genes expression in CD14+ monocytes across healthy controls and sepsis conditions at T0 and T6. Violin plots are ordered with the decreasing expression average value of HLA-DR-associated genes. The color saturation indicates the average expression level, the darker the color, the lower the average expression level.

GO:0042113) indicated increased activation over 6 h in surviving sepsis patients (Figure 3.6A; Figure A.4). In contrast, NS ES had significantly decreased activation scores over time in all lymphocytes (Figure 3.6A; Figure A.4). NS LS also exhibited activation scores that decreased significantly, but only in $CD4^+$ T cells, mainly due to the severe CD8+ T cell and B cell lymphopenia Figure 3.6C; Figure A.4).

We next applied DEG analyses between T0 versus T6 in all patients' lymphocytes (Figure 3.6D–F, only the genes with $|\log2FC| > 0.25$ and adjusted P-value < 0.05 are shown). To our knowledge, no studies have investigated transcriptional response changes of individual cell types within hours in sepsis patients; however, a study utilizing cecal ligation and puncture as a mouse model for sepsis reported reduced CD4+ T cell activation in the spleen after 6 h, consistent with our current data.45 We therefore investigated the transcripts that were reported in this mouse study: CCR2, CCR6, CD3, CD48, CD52, CD80, ITGB7, SELL, SLAMF6, and Thy1. Of all these genes, only some of them were significantly changed between T0 and T6 in sepsis (Figure 3.6D-F). Our analysis identified that CD52, a surface glycoprotein involved in lymphocyte activation, was the most relevant biomarker to predict lymphocyte status and disease outcome. CD52 expression was increased over 6 h in the T and B cells of survivors, but not in sepsis nonsurvivors (Figure 3.6D-F). We investigated the other markers, however, did not observe consistent trends associated with protection. To validate that CD52 is correlated with improved lymphocyte function, we plotted CD52 expression against GO term T cell activation and B cell activation module scores. In $CD4^+$ T cells, we observed significant positive correlations in

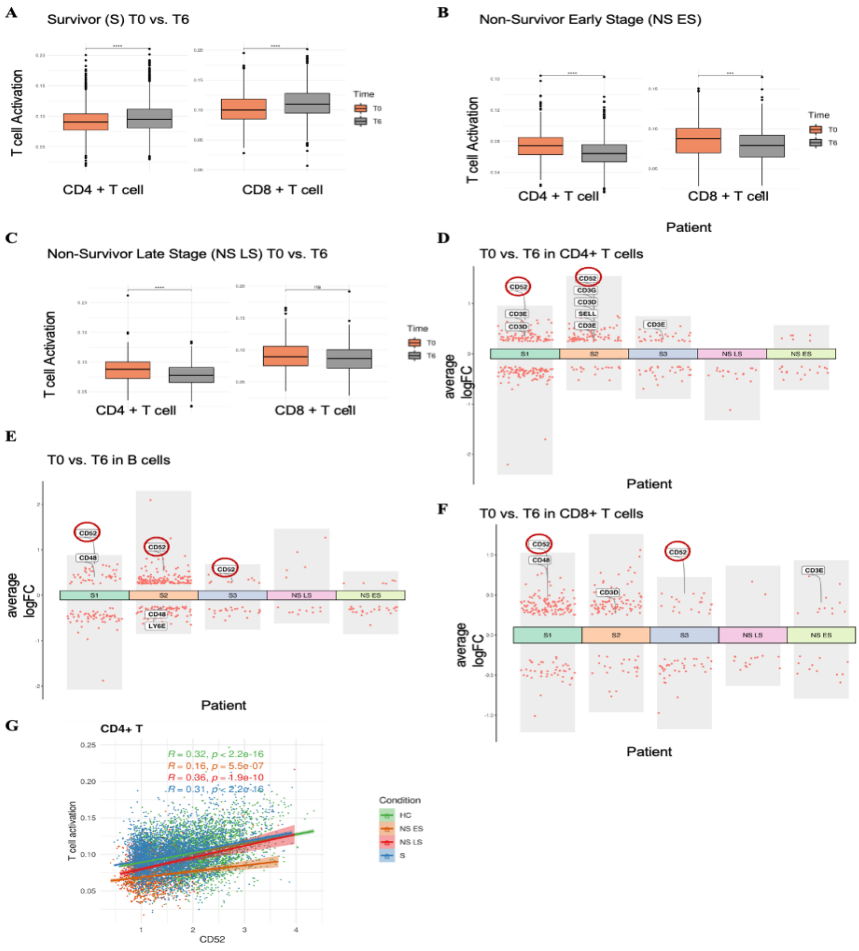

Figure 3.6 CD52 expression correlates with lymphocyte activation.

(A-C) T cell activation pathway module score comparison between T0 and T6 in T cells. (A) Survivors (S), (B) Nonsurvivor early stage (NS ES), and (C) Nonsurvivor late stage (NS LS). The differences in scores associated with adjusted P-values below 0.05, 0.01, 0.001, and 0.0001 are indicated as *, **, ***, and ****, respectively. The significance analysis was performed using Wilcoxon tests. (D-F) Differential gene expression analysis showing up- and down-regulated genes with |log2FC| > 0.25 and adjusted P-value < 0.05 across all 5 sepsis patients between T0 and T6 in (D) CD4+ T cells, (E) B cells, (F) CD8+ T cells. (G) CD52 expression and its correlation with the T cell activation pathway module score in CD4+ T cells across four conditions. R-values from Pearson's correlation, exact 2-sided P-values, and the 95% confidence intervals are shown on each graph. Each dot represents a single cell. Only cells with CD52 expression ≠0 were included in the analysis. Green, orange. red and blue points represent cells from HC, NS ES, NS LS, and S samples, respectively.

all conditions (Figure 3.6G), while in B cells, there were significant positive correlations in the HC, NS LS, and S, but not in the NS ES patient (Figure A.4). Together, our data indicate that increased CD52 expression within hours of sepsis recognition is associated with improved sepsis outcomes. Here, CD52 may act to promote restoration of protective lymphocyte responses, and therefore may serve both as a biomarker for sepsis progression, or as a therapeutic target to promote immune homeostasis.

3.4 Discussion

Sepsis is a dysregulated systemic inflammatory response, which results in organ injury with mortality rates of 15–25% (249, 284). The molecular level heterogeneity of sepsis makes the study of the dynamics of the individual cell types the ideal tool for understanding sepsis progress and response. However, from over 1000 single-cell transcriptomics studies that have been published to date (285), only 3 have studied sepsis (244, 286, 287). These studies focused specifically on only two groups of cells: monocytes and myeloid-derived suppressor cells. In contrast, our study used the centrifuge gradient-based approach to isolate PBMC before performing single-cell RNA-seq, which expanded the cell subsets investigated. We additionally collected samples at two different time points from differing outcomes in sepsis, which provided temporal details of the immune response in severe sepsis. These focused analyses were able to identify specific immune cell subsets and gene expression patterns over time that correlated with beneficial or fatal outcomes. Our results are consistent with the previous studies both in single-cell and bulk sepsis transcriptomic studies, but also bring details not seen in other studies, notably molecular changes that

occur within hours. Based on this data, future studies evaluating additional time points (such as 24 h and post-discharge) would be relevant to investigate whether the first 6 h are predictive of recovery in surviving sepsis patients. We report in this study that the peripheral blood cell composition of nonsurvivors is more "distant" from healthy controls than the blood cells of survivors, and that severe lymphopenia occurs in fatal sepsis. We also explored cell types that were not previously investigated in sepsis single-cell studies, such as platelets and erythroid precursors, and observed distinct changes in monocytes and lymphocytes within 6 h. Neutrophils are a dominant cell subset in the blood with established antimicrobial but also inflammatory roles in sepsis (288, 289); however, we were unable to investigate this subset given that most neutrophils are not recovered in the PBMC fraction, and technical issues exist with scRNA-seq of this cell type.

We found that platelets were expanded in sepsis patients, especially in fatal outcomes. Examination of transcriptional changes over time in platelets from the nonsurvivor sepsis patient revealed increased expression of genes related to coagulation, platelet activation, and ATP production modules, including OXPHOS genes and glycolysis genes. These changes were also reported in a COVID-19 MK study (264), suggesting that platelet dysfunction is a shared feature of both diseases, and indicative of clinical severity. This study identified increased metabolic activity of MKs compared to healthy controls. Another study from Holmsen et al.(290) demonstrated a correlation between platelet energy demand and aggregation. Consistent with this, we found that the platelets in nonsurvivor sepsis patients had dramatically reduced translation initiation pathways along with the induction of the IFN response pathways, suggesting a general suppression of the

protein synthetic apparatus by IFN (263). We also observed that the transcriptional changes in platelets in fatal sepsis were similar to severe COVID- 19 patients, including lasting IFN responses, increased metabolic activity, and elevated circulating platelet-monocyte aggregates. Several studies focus on using antiplatelet agents such as aspirin with sepsis (291, 292), and a recent COVID-19 study demonstrated that aspirin prescription was associated with decreased mortality rates for COVID- 19 positive patients enrolled at the Veterans Health Administration (293). Coagulation disturbances (bleeding and/or clotting) are prominent clinical concerns in sepsis and COVID-19 and deserve further inquiry. Our results support the potential for antiplatelet therapies for the treatment of severe sepsis.

Further investigation of the erythroid precursor subset that was expanded in fatal sepsis revealed the upregulation of genes related to hypoxic stress and apoptosis, reflective of the hypoxic environment in severe sepsis that leads to emergency erythropoiesis. Interestingly, in a longitudinal COVID-19 study (264), erythroid cells were also identified as a hallmark of severe disease with defined molecular signatures linked to a fatal COVID-19 disease outcome. We also observed that erythrocyte expansion and expression of genes related to hypoxic stress were significant predictors of fatal outcomes. Within the erythroid precursor subset, we identified that inflammatory alarmin S100A9 expression dynamically changed in fatal sepsis, with significant increases at T6. S100A9, together with and S100A8 and S100A12, were previously reported as biomarkers for higher risk of death in septic shock patients (294). S100A9 was also identified in a human bone marrow erythropoiesis study (295), which reported its up-regulation at the last stage of maturation of nucleated red blood cell precursors. Together with these studies, our data suggests that

S100A9 expression reflects stress erythropoiesis and is associated with rapid fatality in sepsis. S100A9 may serve as a valuable biomarker to stratify sepsis severity, in addition to the clinical scores and other plasma markers.

Consistent with previous studies showing that monocytes play a significant role in the sepsis pathogenesis (269, 270), we observed aberrant gene expression and pathway changes in the monocytes of sepsis patients. Overall transcriptional profiles indicated that monocytes were in a hyperinflammatory state in sepsis nonsurvivors. However, focused analyses within the 6-h time window revealed that monocytes from NS patients were undergoing immune suppression, including decreased pro-inflammatory cytokine and HLA-DR expression, and reduced glycolysis energy consumption at T6. Within the 6-h time window, we also identify CD52 as the biomarker for B and T cell activation that correlates with beneficial outcomes. CD52 is a glycoprotein expressed on the surface of mature lymphocytes, monocytes, dendritic cells, and NK cells (296), therefore surface expression could be quantified to predict sepsis progression. Most CD52 targeted therapeutic approaches aim to delete CD52-expressing cells, such as the monoclonal antibody alemtuzumab, which treats chronic lymphocytic leukemia and multiple sclerosis (297). However, our study suggests that promoting CD52 signaling may be beneficial to improving lymphocyte function. In fact, in a study using alemtuzumab, one patient with aggressive multiple sclerosis developed sepsis after treatment (298). Additionally, CD52 had been proposed as a prognostic biomarker in breast cancer, where it is correlated with improved outcomes likely due to increased immune tumoricidal activity (299). CD52 might therefore serve as a biomarker for sepsis prognosis and provides a new therapeutic

target for sepsis patients, however, determining its influence on sepsis therapy would require an expanded sample size study.

In conclusion, results from this study indicate that the initial status of the sepsis patient and the dynamic changes in cell behavior during the critical period following diagnosis significantly affect sepsis outcome. Therapeutic intervention to modify these immune trajectories may therefore lead to improved outcomes in these patients that could be identified by biomarkers reported in this study, such as CD52 or S100A9. Future studies that focus on these dysfunctional cell subsets at the individual level, addressing their metabolic dysfunction or how to promote their recovery from exhaustion, may provide therapeutic and prognostic strategies for sepsis, which could be applicable to other fatal diseases such as COVID-19.

3.5 Acknowledgments

In this chapter, I acknowledge the collaborative efforts of a group of esteemed professionals in the field of sepsis research. My co-authors on this work include Dr. Jiang Li, Dr. Jeff Bonenfant, Dr. Lukasz Jaroszewski, Dr. Aarti Mittal, Dr. Walter Klein, Dr. Meera G. Nair and Dr. Adam Godzik, who all contributed in various ways to our study on dynamic changes in human single-cell transcriptional signatures during fatal sepsis (99).

I am grateful to Dr. Meera G. Nair, Dr. Adam Godzik, Dr. Jiang Li, Dr. Jeff Bonenfant, Dr. Aarti Mittal, and Dr. Walter Klein for their role in conceptualizing the study. The methodology was developed with the assistance of Dr. Jiang Li, Dr. Aarti Mittal, Dr.

Meera G. Nair, and Dr. Adam Godzik, while the investigation itself was carried out with the help of Dr. Meera G. Nair, Dr. Adam Godzik, Dr. Jiang Li, and Dr. Jeff Bonenfant.

Dr. Meera G. Nair, Dr. Adam Godzik, Dr. Jiang Li, and Dr. Lukasz Jaroszewski provided invaluable support in providing the formal analysis, and Dr. Walter Klein, Dr. Jeff Bonenfant, and Dr. Aarti Mittal contributed their clinical expertise to the study. Finally, I want to express my gratitude to Dr. Meera G. Nair and Dr. Adam Godzik for their outstanding leadership in supervising the study.

CHAPTER 4 Platelets in Immune Dysregulation: Identifying Novel Platelet Subtypes in Severe and Fatal Cases of COVID-19 and Sepsis

4.1 Introduction

Until recently, platelets were regarded as simplistic cells with a singular purpose of facilitating hemostasis, the cessation of blood flow at the site of disrupted endothelial lining, through the formation of blood clots. However, in the last few years, platelets emerged as critical elements of the immune system, being first responders in infections and regulating both the activation as well as post-infection deceleration of the host immunity (302-304). Platelets have been proven to modulate the function of immune cells by physically adhering to them or by releasing cytokines and chemokines that regulate them. Platelets affect the activation (305, 306), proliferation (307), differentiation (308, 309), pathogen clearance (310), and cytokine response of other immune cells (311, 312). Additionally, abnormal platelet status was found to be an important component of many diseases that involve dysregulation of the immune system (99, 313-315).

Thrombosis and abnormal coagulation associated with the coronavirus disease 2019 (COVID-19) brought more attention to the role of platelets in immune system dysfunction caused by infections. A higher percentage of platelets purified with the Peripheral Blood Mononuclear Cell (PBMC) fraction was shown to be an indicator of both COVID-19 (316) and sepsis (317) severity. The prothrombotic profile that is now a hallmark of COVID-19 is believed to be mediated in part by platelet activation and aggregation (313). In addition to COVID-19 and sepsis pathogenesis, platelets are

implicated in autoimmune diseases such as systemic lupus erythematosus (SLE) (318); therefore, we hypothesize that targeting platelets or some subsets thereof may be a potential strategy in the treatment of these diseases and possibly even a broader range of immune disorders. To evaluate this hypothesis, we undertook a deeper analysis of the available single-cell RNA-seq datasets from these three diseases, building upon the results of our previous analysis of sepsis patients (99).

The question of how to prevent healthy platelets from transforming into abnormal ones is difficult to answer with bulk transcriptomics, as this doesn't allow for a detailed study of the changes that occur specifically within the platelet subsets in these diseases. Single-cell RNA seq can potentially answer such questions, therefore we collected datasets from over four hundred samples, spanning a range of diseases, including sepsis, COVID-19, and Systemic Lupus Erythematosus (SLE). And disease seveverities, from mild to severe and fatal. Using single-cell level analysis, we can identify and characterize subpopulations of abnormal platelets, study molecular details of their pathogenic changes, and develop testable hypotheses for how they contribute to patient death or survival.

4.2 Materials and Methods

4.2.1 Integrated Single-cell Transcriptome Atlas of Peripheral Blood Mononuclear Cells (PBMCs) from COVID-19, Sepsis, and Systemic Lupus Erythematosus Patients

We gathered single-cell RNA-seq datasets of peripheral blood mononuclear cells (PBMCs) from COVID-19 (202, 203, 264, 319-324), sepsis (99, 286), and Systemic Lupus Erythematosus (SLE) (325) patients, for a total of 413 samples, in order to study and

compare the immunological dysregulation caused by these diseases. In the COVID-19 studies, individuals with severe influenza, lung infections but testing negative for COVID-19 were identified as the "hospitalized patients with similar symptoms" (SSH) group and we included them in our analysis. We categorized the samples into six groups based on the severity of the disease: healthy control (HC), convalescence (CV), mild (ML), moderate (MD), severe (SV), fatal (FT) and created a separate group for SLE. For the outcome analyses we grouped patients who survived from the CV, ML, MD and SV groups into a survivor (S) group, with the patients with unannotated outcomes forming another group labeled as the unknown group (Figure 4.1A).

From the 413 samples, a total of 47,977 individual platelet datasets were extracted. Classifying the platelets by the disease severity group, there were 3,205 cells from HC, 3,695 cells from CV, 7,359 cells from ML, 4,330 cells from MD, 19,805 cells from SV, 9,414 cells from FT, and 169 cells from SLE. By the outcome group, there were 3,205 from HC, 9,414 from FT, 25,750 from S, and 9,608 from unknown outcomes. Overall, there are 38,673 platelet datasets from COVID-19, 2,508 from SSH, 3,422 from sepsis, and 169 from SLE.

4.2.2 Integrating the Datasets and Identifying the Cell Types

Expression data for platelets was extracted from all the datasets identified in the previous step. For the datasets with the assigned cell type, we used the existing annotations. We used SingleR (326) and Seurat v4 (77) to identify platelets in the datasets that lacked cell type annotations. Using HGNChelper (327), we harmonized the nomenclature of the genes from

different datasets and identified the 3000 most variable genes from each. For the integrated analysis, the top 3000 highly variable genes were retained. Harmony (65) was executed with PCA embeddings (30 PCs) as input and the default parameters to eliminate batch effects among twelve datasets (nine COVID-19 datasets, two sepsis datasets, and one SLE dataset). Seurat was then applied to the Harmony embeddings to determine the clusters. In the round of clustering, the resolution for Louvain clustering was set to 0.4.

4.2.3 Differential Expression Analysis

The MAST method (253) from the Seurat package (implemented in the FindAllMarkers function) was used with the default parameters to perform the differential gene expression analysis and identified the DEGs. We used patients' categories, as described before, to compare DEGs between disease severity states.

4.2.4 Comparison of Module Scores

We measured the extent to which individual cells expressed certain predefined expression gene sets using cell module scores. All calculations and comparisons of module scores were performed using the AddModuleScore function from the Seurat package with the default parameters. We compared the expression of modules extracted from the GO.db database (328). Additionally, we used literature-based modules for oxidative phosphorylation (OXPHOS), glycolysis (260), and MHC class II. We disregarded genes without detectable expression in our data.

4.2.5 Pathway Enrichment Analysis

We analyzed Gene Ontology biological process terms (GO-BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using R package clusterProfiler (88) to identify the genes up- and down-regulated in disease samples as compared to HC. The "common up" and "common down" gene sets at the disease level were identified by intersecting the upregulated or downregulated DEGs for COVID-19, sepsis, and SLE compared to HC. We performed Medical Subject Headings (MeSH) (329) enrichment using the R package DOSE (330) for each cluster's upregulated genes. We calculated scores for the Molecular Signatures Database (MSigDB) signature gene set collection (331) using the R package GSVA (85). To visualize the upregulated genes and enriched diseases associated with the platelets subtypes over-represented genes, we also employed the Gene, Disease Features Ontology-based Overview System (gendoo) (332).

4.2.6 Calculation of Ligand-Receptor Interaction Scores Between Platelets and Other Cell Types

Cell-cell ligand-receptor interactions were evaluated using the scoring system proposed by Kumar et. al. (89). The score uses ligand-receptor interactions from iTALK ligand-receptor database (92). The ligand-receptor interaction score between cell type A and cell type B is calculated as a product of average receptors' expression across all cells of cell type A and the average ligands' expression across all cells of cell type B.

$$Interaction\ score(receptor, ligand, cell\ type\ A, cell\ type\ B) = \frac{1}{n_{cell\ type\ A}} \sum_{i \in cell\ type\ A} e_{i,receptor} \times \frac{1}{n_{cell\ type\ B}} \sum_{j \in cell\ type\ B} e_{j,ligand}$$

$e_{i,j}$ = expression gene j in cell i

$n_{celltypeA}$ = number of cells in the cell type A

4.2.7 Trajectory Inference of Transition in Platelets Subclusters

We presented the platelets trajectory analysis using monocle3 (333). Trajectories were calculated and the cells displayed based on monocle3 pseudotime approach rooted in the previously identified transitional platelets. Integrated gene expression matrices from each dataset were first exported from Seurat into monocle3 to construct a new_cell_data_set. The exported matrices were aligned using the single-cell batch correction method package batchelor (67). Then, they were subjected to the standard PCA to preprocess the data with the number of dimensions set at 100. The dimension reduction and clustering of cells were all set to monocle3 default settings. Finally, we used reversed graph embedding to learn the principal graph from the reduced dimension space using the learn_graph function.

4.3 Results

4.3.1 PBMC Composition Changes with Patient Disease Severity and Outcome.

First, we analyzed the relative populations (fractions) of immune cell subsets in peripheral blood mononuclear cells (PBMCs) in relation to disease severity, outcome, and disease types (Figure 4.1A). The population proportions of platelets, precursor cells, and

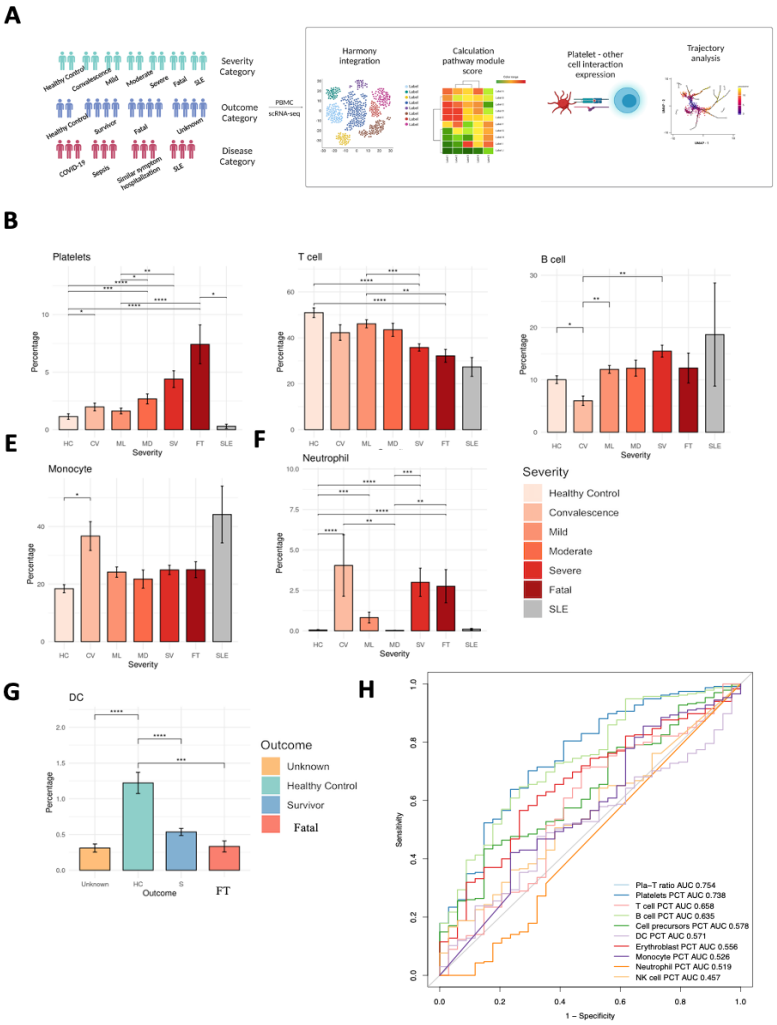

Figure 4.7 PBMC profiling from healthy controls, sepsis, similar symptom hospitalized, COVID-19 and SLE patients.

(A) Schematic diagram illustrating the process for data acquisition from published literature and the subsequent integrated analysis. Generated using biorender.com. **(B-F)** Bar charts displaying the proportions of various cell types under different disease severity conditions. B) Platelets, C) T cells, D) B cells, E) Monocytes, F) Neutrophils. **(G)** Dendritic cells (DCs) in distinct outcome scenarios. Statistically significant differences in percentages with adjusted P-values below 0.05, 0.01, 0.001, and 0.0001 are denoted as *, **, ***, and ****, respectively; non-significant results are not indicated. The statistical analysis was conducted using Wilcoxon tests. Standard error bars are included. **(H)** Receiver operating characteristic (ROC) curves for the platelet-to-T cell ratio and other cell type percentages were employed to differentiate non-survivors from survivors.

erythroblasts in PBMCs increased with the disease and outcome severity (Figure 4.1B, Figure A.5), while the T cell fraction decreased (Figure 4.1B). B cell fractions were significantly increased in SLE patients, which is a known phenotype in SLE patients (Figure 4.1C); monocyte and neutrophil fractions were significantly increased in convalescence patients (Figure 4.1E, F); however, fractions of the dendritic cells decreased as outcome severity increased (Figure 4.1G). We evaluated ratios of different cell types as a criterion for separating fatal (FT) and surviving (S) patients and found that the ratio of platelets percentage to T cells percentage, (Pla-T ratio) had the highest area under the curve (AUC) at 0.754, and platelet percentage (Platelets PCT) was closely overlapping with Pla-T ratio at 0.738, with the 0.063 ratio between FT and S patients (Figure 4.1H). Together, these results identify platelets as an important single criterion that increases with disease severity, while circulating T cells follow the opposite trend with decreasing frequency in severe disease.

4.3.2 Platelets Amplify Endotheliopathy and Lead to Potential Increase in Disseminated Intravascular Coagulation in Fatal Patients

Platelets are critically involved in the development and progression of endotheliopathy and disseminated intravascular coagulation (DIC) (334, 335). Platelets can bind to and activate endothelial cells, releasing pro-inflammatory and pro-coagulant molecules that can contribute to the development of DIC. Additionally, platelets can also contribute to thrombus formation and further damage to the endothelium (336).

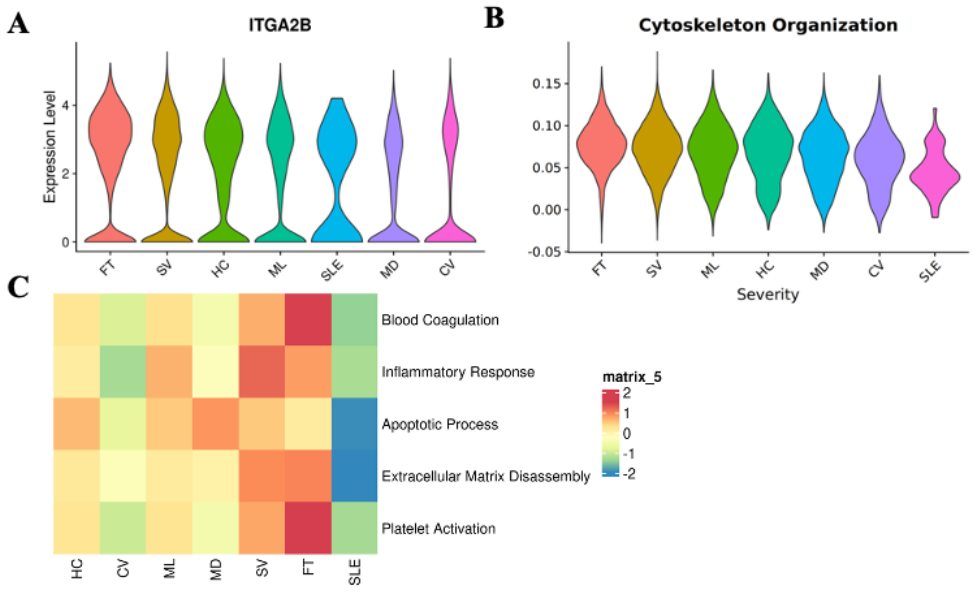

Figure 4.8 Differential expression of platelets affecting endotheliopathy across disease severity states.

(A) Expression of the ITGA2B gene in platelets across various severity states. **(B)** Expression of the Gene Ontology (GO) module related to cytoskeleton organization (GO:0007010) in platelets across different severity states. Violin plots are arranged in descending order based on the mean expression values. **(C)** Comparison of expression levels for GO terms, including blood coagulation (GO:0007596), inflammatory response (GO:0006954), apoptotic process (GO:0006915), extracellular matrix disassembly (GO:0022617), and platelet activation (GO:0030168). The heatmap coloring denotes z-scored values averaged across all cells from a specific sample.

We examined platelet expression of the Integrin Subunit Alpha 2b (ITGA2B) gene, which is involved in platelet-endothelial cell interactions by binding to the Integrin Subunit Alpha V (ITGAV) (337). The patients with fatal outcomes had the highest levels of ITGA2B expression, followed by patients in the severe disease group (Figure 4.2A). On the pathway level, the reorganization of actin cytoskeleton (GO:0007010), also followed

the same severity trend (Figure 4.2B). DIC is a complex condition characterized by abnormal clotting and bleeding due to the activation of the coagulation cascade and the depletion of clotting factors and platelets (338). Several other GO terms associated with DIC, such as blood coagulation (GO:0007596), inflammatory response (GO:0006954), extracellular matrix disassembly (GO:0022617), and platelet activation (GO:0030168) expression had highest expression in FT/SV patients. And all modules had lower expression in CV and SLE patients (Figure 4.2C).

4.3.3 Platelet Subpopulations Associated with Disease Severity

In the conventional evaluation of single-cell RNA sequencing (scRNA-seq), the primary objective is to pinpoint clusters that correspond to established cell types. In this study, we extended the clustering process to delve deeper into the intricacies of platelet population alterations in disease states. Through an iterative process of trial and error, we ascertained similarity thresholds that yielded stable clusters, exhibiting optimal separation and minimal overlap in signature pathways. Consequently, we identified thirteen distinct clusters, denoted as C0 through C12 (Figure 4.3A, B). We analyzed the composition of each cluster with respect to the contributions from various outcome groups. A strong association was observed between the C11 cluster and the FT group, as 78% of all C11 cells were identified within this group. Clusters C3, C5, and C9 exhibited the highest proportion of HC at 17%, while clusters C6 and C11 displayed the lowest at 1%. Cluster C3 demonstrated the highest proportion of the mild group, accounting for 25%. In clusters C6, C7, C10, and C12, over half of the total platelets were comprised of the severe group (refer to Table B.3). Clusters

C6, C8, and C10 had more than 70% of platelets originating from survivor samples (see Table B.4). Based on these findings, we categorize clusters C4, C9, and C11 as "fatal," C8 as "convalescent," and C6 and C10 as "survival" (Figure 4.3C, D). We subsequently refer to the clusters identified in this study as "subpopulations."

We used DEG and Gene Set Enrichment Analysis (GSEA) to identify molecular factors and pathways that could differentiate the platelet clusters (subpopulations) identified in the previous step. The sets of genes enriched in the fatal platelet clusters C4 and C11 were significantly different than the other platelet clusters (Figure 4.4A) and in some cases show opposite trends from the convalescent and survival clusters. The fatal cluster C4, whose signature module comprises HPSE, WFDC1, and PF4 genes, and the convalescence cluster C8, whose signature module consists of RPLP0, RPS6, and RPS23, have several pathways that trended in the opposite direction. Fatal cluster C11, whose signature module includes TPT1, POLR2L, and CSRP1, had the highest energy consumption pathway scores, including oxidative phosphorylation (OXPHOS) and glycolysis (Figure A.6), but the lowest inflammatory response score (Figure A.6). The other fatal cluster C9, whose signature modules include HBB, HBA2, and HBA1, showed the weakest interferon response, including alpha and gamma interferons (Figure A.6) Coagulation, epithelial-mesenchymal transition (EMT), and the apical junction are all at their highest in C4 but at their lowest in C8 (Figure 4.4B-D). At the same time, MYC targets v1 and v2, which contain nuclear-encoded genes involved in mitochondrial biogenesis (339), are the lowest in C4 but highest in C8 (Figure A.6). Based on these findings, we can conclude that platelets from the fatal cluster C4 are highly active in

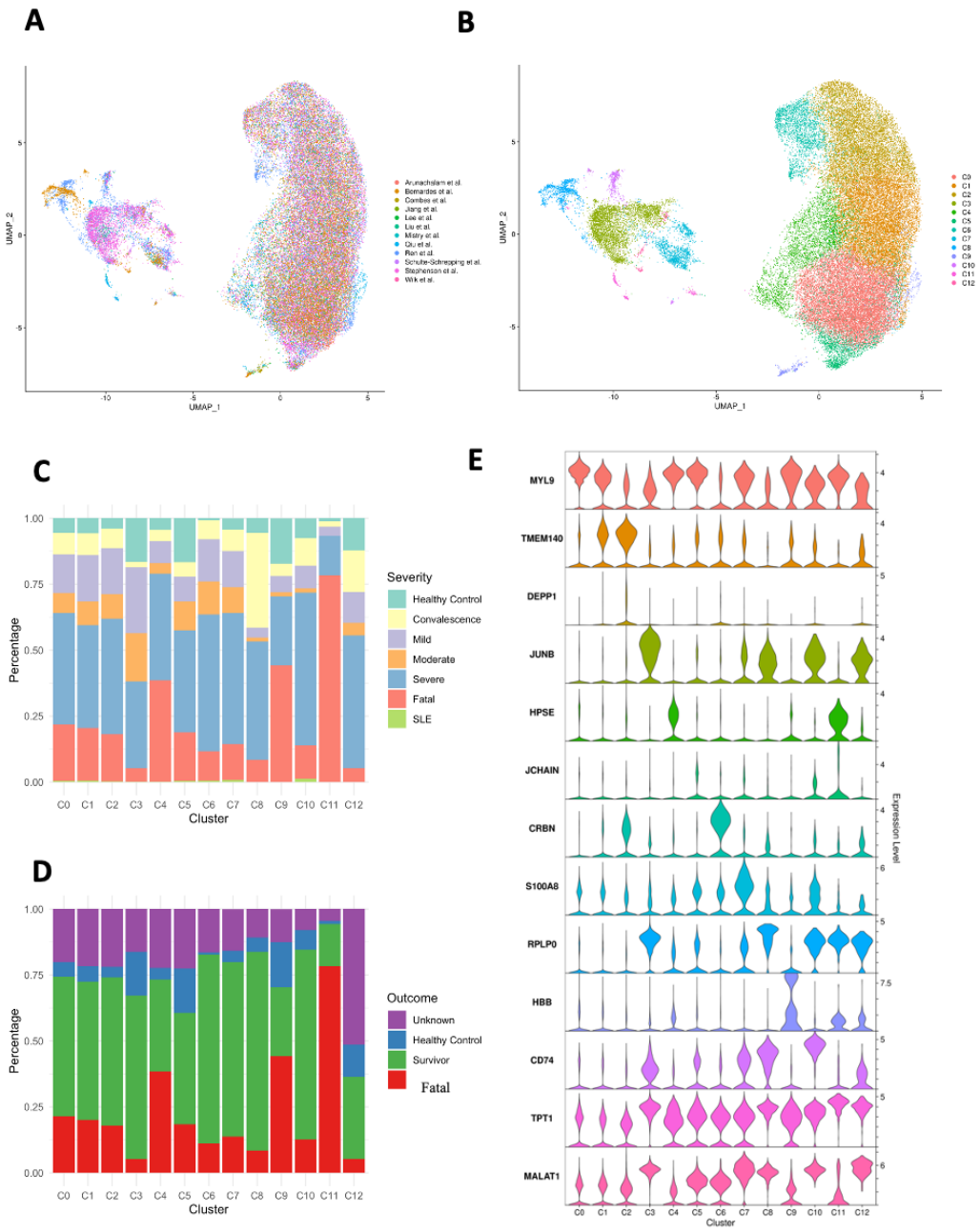

Figure 4.3 Integrated analysis of single-cell transcriptional profiles in platelet clusters.

(A, B) Uniform Manifold Approximation and Projection (UMAP) representation of merged platelet cell clusters. A) Color-coded by data source, B) Color-coded by cluster assignments. **(C, D)** Stacked bar plots illustrating the proportions of platelet clusters under C) varying disease severity levels, and D) distinct outcomes. **(E)** Violin plots depicting the expression levels of marker genes for each platelet cluster.

angiogenesis, coagulation, and endotheliopathy, while having the lowest RNA processing, cell division, and mitochondrial biogenesis, while those from the convalescence cluster C8 had the opposite trend in all these pathways.

C6, a survival cluster enriched in CRBN, CD58, and SRSF3, had the highest notch signaling pathway expression and the lowest IL6/JAK/STAT3 signaling pathway expression (Figure A.6). Allograft rejection and MYC targets v2 (Figure A.6) had higher scores in the other survival cluster C10, whose signature module includes the antigen presentation genes CD74, HLA-DRA, and CD79A. C6 and C10 also had lower angiogenesis (Figure A.6), coagulation, EMT, and apical junction (Figure 4.4 B-D) scores, all of which are signature pathways in fatal cluster C4.

Then, we examined the GO terms up-regulated in the fatal clusters C4, C9, and C11. The key pathways up-regulated in fatal cluster C4 include wound healing, platelet activation, hemostasis, and coagulation, which is consistent with the known observations that thrombotic problems are a major cause of morbidity and mortality in COVID-19 patients (340). Oxygen transport, hydrogen peroxide metabolism, gas transport, and erythrocyte development are enriched terms for C9, reflecting the hypoxic environment of C9 platelets (Figure 4.4F). ATP metabolic activities such as oxidative phosphorylation, cellular respiration, and aerobic respiration are among the enriched terms in C11; this is consistent with the C11 GSEA results, which suggested that C11 platelets are inactive (Figure A.6). Platelets in quiescence are known for requiring ATP for their basic function. According to a study, glycolysis produces up to 65% of the required ATP in inactivated platelets, with mitochondria providing the remainder (341). C4 is the most active

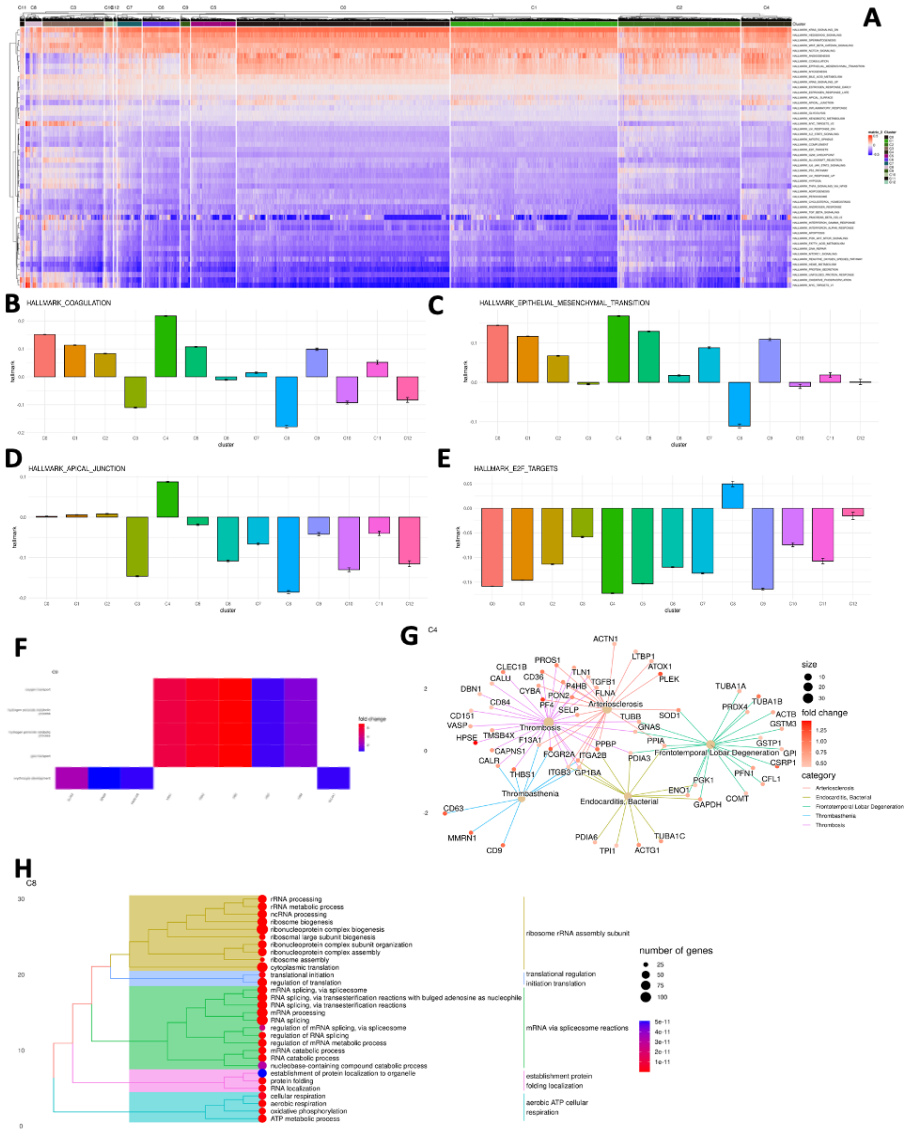

Figure 4.4 Characterization of platelet clusters and their distinct pathway expression changes.

(A) Enrichment analysis of human hallmark gene sets for each platelet cluster. Expression levels represent the normalized enrichment scores in the Gene Set Enrichment Analysis (GSEA) algorithm. **(B - E)** Bar plots depicting the expression of hallmark gene sets among clusters. B) Coagulation, C) Apical Junction, D) Epithelial-Mesenchymal Transition, E) E2F Targets. **(F)** Heatmap illustrating upregulated genes and enriched pathways in Gene Ontology (GO) for the fatal cluster C9. **(G)** Gene-concept network displaying upregulated genes and enriched diseases for the fatal cluster C4. **(H)** Tree plot showcasing enriched pathways in GO for the convalescent cluster C8.

cluster in FT groups. Therefore, we examined C4-enriched genes associated with diseases, such as arteriosclerosis, bacterial endocarditis, thrombosis, and frontotemporal lobar degeneration, a neurodegenerative disorder (Figure 4.4G). Among the mechanisms shared by survivor clusters C6, C8, and C10 are platelet translation regulation, mRNA processing, and ribosomal RNA biogenesis (Figure 4.4H, Figure A.6). Additionally, C10 was enriched for the GO terms homeostasis lymphocyte activation cells, and leukocyte antigen cell-cell adhesion (Figure A.6).

4.3.4 Pseudotime Trajectory Analysis Identifies Platelet Signature Dynamics in Survival or Fatal Disease Outcomes

The fatal cluster C4 appears to be critical for the negative outcome groups, so we evaluated possible events leading to its emergence. Pseudotime analysis, although not providing real time dynamics of cell populations, gives hints as to their order along the developmental trajectories (342). The possible precursor of C4 is the C0 cluster, which is also close to the C1 and C6 clusters in the trajectory map (Figure A.7). C0 cluster has two possible developmental routes, fatal C4 or survival C1and C6 (Figure 4.5A, B), indicating that the platelets in the C0 could be targeted by early intervention to inhibit their development into the C6 phenotype.

Then, we examined DEGs between C4 and C0 and C0 and C1. AKR7A2, CALD1, CALR, CD36, CSRP1, CYBA, ENSA, FCGR2A, HBG2, HCST, HMGN1, HPSE, MMRN1, NDUFA4, PF4, RPLP1, RPS9, SAMD14, and TPT1 genes are the consistently up-regulated, with a log2Fold change greater than 0.4 and an adjusted p value less than

0.05, from C0 to C4, and C1 to C0 (Figure 4.5C, D). For the further analysis we defined a fatal platelet module consisting of these genes. Among the fatal module enriched GO terms are regulation of body fluid levels, coagulation, phagocytosis, and tumor necrosis factor (TNF) production (Figure A.7). The DEGs consistently down-regulated between C4 vs. C0 and C0 vs. C1 were ADIPOR1, CDKN1A, MAP3K7CL, MMD, NEAT1, NUTF2, PTMA, RAB31, SLC40A1, TMEM140, TSC22D3, they were used to define the platelet survival module. However, no GO, KEGG, or Reactome pathways were enriched in the survival module. The scores for the fatality and survival modules were then calculated for all disease clusters and disease severity levels. As expected, cluster C2 had the lowest fatality score, followed by C1 and then survival cluster C6. C11 had the greatest fatality score, followed by C4 and C9 (Figure 4.5E). C2 had the greatest survival module score, followed by C1 and survival cluster C6, while C11 and C4 had the lowest scores for the fatal modules (Figure 4.5F).

The fatality and survival module scores could also serve as indicators of the disease outcome. The patient group with the highest score was the fatal group, followed by the severe group, while the mild group received the lowest score on the fatal module, followed by the moderate group. The HC group was positioned in the center of the fatality module (Figure 4.5G). The groups with the lowest survival module ratings were the fatal and severe groups. However, the highest scores were in the SLE and mild groups (Figure 4.5H). When platelets encounter an immunological disorder response, their expression shifted away from HC, as demonstrated by the findings. In addition, we investigated the composition of platelet subclusters and discovered that the fraction of fatal cluster C4 was the best

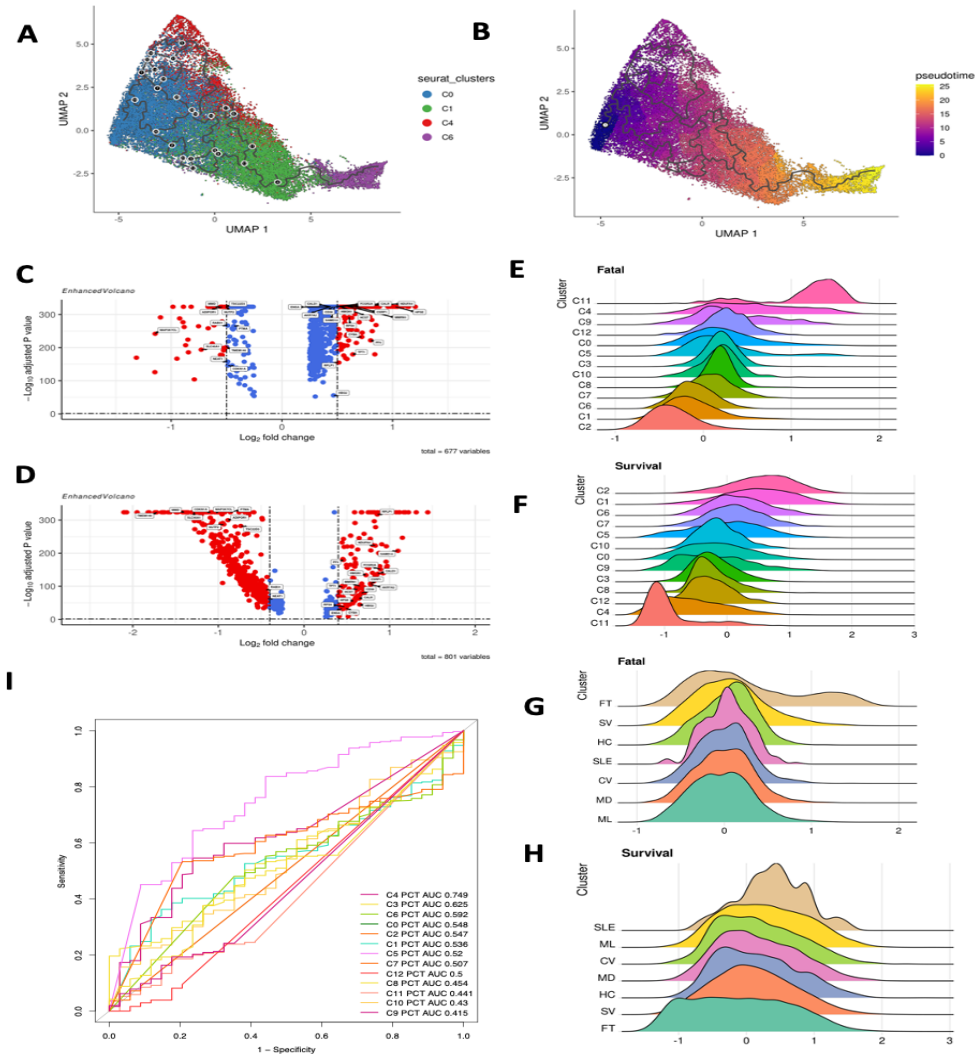

Figure 4.5 Platelet module signatures in patients associated with survival and fatal dynamic trends.

(A, B) Pseudotime plots of platelets from clusters C0, C1, C4, and C6, illustrating trajectory fates. (C, D) Differentially expressed genes in C) C4 vs. C0; D) C0 vs. C1. Volcano plots display the genes consistently upregulated or downregulated with increasing disease severity. (E - H) Ridge plots depicting the density of expression levels for E) fatal module expression across platelet clusters, F) fatal module expression across disease severities, G) survival module expression across platelet clusters, and H) survival module expression across disease severities. Ridge plots are ordered in descending order. (I) Receiver operating characteristic (ROC) curves for platelet cluster percentages within peripheral blood mononuclear cells (PBMCs), utilized to distinguish non-survivors from survivors.

indicator among clusters for distinguishing between S and FT patients (Figure 4.5I), with an AUC of 0.749. When C4 platelets exceeded 3.36 percent of the PBMC platelet composition, patients were at risk of death. In this case, detecting the presence of platelets C4 in a timely manner could help the patient's survival.

These findings show the complexity of COVID-19 and sepsis in relation to gene and pathway signatures in platelets. This is especially true for patients with severe active disease, as well as those recovering from it and surviving it. This research could lead to a better understanding options for both patient subgroups.

4.3.5 Unique and Shared Gene Expression Changes in Platelets from COVID-19, SSH, Sepsis, and SLE Samples

Comparing COVID-19, SSH, sepsis, and SLE platelets to HC platelets, we identified differentially expressed genes (DEGs) and related pathways (whose up- and down-regulations were analyzed using the GO and KEGG databases). There were 45 DEGs that were up-regulated in all four diseases studied here, including IFI27L2, IFITM2, IFITM3, and S100 family genes (S100A8 and S100A9) and genes for ATP synthase, such as ATP5E and C9orf16. BEX3, LCN2, RHEB, and TMEM219 are genes associated with apoptosis (Table B.6) (Figure A.8). 129 genes were downregulated in all four diseases (Figure A.8). The set of 129 genes analyzed comprised 68 genes related to ribosomes. Additionally, CD52, a predictive biomarker and therapeutic target for sepsis (14), was among the genes included. The genes CD3E, CD48, CD7, LCK, LEF1, PTPRC, and TCF7, essential for T cell activation, were also part of the set (Table B.6).

There were 47 pathways that were consistently up-regulated in COVID-19, SSH, sepsis, and SLE as compared to HC (Figure 4.6A). And 81 pathways were consistently down-regulated by all in the four respective diseases (Figure 4.6B). To better interpret the functional meaning of the consistently up/down-regulated pathways among the four diseases, we used Revigo (300) to summarize the 47 up-regulated pathways and 81 down-regulated pathways from COVID-19 vs. HC, SSH vs. HC, Sepsis vs. HC, and SLE vs. HC. Pathways up-regulated in platelets consistently among the four considered diseases include GO: neutrophil mediated immunity, KEGG: Parkinson disease, and GO: ATP metabolic process (Figure 4.6D). In Figure 4.6D, it can be observed that various translation processes are experiencing down-regulation. These processes include GO: nuclear-transcribed mRNA catabolic process, nonsense-mediated decay, GO: regulation of translational initiation, GO: protein localization in the endoplasmic reticulum, KEGG: Ribosome, and GO: response to interleukin-4. Interleukin-4 is known to have multiple biological effects, such as promoting the activation of B cells and T cell proliferation through pathways like T cell activation regulation and response (343).

We then focused on the pathways that were up-regulated in both COVID-19 and sepsis relative to HC. KEGG: Parkinson disease, KEGG: Huntington disease, KEGG: Prion disease, KEGG: Alzheimer disease, KEGG: Pathways of neurodegeneration - multiple diseases, and KEGG: Amyotrophic lateral sclerosis were consistently up-regulated in platelets from COVID-19 and sepsis (Figure A.8). COVID-19 has been recently associated with neurodegenerative disorders (344), even though the molecular mechanism of this association is not clear. Similar to the effects observed in sepsis,

COVID-19 may serve as a considerable inflammatory stimulus, heightening the brain's vulnerability to neurodegenerative disorders, cognitive deterioration, and an increased probability of developing dementia in later stages of life (345). Nonetheless, the observed up-regulation of pathways in platelets, which are typically present in neurons affected by neurodegenerative diseases, might indicate the involvement of shared regulatory processes rather than a direct influence of aberrant platelets on cerebral function. KEGG: Endocytosis, KEGG: Fc gamma R-mediated phagocytosis pathways were likewise up-regulated. Platelets endocytose virions after Toll-like receptors (TLRs) attach to virion released lysosomal ligands, such as ssRNA, dsRNA, and CpG DNA, downstream signaling induces platelet activation and granule release, exposing P-selectin, and subsequently generating platelet leukocyte aggregates (345). In addition to pathways associated with protein targeting, protein translation, and ribosomes, KEGG: Coronavirus disease- COVID-19, was also down-regulated in sepsis. However, all of the genes that were enriched for KEGG: Coronavirus disease - COVID-19 were ribosomal-associated proteins, which was mostly due to the down-regulation of ribosomal-related proteins in both COVID-19 and sepsis (Figure A.8).

Then, we investigated the pathways that are enriched in COVID-19 and sepsis fatal patients (FT) in comparison to survivors (S) (Figure 4.6C). GO: neutrophil mediated immunity and GO: ATP metabolic process were up-regulated in FT patients and were likewise up-regulated in disease vs. HC pathways. GO: response to endoplasmic reticulum stress, GO: response to hypoxia, and GO: intrinsic apoptotic signaling pathway in response to oxidative stress are up-regulated stress pathways. While the KEGG: Bacterial invasion

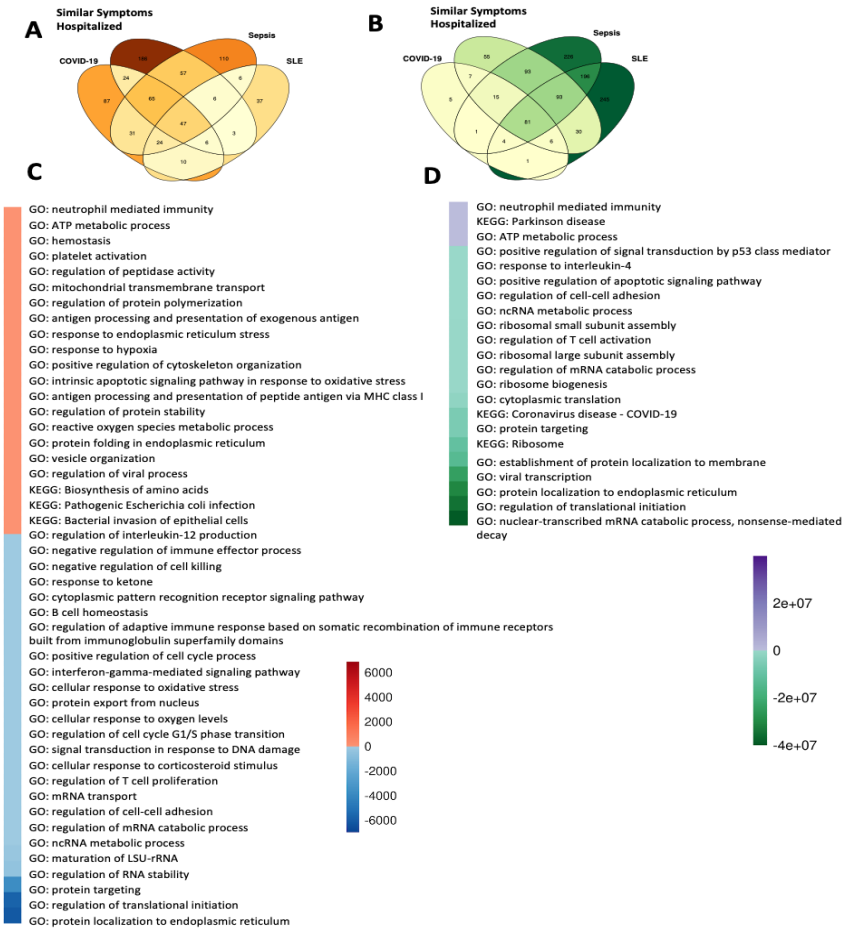

Figure 4.6 Platelet pathway expression among healthy controls, sepsis, similar symptom hospitalized, COVID-19, and SLE patients.

(A, B) Venn diagrams illustrating changes in Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways among COVID-19, similar symptom hospitalized (SSH), sepsis, and systemic lupus erythematosus (SLE) compared to healthy controls (HC). A) Upregulated pathways. B) Downregulated pathways. Pathways were filtered for those with an adjusted P-value below 0.05. (C) Heatmap representation of non-survivor vs survivor upregulated/downregulated pathways. Colors are determined by the product of the COVID-19 and sepsis upregulated/downregulated enriched pathway log10 (adjusted P-value). GO terms were reduced to representative ones using Revigo (with cutoffs set at similarity > 0.4) and then overlapped. (D) Heatmap representation of disease vs healthy control upregulated/downregulated pathways. Colors are determined by the product of COVID-19, SSH, SLE, and sepsis upregulated/downregulated enriched pathway log10 (adjusted P-value). GO terms were reduced to representative ones using Revigo (with cutoffs set at similarity > 0.1) and then overlapped.

of epithelial cells was also among the up-regulated pathways, we concluded that only the part of this pathway involved in cytoskeleton rearrangement is upregulated (see the discussion section). GO: antigen processing and presentation of peptide antigen via MHC Class I were also up-regulated in fatal patients of COVID-19 and sepsis. Platelet MHC Class I mediates CD8+ T-cell suppression during sepsis, according to a previous study (346). Pathways such as GO: protein localization to the endoplasmic reticulum, GO: regulation of translational start, and GO: regulation of RNA stability were down-regulated. In addition, lymphocyte activation pathways, such as GO: regulation of T cell proliferation, GO: interferon-gamma-mediated signaling pathway, and GO: regulation of interleukin-12 production, were down-regulated in FT patients. Interferon-gamma is mostly secreted by activated lymphocytes, including CD4 T helper type 1 cells and CD8 cytotoxic T cells (347), whereas interleukin-12 is known as T cell-stimulating factor (348) (Figure 4.6C). These data indicate that platelets from the disease cohorts exhibited less immunological activation, fewer translational activities, and more neurodegenerative tendencies, such as the KEGG enrichment for Parkinson's disease (Figure 4.6D). Compared to survivor platelets, the aforementioned tendencies became more pronounced in FT platelets, which began to exhibit the ability to invade epithelial cells (Figure 4.6C).

4.3.6 Pathway Enrichment Related to Disease Severity in Platelets

Given the dynamic gene expression changes in platelets in multiple diseases, we evaluated what gene expression modules were significantly changed depending on disease severity. Two modules that diminish with disease severity are MHC class II genes (Figure 4.7A)

and translation initiation (Figure 4.7B). MHC Class II scores are much higher in convalescent patients than in healthy controls, indicating that MHC Class II can be utilized as an indicator of patient recovery. High case fatality rates of COVID-19 reported in some countries have been linked to inadequate MHC class II presentation and, consequently, a weak adaptive immune response against these viral envelope proteins, according to studies (349). The lowest score for translation initiation modules (GO:0006413) was found in the platelets of patients who did not survive the diseases, indicating a halt in protein translation due to fatal illness. An interesting observation is the presence of pathways implicated in neurodegeneration (GO:0070843) in severely sick COVID-19 patients (Figure A.9). Except for axonal transport modules, platelets from severe and fatal COVID-19 patients show all the trends observed in neurodegeneration diseases' major biological processes. In both sepsis and COVID-19 patients, neurodegeneration-related pathways became more severe as the disorders advanced. The scores for blood coagulation (GO:0007596), platelet activation (GO:0030168) (Figure 4.2D), oxidative phosphorylation (OXPHOS) (Figure 4.7C), and glycolysis (260) (Figure 4.7D) modules confirmed the hypothesis that platelet coagulation and energy consumption are functionally linked to the severity of sepsis disease and the progression of COVID-19 disease (236, 350). Moreover, platelets in convalescent patients had higher glycolysis scores, which corresponded to module scores in response to oxygen radicals (Figure A.9), indicating that platelets in convalescent patients were also hypoxic. Hypoxia induces oxidative damage to neural cells and causes widespread neurodegeneration (351).

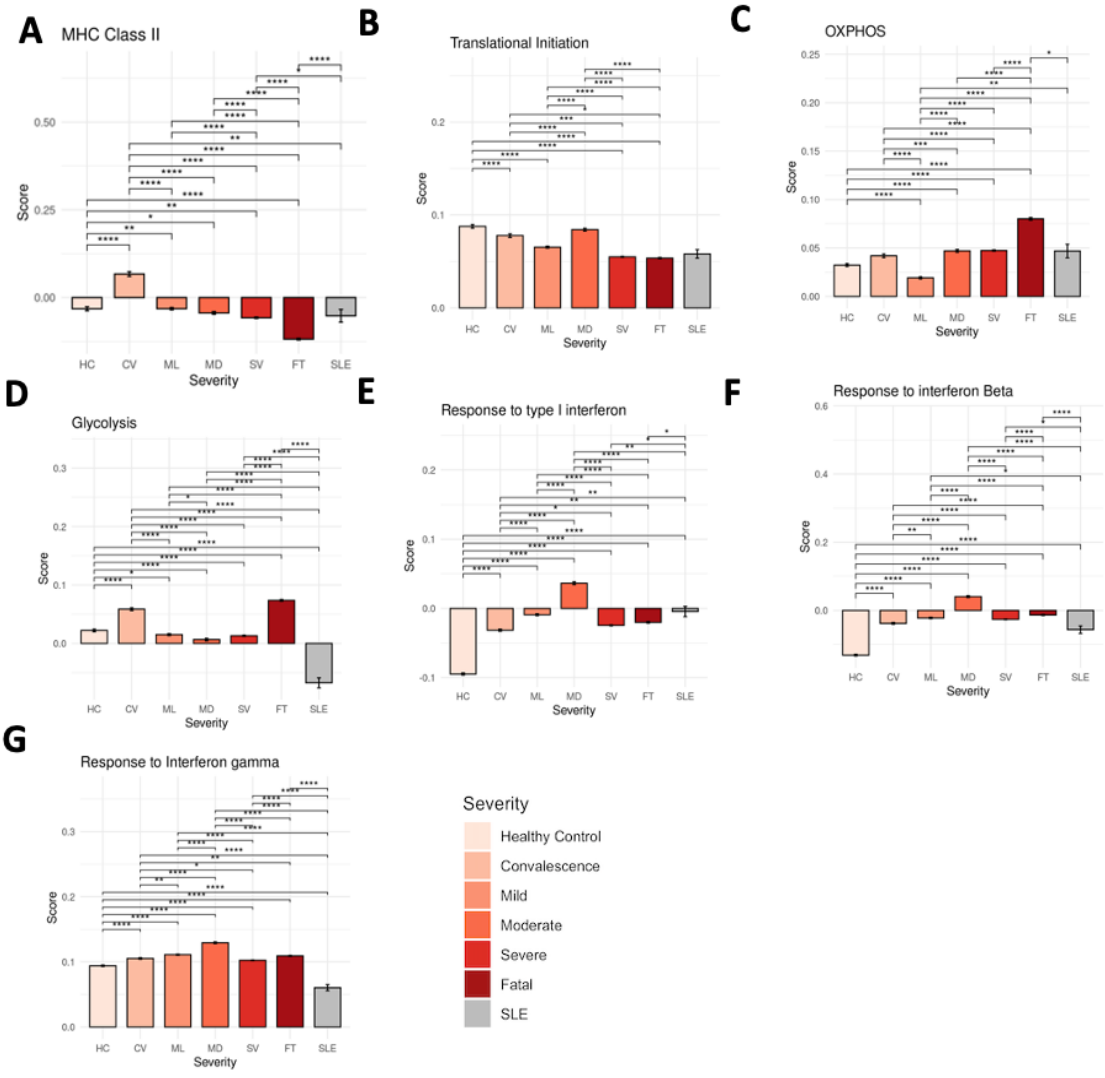

Figure 4.7 Platelet transcriptional alterations across varying disease severity levels.

(A–I) Assessment of pathway module scores among different disease severity levels in platelets. The analyzed modules include genes associated with: A) MHC Class II, B) Translation Initiation, C) OXPHOS, D) Glycolysis, E) Response to Type I IFN, F) Response to IFN gamma, and G) Response to IFN beta. Differences in scores with adjusted P-values below 0.05, 0.01, 0.001, and 0.0001 are denoted as *, **, ***, and ****, respectively, while non-significant differences are not shown. Significance analysis was conducted using Wilcoxon tests.

Interferon response modules, such as response to type I IFN (GO:0034340) (Figure 4.7E), IFN- β (GO:0035456) (Figure 4.7F), and IFN-γ (GO:0003341) (Figure 4.7G), had the highest scores in moderate patients and the lowest scores in HC. The interferon (IFN) protein family is crucial for the immune response against viruses and other infections. IFNs have been demonstrated to play a significant role in preventing SARS-CoV-2 infection in the context of COVID-19, but they have also been linked to severe symptoms (352). The findings may explain the contradictory reports of COVID-19 patients with impaired and robust type I IFN responses. Although robust type I IFN responses have been reported in patients with severe COVID-19 (272), studies have demonstrated that type I IFN responses are severely impaired in the peripheral blood of patients with severe or critical COVID-19, as indicated by low levels of type I IFNs and interferon-stimulated genes (353). Our data indicate that severe and fatal patients had IFN levels much lower than moderate patients, but greater than mild patients and healthy controls. In contrast to individuals with other disorders, COVID-19 patients have continuously elevated IFN levels. The above data conclude that the higher expression of MHC Class II and translational initiation expression in platelets were associated with better outcomes in patients. Coagulation and higher ATP synthesis from platelets, were linked to worst outcomes for patients. As for interferon response, with both protective and deleterious effects being reported, we confirmed the theory that severe COVID-19 is associated with decreased IFN signaling (281, 354).

4.3.7 Platelet Crosstalk with Monocytes and Lymphocytes

One of the important roles of platelets is communication with other cell types during formation of the thrombi, both regular ones forming during the hemostasis or potential abnormal ones forming during sepsis and COVID-19 disease. Using the ligand and receptor database from iTalk (92), we evaluated these interactions by computing the product of average ligand and receptor expressions in the corresponding cell types from peripheral blood mononuclear cells (PBMC) (see Materials and Methods). Platelet-monocyte interaction was evaluated and found to have the highest score in fatal patients relative to other outcomes (Figure 4.8A). Consistently with our previous sepsis study (99), platelet-monocyte interaction scores were significantly elevated in FT sepsis patients. This phenomenon was also noticed in COVID-19-severe patients. Compared to control participants and mildly infected individuals, ICU-admitted COVID-19 patients had higher platelet-monocyte aggregate levels (350). Thus, we postulate that the aggregation of platelets and monocytes is linked with the severity and mortality of sepsis and COVID-19.

Chemokine receptor 2 (CCR2), CCR5, and their selective ligands, chemokine ligand 2 (CCL2), and CCL3, have been found to promote the trafficking of leukocytes to sites of inflammation and regulate their activation (355). Also, the CCL2-CCR2 and CCL3-CCR5 ligand-receptor systems in differentiating T cells have been identified (356). CCL2-CCR2 and CCL3-CCR5 interactions between platelets and T cells revealed that SLE had the highest expression of the CCL3-CCR5 system between platelets and T cells, a characteristic of SLE (357). Compared to other outcome groups, the CCL3-CCR5 system's expression was the lowest among patients who did not survive the diseases (Figure 4.8B).

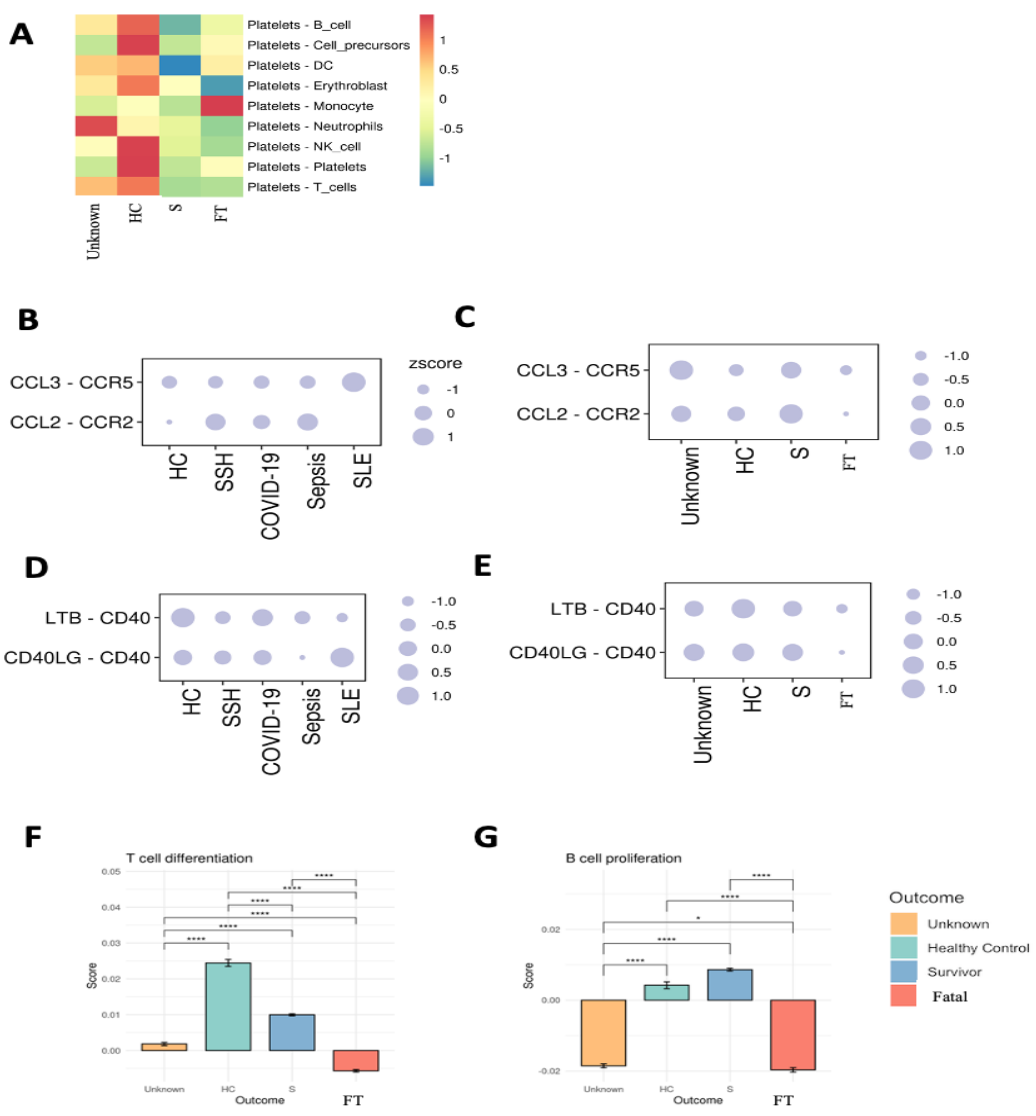

Figure 4.8 Alterations in platelet and other cell type interactions among healthy controls, sepsis, similar symptom hospitalized, COVID-19, and SLE patients.

(A) Comparison of ligand-receptor interaction scores between platelets and other cell types. Heatmap coloration corresponds to z-scored, log-normalized mean interaction scores averaged across all cells from a specific sample. **(B-E)** Ligand and receptor interaction scores between platelets and B) T cells across various disease severity levels, C) T cells across different outcomes, D) B cells across various disease severity levels, and E) B cells across different outcomes. Circle size represents the z-scored interaction scores. **(F, G)** Pathway module scores across different outcome situations in platelets, including F) T cell differentiation and G) B cell proliferation.

The results are associated with T cell differentiation module scores (GO:0030217) (Figure 4.8F). The CCL2-CCR2 and CCL3-CCR5 ligand-receptor systems are more prevalent in surviving patients than in HC (Figure 4.8C).

CD40 is recognized to play a crucial role in B lymphocyte proliferation and differentiation (358). And CD40 ligand CD40LG expression is low or undetectable on the surface of resting platelets but is highly expressed upon platelet activation (359). Also, platelets are noted to directly influence adaptive immune responses via the secretion of CD40 and CD40L molecules (2). Platelets and B cell interaction analysis revealed that healthy controls had the highest CD40 induction relative to other groups, whereas patients who did not survive had the lowest CD40 induction from platelets – B cell interaction (Figure 4.8E). The module score for B cell proliferation (GO:0042100) (Figure 4.8G) also supported this conclusion. SLE also had the highest expression of CD40LG and CD40 interaction between platelets and B cells compared to other diseases (Figure 4.8D). These data implicate that in SLE, platelets may induce B cell activity,

consistent with previous studies reporting the putative effects of activated platelets in SLE pathogenesis (315).

4.4 Discussion

For this study, we compiled single-cell transcriptome datasets from multiple cohorts of patients with inflammatory diseases with infectious and non-infectious roots (COVID-19, sepsis, SSH, and SLE). Extensive evidence points to a key role for platelets in all these diseases, severe COVID-19 (313, 340), fatal sepsis (99) and SLE (315) respectively. In our

analysis we identified both shared and unique signatures in platelet transcriptomics in patients with these diseases. We analyzed alterations in immune cell subsets in peripheral blood mononuclear cells (PBMCs) and found many features related to disease severity, outcome, and disease type. The ratio of platelets to T cells (Pla-T ratio) was shown to be a useful criterion for separating patients with fatal and survival outcomes (Figure 4.1H). Analysis of the upregulated pathways shows that pathways characteristic of the development and progression of endotheliopathy and disseminated intravascular coagulation (DIC) are upregulated in platelets. The patients who passed away had the highest expression of the gene involved in platelet-endothelial cell interactions and highest expression in the modules related to DIC, including actin cytoskeleton reorganization (Figure 4.2). Platelet actin cytoskeleton reorganization plays a role in this process, as the cytoskeleton helps to determine the shape and function of platelets and is essential for platelet activation and aggregation and changes observed here contribute to the pathological variants of this process.

We identified subgroups of platelets overrepresented in the convalescent, surviving, and fatal patients. Specifically, the three types of platelets were highly active in FT patients - the coagulation cluster C4, hypoxic cluster C9, and quiescence cluster C11. This suggests that a combination of anticoagulants, anti-hypoxic therapies, arteriosclerosis treatments, and selective serotonin reuptake inhibitors (SSRI) should be considered for the removal of fatal platelets. We examined the pseudotime trajectory of platelet clusters and identified C0 as the cluster preceding fatal cluster C4. C0 can be differentiated by following the path leading to the emergence of either fatal (C4) or survival (C1 leading to C6) cluster. We

propose that targeting C0 and modifying its developmental path may prove to be an effective strategy to treat immune instability in sepsis or COVID-19. Expression of MYL9, FCER1G, and PARVB are the signatures of cluster C0 (Figure 4.3E). MYL9 was investigated in relation to platelet dysfunction (360). Comparing the fatal cluster C4, its ancestor C0, and its alternative, survival cluster C1 to each other, we identified two gene modules associated with fatal outcome and survival and we labeled them accordingly (Figure 4.5C, D). Some genes in the fatal module, such as heparanase (HPSE) had already been studied as a target of therapy aimed at selectively neutralizing platelets to mitigate disease severity. HPSE expression and activity was shown to increase in platelets during clinical sepsis at both transcriptomic and proteomic level (361). The same phenomenon was also observed in COVID-19 patients' plasma (362). Thus, the fatal and survival module genes identified in this study provide potential clinical therapeutic targets, and diagnostic and prognostic biomarkers in platelets.

This study also investigated gene expression in four diseases (COVID-19, SSH, Sepsis, and SLE) compared to healthy controls (HC) and found consistent changes in gene expression, including stress pathways (Figure 4.6D) and decreased levels of MHC class II gene expression in fatal cases (Figure 4.7A). The presence of pathways also seen in neurodegeneration processes in the brain are also upregulated in severe COVID-19 patients and in all immune response disorders. Similarly, the ATP metabolic process are up-regulated but ribosome biogenesis and lymphocyte activation are commonly down-regulated (Figure 4.6C).

COVID-19 (363-366), sepsis (367, 368), and SLE (369) are all strongly associated with neurodegenerative diseases, but platelets have not been evaluated for their potential involvement. Platelets could potentially access the central nervous system by breaking the blood-brain barrier for a number of reasons, such as infection, neuroinflammation, neurodegeneration, and traumatic brain injury (370). Platelet changes in activation and aggregation have also been documented in Lewy body diseases, amyotrophic lateral sclerosis and multiple sclerosis (371-373). However, whether abnormal platelets are directly involved in triggering neurodegeneration, or if they instead respond to the overall immune environment in way similar to neurons undergoing degeneration, is unclear. This study illustrates the potential clinical significance of platelets as peripheral diagnostic biomarkers and, potentially, therapeutic targets for neurodegenerative diseases.

Platelet-monocyte aggregation have been studied as hallmark of severe COVID-19 (340, 374, 375) and sepsis (99, 376, 377). Those previous findings are consistent with our cell-cell interaction study (Figure 4.8A), which again provides the molecular mechanism for this effect. In addition, we discover that platelets may influence lymphocyte activation, proliferation, and differentiation through interactions with other immune cells, implying that platelets can modulate lymphocyte function and contribute to inflammatory and immune responses.

Overall, the study highlights the role of platelets in the development and progression of disease and the importance of monitoring changes in immune cell subsets and gene expression for predicting disease outcome and severity. We were able to study the etiology of dysfunctional platelets in sepsis, COVID-19, and SLE patients due to the

high number of clinical samples included in our integrated analysis. Our research demonstrates the significance of platelets' dysfunction in immunological imbalance disorders and supports the utility of platelet-directed therapies to treat multiple immune disorders.

CHAPTER 5 Impact of Smoking on Immune Cell Populations in Lung Cancer Patients

5.1 Introduction

Tobacco use remains the most significant preventable cause of death and disease in the United States, with over 480,000 Americans losing their lives to its effects annually. Although a plethora of studies have delved into the mutations resulting from nicotine and cigarette smoke genotoxicity (378-380), the ramifications of smoking on the immune system could be even more critical. Current research indicates that smoking has far-reaching consequences on chronic inflammation and autoimmunity at a systemic level (381-383). Nonetheless, the impact of smoking on individual immune cell types especially the tumor immune microenvironment (TIME) is an area that warrants further exploration.

Smoking exerts its influence on the immune system response in two main ways: by altering the composition of immune cells and by modifying the behavior of individual immune cell types (381, 384). While techniques such as cell sorting can identify changes in immune cell composition, the investigation of modifications in individual cell behavior remains underdeveloped. We hypothesize that single-cell sequencing techniques can shed light on smoking-induced changes within each cell type. Moreover, we postulate that the effects of smoking on immune cells could increase heterogeneity on the immune system. This increased heterogeneity can have detrimental effects on the immune system's ability to fight off cancer. For instance, in the context of lung cancer, the immune system of a smoker may display a higher degree of heterogeneity in immune cell populations, including

T cells. This could potentially result in a more diverse and complex TIME, making it harder for the immune system to effectively target and eliminate cancer cells (385).

In this chapter, we employed single-cell transcriptomics to examine the understudied question of smoking's effects on individual immune cells. We analyzed how smoking influences the behavior of each specific immune cell type and how the functional state transitions of immune cells are impacted by smoking. By doing so, we aimed to provide a more comprehensive understanding of the cellular and molecular mechanisms underlying the detrimental effects of smoking on the immune system. This knowledge can potentially inform the development of targeted therapies and preventive measures to mitigate the adverse health outcomes associated with tobacco use, particularly in the context of immune-related diseases and conditions.

5.2 Methods

To compare the gene expression pattern of smokers and never-smokers at gene expression level we used the previously generated scRNA-seq data, which have been deposited in the Sequence Read Archive (SRA, NCBI) /Gene Expression Omnibus (GEO) (GSE99254). The patient samples from Guo et al. (227) included 14 patients with lung cancer, 5 smokers (1 female and 4 males) and 9 never smokers (7 females and 2 males). We assessed the smoking-induced changes in tumor immune microenvironment at single cell level by pooling the data from both genders (due to the lack of enough cases for statistical analysis). The trajectory inference/pseudo-time for data from Guo et al. (227) was obtained using a Monocle2 (101) algorithm, which applied reverse graph embedding algorithm for the

trajectory construction. This is a mathematical approach that can order cells within a single cluster along a trajectory that can we interpreted as time but has an unknown relation to actual time (hence it is called a pseudo-time). To compare T cells CD4+ memory activated and T cells CD4+ memory resting ratio between smokers and never smokers, we used the AddModuleScore function from the Seurat package (57) with default settings to perform all calculations and comparisons of module scores. 22 functionally defined human immune subsets signature genes (LM22) for T cells CD4 memory activated, and T cells CD4 memory resting was applied to the calculate CD4 memory activate and resting modules scores (386). Genes without detectable expression in our data were ignored. The activated to resting ratio was calculated based on the module scores of CD4+ memory activate divided by the module scores of CD4+ memory resting. Additionally, we used the same data to compare the gene expression between smokers and never smokers in each predefined T cell category (CD4+ and CD8+) and tissue type. All comparison gene expressions between smokers and never smokers were conducted using the Seurat package the FindAllMarker function with default parameters. The feature plots were generated using R package Seurat FeaturePlot function. We subset T cells on the expression level of GPR15 > 0 as GPR15+ T cells, and compared the relative abundance of GPR15+ T cells between smokers and never smokers using Wilcoxon–Mann–Whitney test in R.

5.3 Results

5.3.1 Increased T cell heterogeneity in smokers

In a comprehensive analysis of single-cell RNA-sequencing (scRNA-seq) data, we utilized the dataset generated by Guo et al. (227), which has been made available in the Gene Expression Omnibus (GEO) databases (GSE99254). The original study by Guo et al. focused on patient samples collected from 14 individuals diagnosed with lung cancer, encompassing nine never-smokers and five smokers.

Guo et al.'s study identifies multiple T cell clusters through in-depth single-cell RNA sequencing. For CD4+ T cells, clusters CD4-C1 to CD4-C7 represent various functional states, ranging from naïve to exhausted. In never-smokers, naïve T cell cluster CD4-C1-CCR7 is observed, with conventional CD4+ T cell clusters CD4-C2-ANXA1 and CD4-C4-CD68 developing into two trajectories. One trajectory develops into effector CD4+ T cells CD4-C3-GNLY, which are characterized by the expression of cytotoxic molecules GNLY, enabling them to eliminate infected or cancerous cells. The other trajectory develops into exhausted T cells CD4-C7-CXCL13, marked by high expression of exhaustion markers PDCD1, CTLA4, HAVCR2, and TIGIT. In smokers, however, CD4-C1, C2, and C4 primarily develop into exhausted T cells (Figure 5.1a, b).

Regarding CD8+ T cells, clusters CD8-C1 to CD8-C6 are identified through single-cell RNA sequencing, representing various functional states from naïve to exhausted. In both smokers and never-smokers, naïve CD8+ T cell cluster CD8-C1-LEF1 and intermediate functional state cluster CD8-C2-CD28 develop into two trajectories. One

trajectory leads to effector CD8+ T cell cluster CD8-C3-CX3CR1, while the other develops into exhausted CD8+ T cell cluster CD8-C6-LAYN, which consists of dysfunctional T cells that have lost their ability to combat infections or cancer. However, in smokers, CD8+ T cells, like CD4+ T cells, exhibit greater heterogeneity, displaying more trajectory subtypes (Figure 5.2).

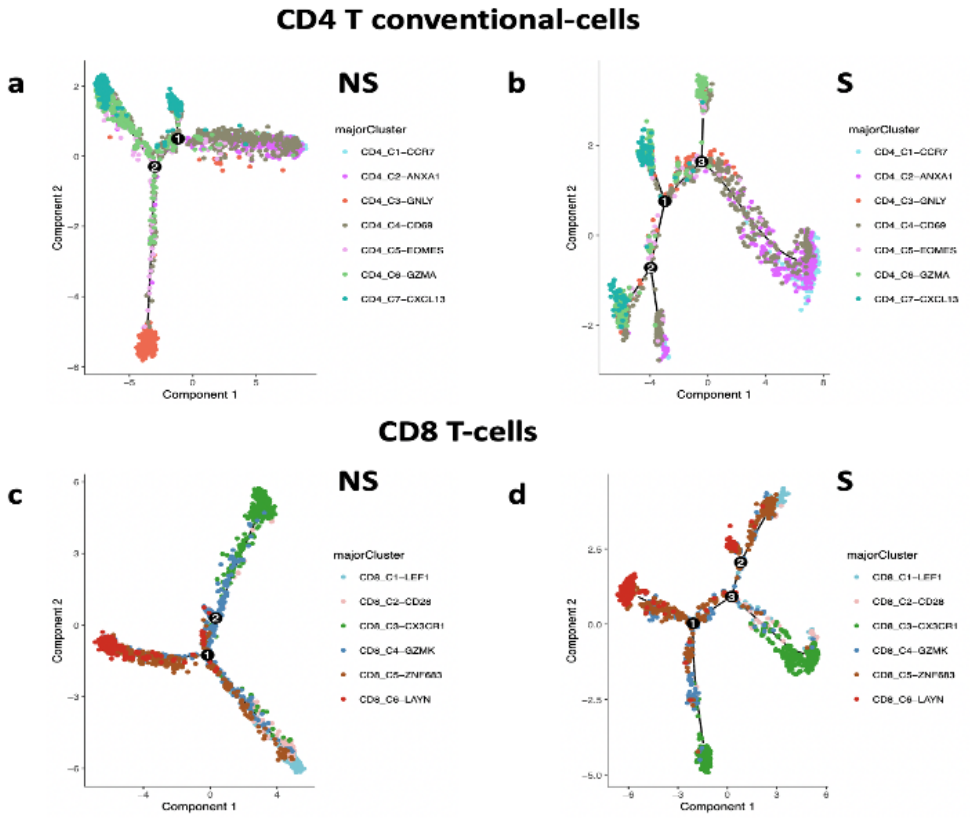

Figure 5.1 Comparison of T cell state transitions between never-smokers and smokers.

(a) Depicts the branched trajectory of CD4+ conventional T cell state transitions in a two-dimensional state-space for never-smokers. (b) Depicts the branched trajectory of CD4+ conventional T cell state transitions in a two-dimensional state-space for smokers. (c) Depicts the branched trajectory of CD8+ T cell state transitions in a two-dimensional state-space for never-smokers. (d) Depicts the branched trajectory of CD8+ T cell state transitions in a two-dimensional state-space for smokers. Each dot within the diagrams represents a single cell, with its color indicating the corresponding cluster label.

5.3.2 Smoking effects on tumor-immune microenvironment at the gene expression level

Furthermore, we run The Cancer Genomic Atlas (TCGA) datasets pan-cancer gene expression analysis for differences in immune-related gene expression. After correcting gender, age, and tumor histology, impact the profile of gene expression, GPR15 was found to be the only significant differentially expressed gene (DEG) between smokers and never smokers, in line with previous reports (387), and suggesting its potential as a smoking biomarker (Figure 5.2a and b). We also reanalyzed the scRNA-seq data from 14 lung cancer patients generated by Guo et al. (2018). We showed a significantly higher proportion of GPR15+ T cells in tumor tissue of smokers compared to never smokers while no significant changes were identified in normal tissue (Figure 5.2c). The higher expression of GPR15 gene was recorded in both CD4+ and CD8+ T cells in smokers compared to never smokers (Figure. 5.2d and e). The results indicate that smoking differently affects the expression of genes in different immune cell types. The significant differential expression of KLRC1 (as biomarker of cell exhaustion) in CD4+ T cells of patients with lung cancer indicates smoking might increase T cell exhaustion only in the tumor tissue compared to normal tissue of smokers (Figure. 5.2f), suggesting smoking increases complications in the tumor immune microenvironment such as increased risk of wound infections, poor healing, and metastasis.

5.4 Discussion

Compared to never-smokers, smokers exhibited more T cell heterogeneity, which has consequences for the tumor-immune microenvironment at the gene expression level. The

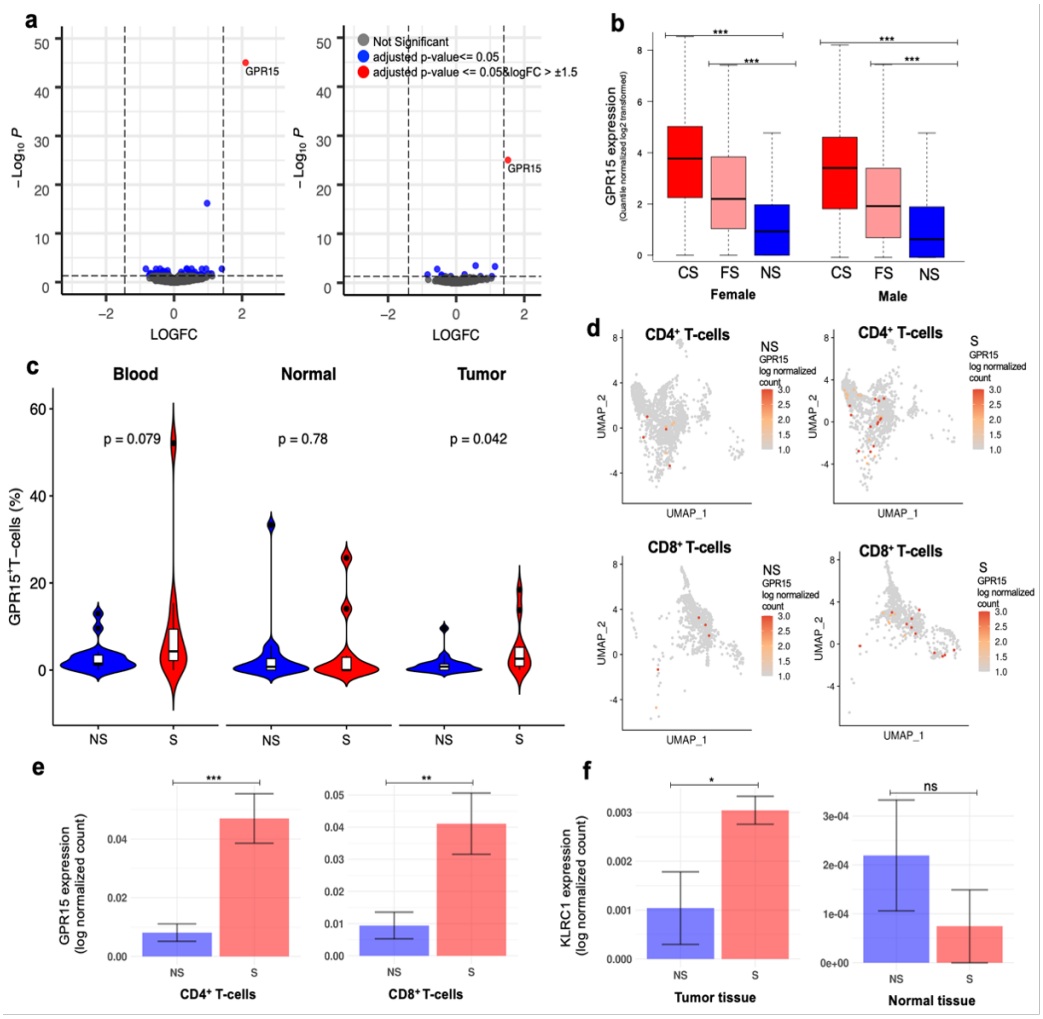

Figure 5.2 Differentially expressed genes (DEGs) between smokers and never smokers.

(a) and (b) DEGs in pan-cancer analysis of female and male cases, respectively. The FDR adjusted p-values (q-values) were obtained from the moderated t-test after controlling for confounding variables, including type of cancer, tumor pathologic stage, ethnicity, and race. (c) The proportion of GPR15+ T cells (both CD4+ and CD8+) of patients with lung cancer (Single-cell RNA-seq data adopted from study of Guo et al., 2018) was compared between smokers and never smokers using Wilcoxon–Mann–Whitney test. (d) UMAP plots showing the expression of GPR15 in CD4+ and CD8+ T cells in smokers and never smokers. (e) Single-cell RNA-seq based GPR15 expression in smokers and never smokers in tumor infiltrating CD4+ and CD8+ T cells of patients with lung cancer, respectively. (f) Single-cell RNA-seq based KLRC1 expression between smokers and never smokers in CD4+ T cells of patients with lung cancer. ns, *, ** and *** represents not significant, q-values < 0.05, 0.01 and 0.001, respectively. S, CS, FS, and NS denote for smoker, current smoker, former smoker, and never smoker, respectively.

observed higher heterogeneity of CD4+ and CD8+ T cells in smokers is expected to impair the immune system's capacity to combat infections and cancers. Furthermore, the increased heterogeneity of T cells in smokers may contribute to the immunosuppressive tumor microenvironment frequently observed in smoking-related malignancies. The increased proportion of exhausted T cells in smokers, as indicated by the elevation of exhaustion markers, may compromise the immune system's ability to efficiently detect and kill cancer cells. This may increase the formation and advancement of tumors. A recent lung cancer scRNA-seq study (388) demonstrated that enhanced immunological heterogeneity had a greater impact on tumor progression and drug resistance in smokers.

In addition to the direct impacts of smoking on T cell heterogeneity, our work focused on the influence of smoking on gene expression levels within the TIME. Smoking-induced alterations in gene expression manifest differently across various immune cell types, further highlighting the complex relationship between smoking and immune system function. The observed alterations in gene expression may also impact treatment outcomes, as some cancer treatments rely on the immune system's capacity to mount an effective response to cancer cells.

5.5 Acknowledgements

In this chapter, I acknowledge the collaborative efforts of a group of esteemed professionals. My co-authors on this work include Dr.Arghavan Alisoltani, Dr.Lukasz Jaroszewski, Dr.Mallika Iyer, Dr.Mayya Sedova, and Dr.Adam Godzik, who all contributed in various ways to our study on smoking-induced changes in the tumor immune

microenvironment (389). I particularly thank for the first author of the study Dr.Arghavan Alisoltani for conducting the TCGA pan cancer analyses and contributed to Figure 5.2a and b.

CHAPTER 6 Conclusion

Over the past decade, single-cell RNA sequencing (scRNA-seq) has revolutionized the field of transcriptomics. This cutting-edge technology has made previously unattainable studies achievable by allowing researchers to profile the gene expression patterns of individual cells within a heterogeneous population. This has provided unprecedented insights into cellular diversity, differentiation, and function. scRNA-seq has enabled the identification of rare and previously unknown cell types, as well as the discovery of unique gene expression patterns within individual cells. This has improved our understanding of the complex cellular landscapes in tissues and organs. scRNA-seq has also been instrumental in uncovering the dynamics of cellular differentiation and developmental processes. By tracking gene expression changes over time in individual cells, researchers have been able to reconstruct lineage trajectories and identify key regulatory factors that drive cell fate decisions. In addition, scRNA-seq has provided valuable insights into disease mechanisms by revealing cell type-specific gene expression changes in various pathological conditions. Also, the application of scRNA-seq in cancer research has allowed for a more comprehensive understanding of tumor heterogeneity, immune cell infiltration, and the identification of novel cellular subpopulations associated with drug resistance or metastasis. Lastly, scRNA-seq has been used to investigate how individual cells within a population respond to drug treatment, revealing the molecular mechanisms behind drug resistance and providing a foundation for personalized medicine.

Despite these advances, scRNA-seq still faces some challenges, including technical issues related to sensitivity, scalability, and data analysis. However, ongoing developments

in experimental techniques and computational methods are expected to further enhance the power and applicability of scRNA-seq in the coming years.

Chapter 1 provides an introduction to the scRNA-seq analysis, outlining the currently available protocols and technologies, including tissue preparation, single-cell isolation, and library preparation. The chapter also covers statistical and computational methods for scRNA-seq data analysis, including data preprocessing, quality control, normalization, imputation, data integration, dimensionality reduction, clustering, cell type prediction and annotation, differential expression analysis, gene set analysis, cell-cell communication networks, and gene regulatory networks. Additionally, the chapter discusses the use of deep learning applications in scRNA-seq analysis, specifically in dimension reduction and clustering, cell type prediction and annotation, and treatment response prediction. The chapter concludes with a discussion of the importance and potential future directions of scRNA-seq analysis.

In Chapter 2 of the dissertation, the focus is on the utilization of scRNA-seq within the immune system. The chapter commences with an introductory section and delves into the diverse nature of immune cells and their respective functions. Topics covered include the identification of immune cell types and subtypes, the characterization of immune cell states and activation, and the analysis of immune cell population dynamics during infectious diseases. Furthermore, the chapter examines immune cell dysregulation in autoimmune disorders and the characterization of inflammatory responses in chronic illnesses. The discussion then shifts to cancer immunology and immuno-oncology, exploring the tumor microenvironment, immune cell infiltration, and individualized

immunotherapy approaches. The chapter wraps up with a comprehensive review of the subject matter discussed.

In Chapter 3, we identified abnormal immune cell subsets, changes in functional pathways, and molecular signatures associated with fatal or surviving outcomes in sepsis by using scRNA-seq data. We observed the emergence and continuous changes in abnormal immune cells, including new types of cells unique to sepsis and classical cell types present in both sepsis and healthy controls but with abnormal gene expression profiles and changes in population ratios. The study provides foundation data and identifies specific cell subsets and molecular pathways that can be further explored to better predict and possibly modify sepsis outcomes. The major findings of this study are that fatal sepsis is associated with the expansion of platelets and erythroid precursors and the immunosuppressive trend of monocytes. Additionally, the study identified CD52 expression in lymphocytes as a potential biomarker and therapeutic target for sepsis, where it correlated with increased lymphocyte activation and survival outcomes. This research could lead to new diagnostic or therapeutic approaches for sepsis. However, further research is needed to fully understand the implications of these findings and their potential clinical applications.

In Chapter 4, we integrated twelve public datasets with platelets' single cell transcriptomics data, they were from sepsis, COVID-19, and SLE. We investigated the molecular pathway changes in each disease compared to healthy platelets. We found that the diseased platelets have lower expression of MHC class II, translation initiation, and have higher expression relative to neurodegenerative diseases. We categorized patients

according to patients' metadata into different disease severities including healthy controls, convalescence, mild, moderate, severe and fatal patients. We found the severe and fatal patients' platelets have higher expression of coagulation, inflammatory response, extracellular matrix disassembly, and cytoskeleton organization, indicating a higher degree of endotheliopathy in patients with severe and fatal outcomes. We also identified the clusters of abnormal platelets have more proportion in fatal patients, and platelets that have higher proportion in recovered patients and survival patients but low in proportion in severe and fatal patients. We looked into the trajectories developed into the fatal clusters and survival clusters, and identified two gene modules, the fatal module genes include AKR7A2, CALD1, CALR, CD36, CSRP1, CYBA, ENSA, FCGR2A, HBG2, HCST, HMGN1, HPSE, MMRN1, NDUFA4, PF4, RPLP1, RPS9, SAMD14, and TPT1. The survival module genes include ADIPOR1, CDKN1A, MAP3K7CL, MMD, NEAT1, NUTF2, PTMA, RAB31, SLC40A1, TMEM140, and TSC22D3. In addition to that, we investigated the cell-cell interaction between platelets and other immune cell types and found that platelet and monocyte aggregation as a hallmark of patient disease severity. Also, abnormal platelets reduce lymphocyte activation and differentiation by ligand and receptor interaction. Our research has the potential to pave the way for novel diagnostic and therapeutic strategies for sepsis and severe COVID-19, specifically focusing on platelets. This study offers foundational data and highlights distinct platelet subsets and molecular pathways that can be investigated further to enhance the prediction and potential alteration of outcomes in critical patients. The discovery of gene module expression in platelets associated with survival and fatality presents a promising biomarker and therapeutic target,

which could contribute to the development of innovative diagnostic tools or treatments. Nevertheless, additional research is required to thoroughly comprehend the significance of these findings and their prospective clinical applications.

In Chapter 5, we investigated the tumor immune microenvironment (TIME) varies depending on the patient's smoking history. Smoking is one of the leading known risk factors for lung cancer, in our study, we attempted to further understand the association between smoking and the heterogeneity in TIME. With high-resolution scRNA-seq data, we investigated the immune system in smokers vs. never smokers in two perspectives, one is about the alterations in the immune cell type changes, the other one is specific gene expression changes vary according to cell types and subtypes. We found smoking increases the effect of T cell heterogeneity in lung cancer patients. We also found smoking-induced changes in gene expression display distinct patterns across diverse immune cell types, underlining the intricate interplay between smoking and immune system functionality. These alterations in gene expression could potentially influence treatment outcomes, as certain cancer therapies depend on the immune system's ability to launch a robust response against cancerous cells.

In conclusion, we have discussed single-cell RNA-seq (scRNA-seq) data analysis computational and statistical methods, also scRNA-seq applications in the immune system study. Investigating immune system heterogeneity in survival and fatal outcomes using single-cell RNA sequencing (scRNA-seq) is crucial for several reasons. First, it provides a high-resolution understanding of the diverse immune cell populations and their functional states, which can have significant implications for a patient's response to cancer or

infections. Second, studying heterogeneity can help identify potential biomarkers and therapeutic targets, enabling the development of more personalized and effective treatment strategies. Furthermore, examining immune cell heterogeneity can uncover previously unknown immune cell subsets or pathways associated with favorable or adverse clinical outcomes. This knowledge can facilitate the development of novel diagnostic tools or therapies aimed at modulating the immune system to improve patient survival. Overall, studying immune system heterogeneity through scRNA-seq holds immense potential for enhancing our comprehension of the immune landscape and its impact on patient prognosis. Hence, by utilizing both collaborations collected data and publicly available datasets, our research examines heterogeneity in the immune system within infectious diseases and the tumor microenvironment. This approach is essential for driving biologically significant discoveries, as it enables a deeper understanding of the complex immune landscape and its impact on disease progression and patient outcomes.

APPENDIX A Supplementary Figures

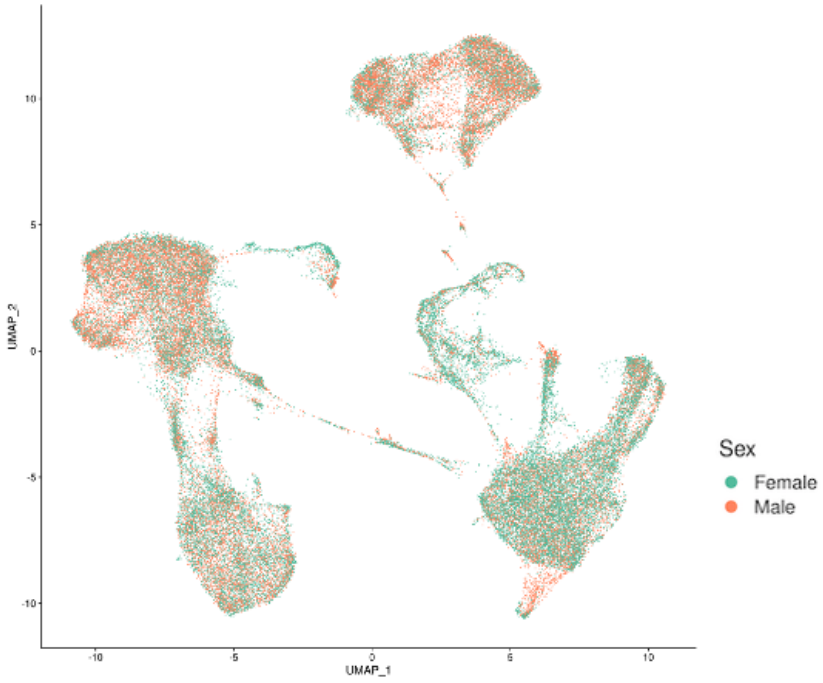

Figure A.1 UMAP representation of all merged samples.

Points representing cells are colored according to the donor's sex.

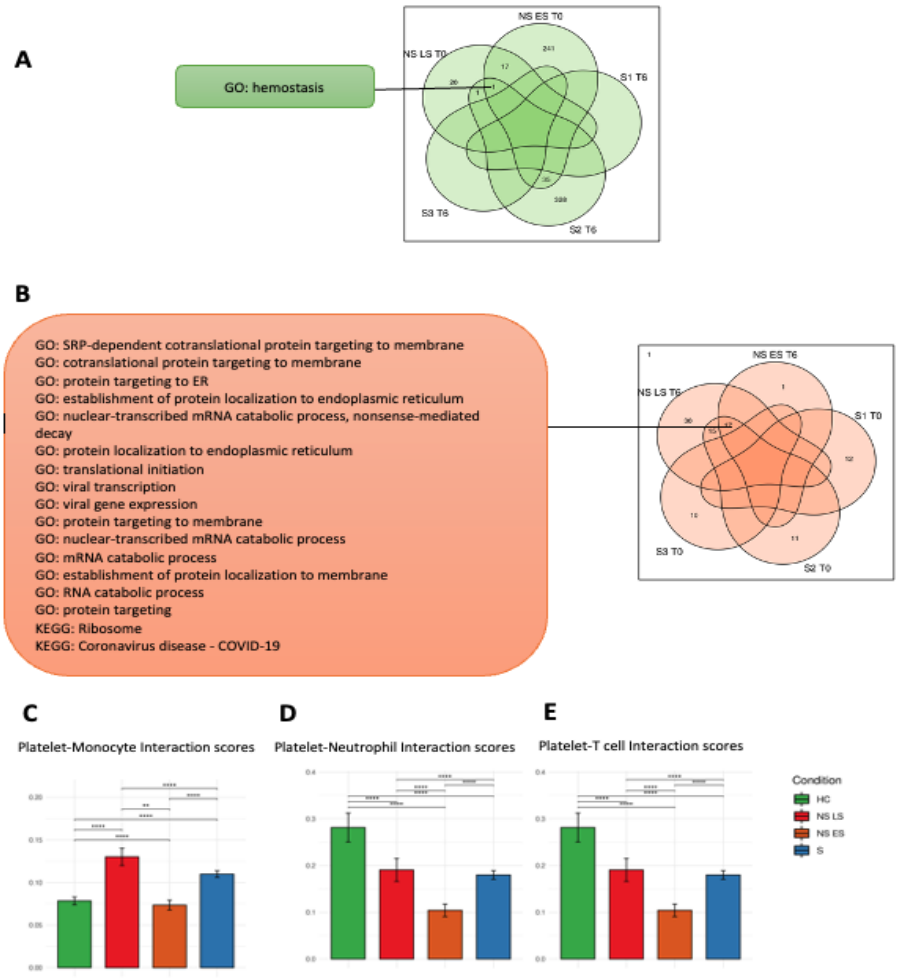

Figure A.2 Aberrant platelet function changes observed in fatal sepsis.

(A, B) Venn diagrams showing the number of pathways up-regulated and down-regulated in temporal changes in platelet cells from sepsis patients. Sets of pathways are labeled with patients' IDs and time. For example, the set labeled NS ES T0 contains pathways decreasing in the sepsis non-survivor early stage between T0 and T6 and the set labeled S2 T6 contains pathways increasing between T0 and T6 in platelets from sepsis survivor #2, etc. A) Sets of pathways decreasing in non-survivors and increasing in survivors (green). B) Sets of pathways increasing in non-survivors and decreasing in survivors (orange). The pathways which showed consistent changes in three or more patients are listed on the left. (C, D, E) Platelet and other cell types interaction scores across conditions. C) Platelet-Monocyte Interaction scores, D) Platelet-Neutrophil Interaction scores, E) Platelet-T cell Interaction scores. The differences in scores associated with adjusted p-values below 0.05, 0.01, 0.001, and 0.0001 are indicated as *, **, ***, and ****, respectively. The significance analysis was performed using Wilcoxon test.

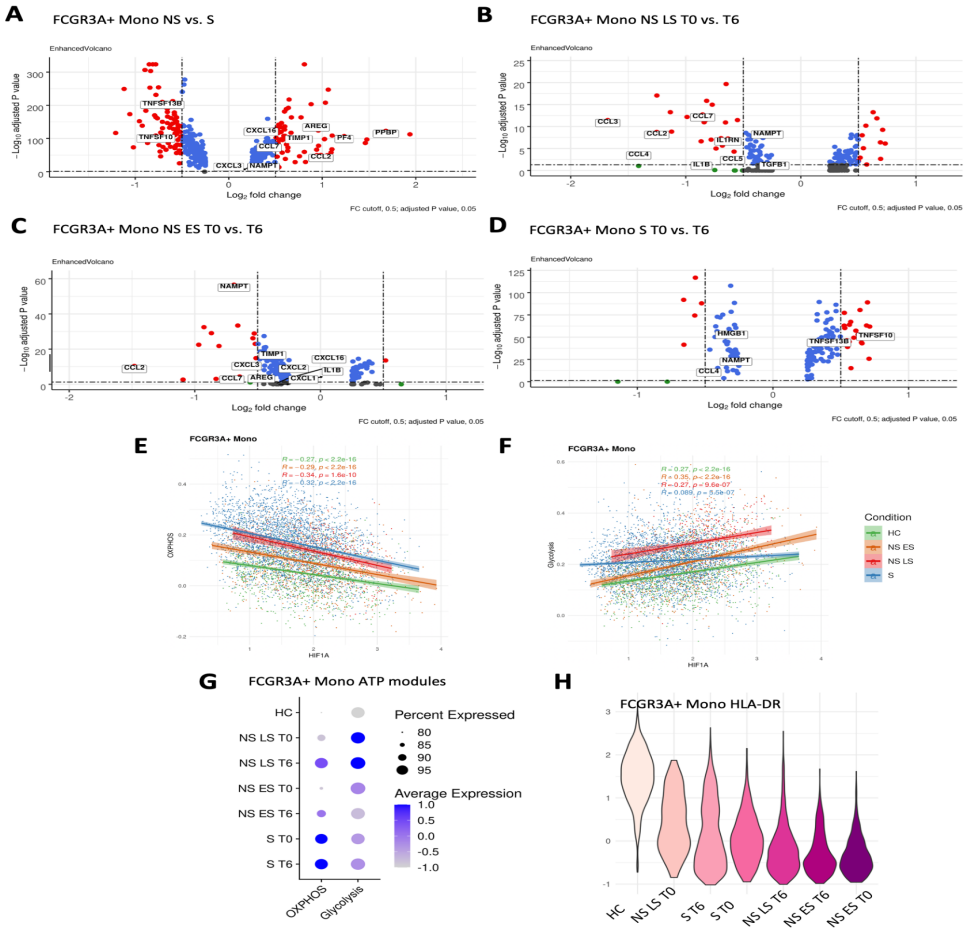

Figure A.3 Immunosuppressive trends in monocytes from sepsis non-survivors.

(A – D) Differential expression of genes in FCGR3A+ monocytes from comparisons: A) NS versus S, B) NS LS T0 versus T6, C) NS ES T0 versus T6, D) S T0 versus T6. (E, F) The correlations between the HIF1A expression and: E) module score for OXPHOS and F) glycolysis module score. Correlations are shown separately for each condition. R-values from Pearson's correlation, exact two-sided p-values and the 95% confidence intervals are shown on each graph. Each dot represents a single cell. Only cells with HIF1A expression above 0 were included in the analysis. Green, orange, red, and blue points represent cells from HC, NS ES, NS LS and S samples, respectively. (G) The expression of OXPHOS and glycolysis pathway modules in FCGR3A+ monocytes from healthy controls and from sepsis patients at T0 and T6. The color saturation indicates the average expression level, and the circle's size indicates the percentage of cells expressing a given module. (H) HLA-DR related genes expression in FCGR3A+ monocytes across healthy controls and sepsis conditions at T0 and T6. Violin plots are ordered with the decreasing expression average value of HLA-DR related genes. The color saturation indicates the average expression level, the darker the color the lower the average expression level.

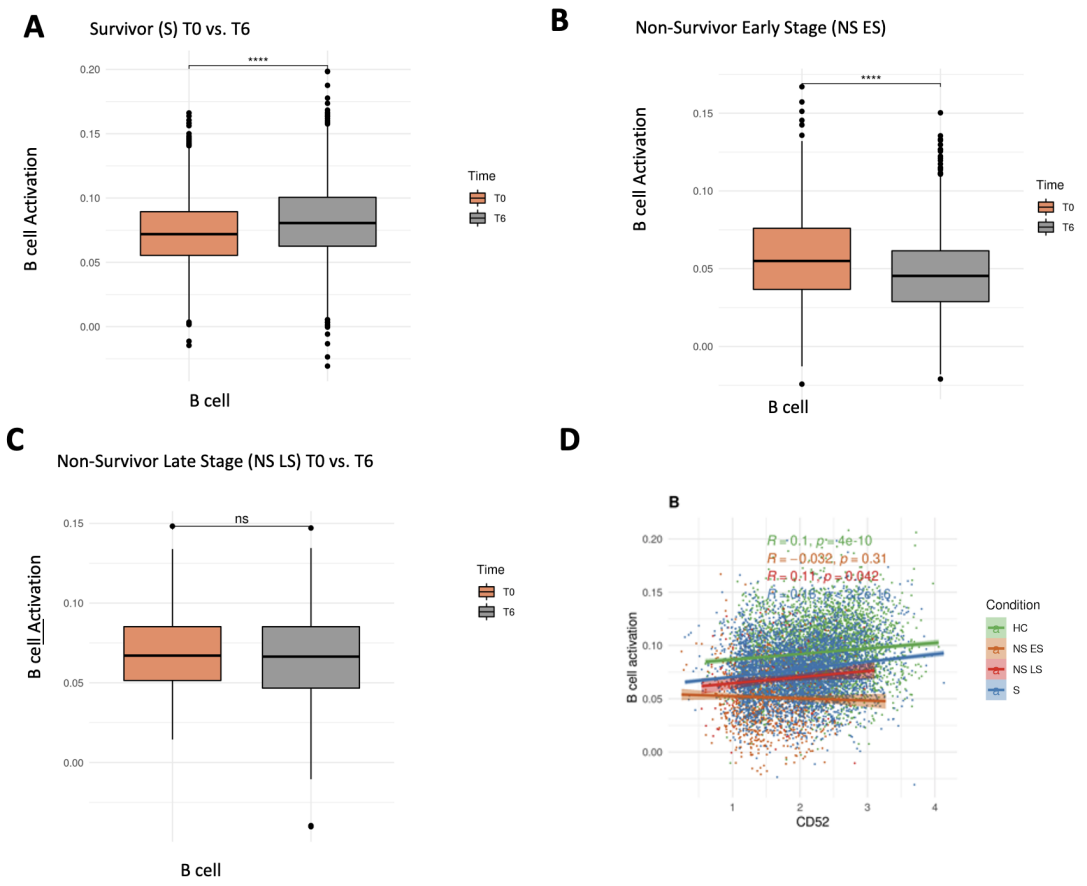

Figure A.4 The association between CD52 expression and B cell activation.

(A-C) B cell activation pathway module scores comparison between T0 vs. T6 in B cells. B) Survivors, C) Non-survivor early stage, D) Non-survivor late stage. The differences in scores associated with adjusted p-values below 0.05, 0.01, 0.001, and 0.0001 are indicated as *, **, ***, and ****, respectively. The significance analysis was performed using Wilcoxon test. (D) CD52 expression and its correlation with the B cell activation pathway module score in B cells across four conditions. R-values from Pearson's correlation, exact two-sided p-values and the 95% confidence intervals are shown on each graph. Each dot represents a single cell. Only cells with CD52 expression above 0 were included in the analysis. Green, orange, red, and blue points represent cells from HC, NS ES, NS LS and S samples, respectively.

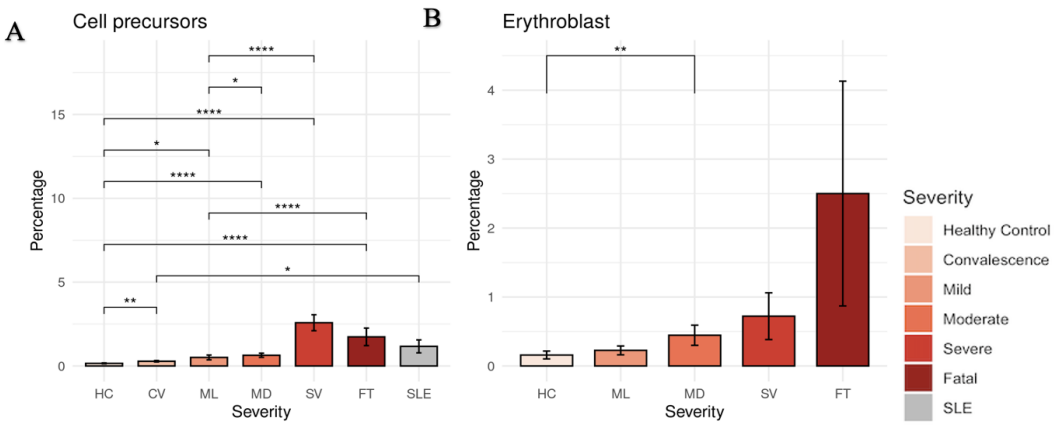

Figure A.5 PBMC profiling from healthy controls, sepsis, similar symptom hospitalized, COVID-19 and SLE patients.

(A-C) Bar plots depicting the percentage of different cell types under different disease severities. A) Cell precursors, B) Erythroblast
The differences in percentages associated with adjusted *P*-values below 0.05, 0.01, 0.001, and 0.0001 are indicated as *, **, ***, and ****, respectively and not significant are not shown. The significance analysis was performed using Wilcoxon tests. Standard error bars were also added.

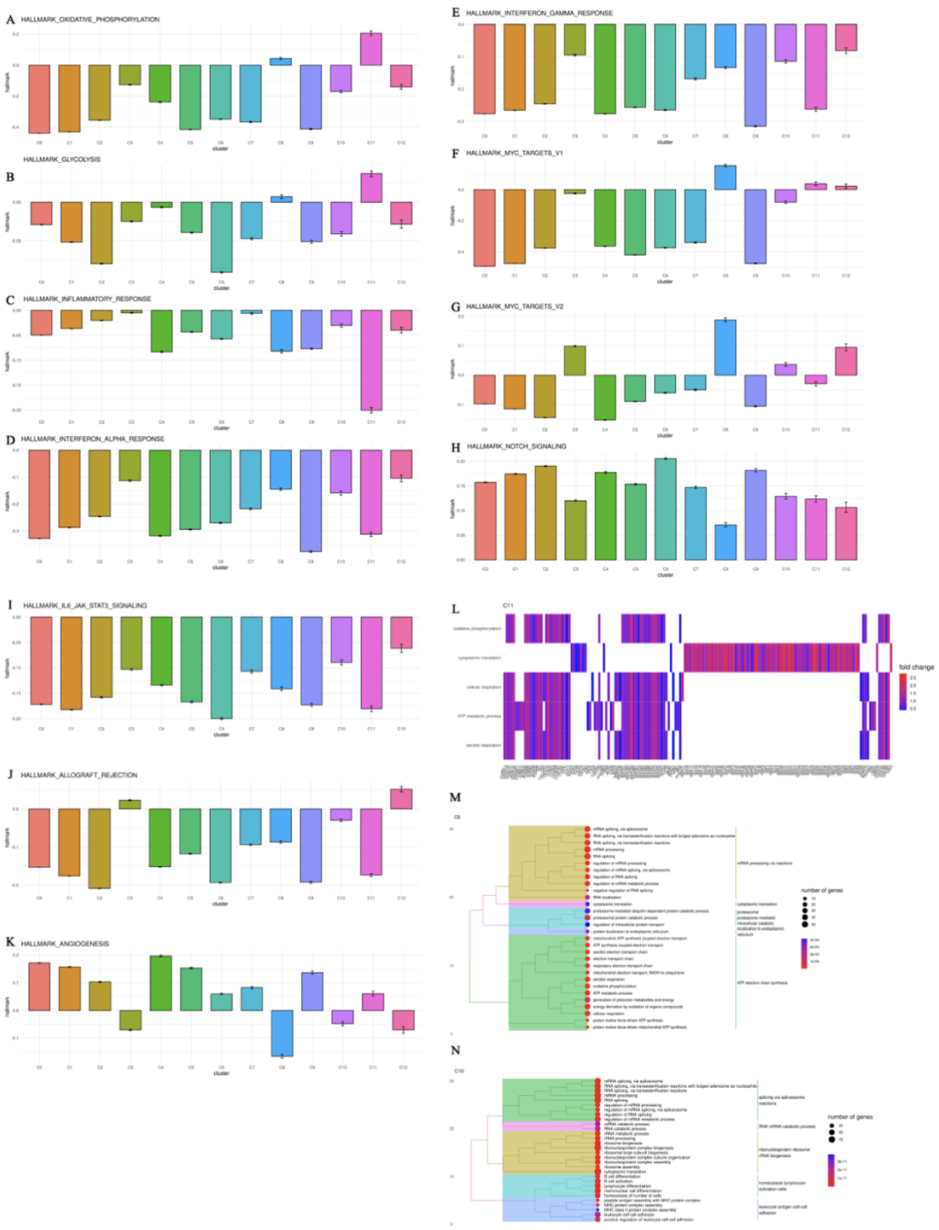

Figure A.6 Clustered platelets and their unique pathway expression changes.

(A - K) Bar plot of hallmark gene sets expression among clusters. A) MYC_targets_v1, B) MYC_targets_v1, C) Oxidative phosphorylation D) Glycolysis E) Inflammatory response F) Interferon alpha response, G) Interferon gamma response, H) Notch signaling I) IL6/JAK/STAT3 signaling. (L) Heatmap display fatal cluster C11 up-regulated genes and enriched pathways in gene ontology (GO). (M, N) Tree plot display survival cluster C6 and C10 enriched pathways in GO.

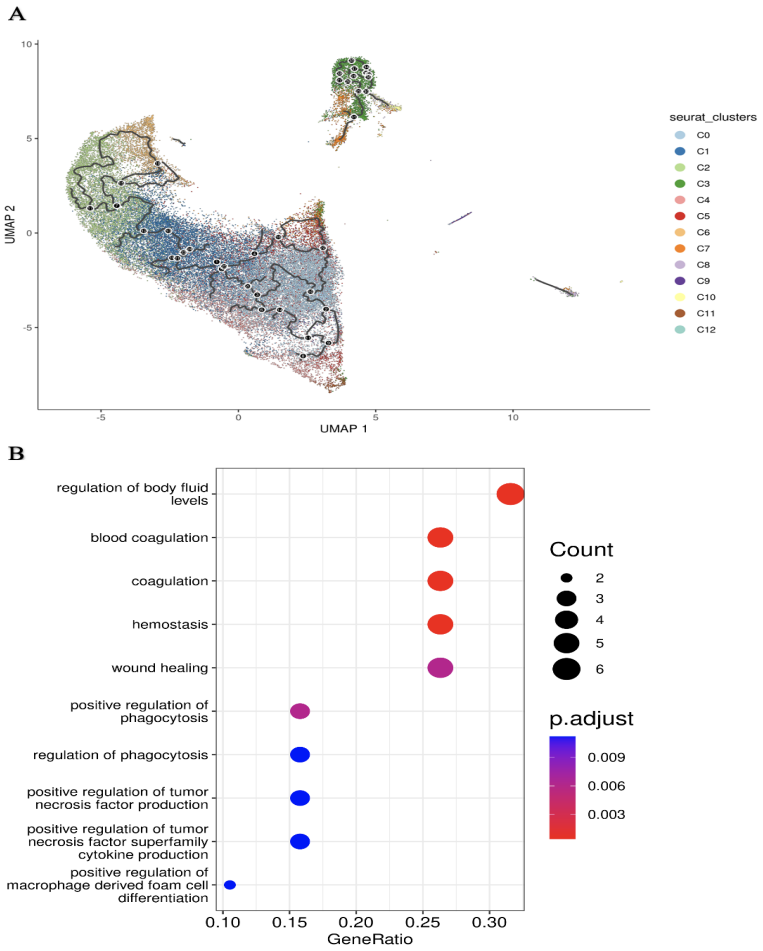

Figure A.7 Pseudo-time plot of platelets from all clusters exhibiting trajectory fates

(A) Pseudo-time plot of platelets from all clusters exhibiting trajectory fates. (B) The fatal gene module list enriched pathways from GO, KEGG, and Reactome databases. The circle size represents the gene counts included in the pathway; color represents the adjusted P value.

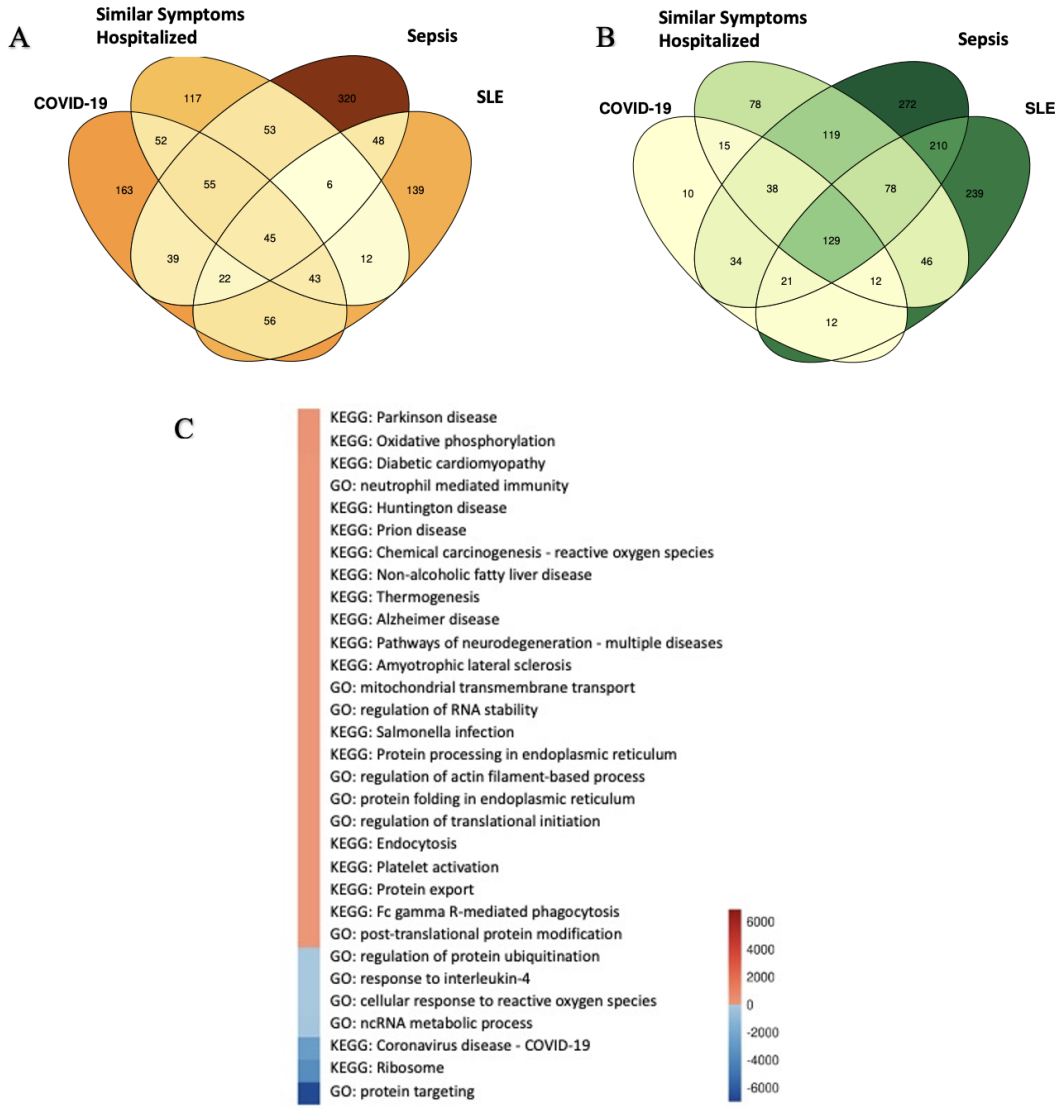

Figure A.8 Platelet's expression changes among healthy controls, sepsis, similar symptom hospitalized, COVID-19 and SLE patients

Venn diagrams describing changes in differential expression genes (DEGs) from COVID-19, similar symptoms hospitalized (SSH), sepsis, and systemic sclerosis lupus (SLE) vs. healthy controls (HC), A) Up-regulated genes. B) Down-regulated genes. DEGs with adjusted p value below 0.05 and log2 fold change over 0.25 were used for generation of the VENN figure. (C) Heatmap illustration of COVID-19 and sepsis patients vs healthy control pathways. Colors are decided by the product of the COVID-19 and sepsis up/down-regulated enriched pathway log10 (adjusted p value). The GO terms were reduced to representative ones using Revigo (300) (the cutoffs were similarity > 0.4) then overlapped the terms.

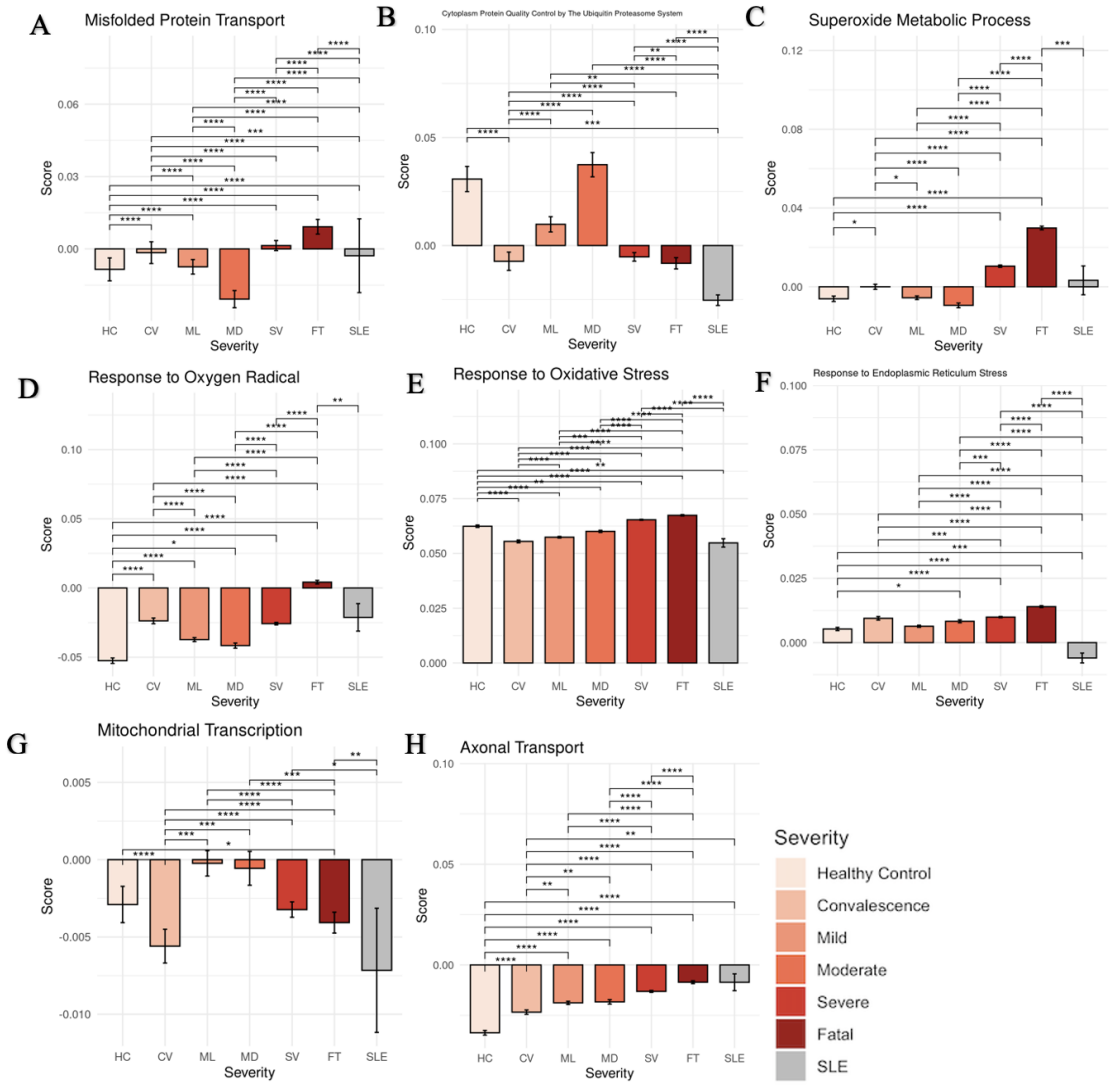

Figure A.9 Platelet transcriptional changes under different disease severity related to neurodegeneration diseases.

(A– H) Comparisons of pathway module scores across disease severities in platelets. The included modules contain genes related to A) Misfolded Protein Transport, (B) Cytoplasm Protein Quality Control by The Ubiquitin Proteasome System, (C) Superoxide Metabolic Process, (D) Response to Oxygen Radical, (E) Response to Oxidative Stress, (F) Response to Endoplasmic Reticulum Stress, (G) Mitochondrial Transcription, (H) Axonal Transport. The differences in scores associated with adjusted *P*-values below 0.05, 0.01, 0.001, and 0.0001 are indicated as *, **, ***, and ****, respectively and not significant ones are not shown. The significance analysis was performed using Wilcoxon tests.

APPENDIX B Supplementary Tables

Table B.1 Absolute cell numbers for each cell type from all samples.

Sample	B	CD14+ Mono	CD4+ T	FCGR3A+ Mono	CD8+ T	NK	DC	Platelet	Erythroid precursors	Neutrophil	CMP	Sample Total
HC1	1474	1091	2549	621	879	662	35	98	27	7	10	7453
HC2	2462	1201	2010	830	520	656	206	126	64	3	15	8093
NSLS_T0	156	201	214	77	328	410	35	779	232	5	3	2440
NSLS_T6	194	441	258	259	205	260	131	1195	364	8	3	3318
NSES_T0	492	753	904	559	215	257	21	286	46	152	27	3712
NSES_T6	998	730	2049	549	272	382	54	272	59	145	36	5546
S2_T0	147	929	337	444	195	229	23	97	42	25	0	2468
S2_T6	514	1095	477	635	363	396	135	198	37	59	4	3913
S2_T0	2744	278	1726	84	278	430	948	81	66	39	1	6675
S2_T6	1061	755	1309	835	491	625	47	32	39	90	9	5293
S3_T0	262	929	409	781	172	163	24	27	3	16	2	2788
S3_T6	760	934	1577	887	638	455	80	69	7	26	1	5434
Cell Type Total	11264	9337	13819	6561	4556	4925	1739	3260	986	575	111	57133

Table B.2 Genes used for gene module scoring.

Module	Genes
T cell activation	ABL1, ABL2, ADA, ADAM8, ADAM17, ADK, ADRM1, AGER, AIF1, ANKLE1, ANXA1, AP3B1, AP3D1, APBB1IP, APC, ARMC5, ATP7A, AZI2, B2M, BAD, BATF, BAX, BCL2, BCL3, BCL6, BCL10, BCL11A, BCL11B, BLM, BMI1, BRAF, BTN2A2, BTNL2, CACNB4, CARD11, CARMIL2, CAV1, CBFB, CBLB, CCDC88B, CCL2, CCL5, CCL19, CCL20, CCND3, CCR2, CCR6, CCR7, CCR9, CD2, CD3D, CD3E, CD3G, CD4, CD5, CD6, CD8A, CD24, CD27, CD28, CD40LG, CD44, CD46, CD47, CD48, CD74, CD80, CD81, CD83, CD86, CD151, CD160, CD274, CD276, CDH26, CDK6, CEACAM1, CEBPB, CGAS, CHD7, CHRNA7, CLEC4D, CLEC4E, CLEC4F, CLPTM1, CORO1A, CRACR2A, CRIP3, CRTAM, CSK, CTLA4, CTNNB1, CTPS1, CTSL, CXADR, CXCL12, CXCR4, CYLD, CYP26B1, DDOST, DHPS, DICER1, DLG1, DLL4, DNAJA3, DOCK2, DOCK8, DPP4, DROSHA, DUSP10, EBI3, EFNB1, EFNB2, EFNB3, EGR1, EGR3, EIF2AK4, ELF4, ENTPD7, EOMES, EPHB4, EPHB6, EPO, F2RL1, FADD, FANCA, FANCD2, FCER1G, FGL2, FKBP1A, FKBP1B, FLOT2, FLT3, FOXN1, FOXP1, FOXP3, FUT7, FYN, FZD5, FZD7, FZD8, GADD45G, GATA3, GBA, GIMAP1, GJA1, GLI3, GPAM, GPR18, GPR89B, GPR183, GSN, HAVCR2, NCKAP1L, HES1, HLX, HS1BP3, HSH2D, HSP90AA1, HSPD1, HSPH1, ICAM1, ICOS, ICOSLG, IFNA1, IFNA2, IFNA4, IFNA5, IFNA6, IFNA7, IFNA13, IFNA14, IFNA16, IFNAR1, IFNB1, IFNE, IFNG, IFNK, IGF1, IGF2, IGFBP2, IHH, IKZF1, IL1A, IL1B, IL1RL2, IL2, IL2RA, IL2RG, IL4, IL4I1, IL6, IL6R, IL6ST, IL7, IL7R, IL12A, IL12B, IL12RB1, IL15, IL18, IL18R1, IL20RB, IL21, IL23A, IL27, IL27RA, IL36B, IRF1, IRF4, ITGAD, ITGAL, ITGAM, ITGAV, ITGAX, ITGB2, ITK, ITPKB, JAG2, JAK3, JAML, JMJD6, KAT2A, KDELR1, KIT, KLRC1, LAT, LCK, LCP1, LEF1, LEP, LEPR, LFNG, LGALS1, LGALS3, LGALS8, LGALS9, LIG4, LMBR1L, LMO1, LY9, MAFB, MALT1, MAPK8IP1, MDK, METTL3, MPZL2, MR1, MSN, MTOR, MYB, MYH9, NCAPH2, NCK1, NCK2, NCKAP1L, NCOR1, NCSTN, NEDD4, NFATC3, NFKBID, NFKBIZ, NHEJ1, NKAP, NKX2-3, NLRC3, NLRP3, OTUD5, P2RX7, PAG1, PATZ1, PAX1, PCK1, PDCD1LG2, PIK3R6, PKNOX1, PLA2G2D, PNP, PPP3CB, PRDM1, PRDX2, PRELID1, PREX1, PRKCQ, PRKCZ, PRKDC, PRR7, PSAP, PSEN1, PSEN2, PSMB10, PSMB11, PTGER4, PTPN2, PTPN22, PTPRC, PYCARD, RAB27A, RAB29, RABL3, RAC2, RAG1, RAG2, RARA, RASAL3,

	RASGRP1, RC3H1, RC3H2, RELB, RHOA, RHOH, RIPK2, RIPK3, RORA, RORC, RPL22, RPS3, RPS6, RSAD2, RUNX1, RUNX2, RUNX3, SART1, SASH3, SATB1, SCRIB, SELENOK, SEMA4A, SH3RF1, SHB, SHH, SIRPA, SIT1, SLA2, SLAMF1, SLAMF6, SLC4A1, SLC7A1, SLC11A1, SLC46A2, SMAD3, SOCS1, SOCS5, SOS1, SOS2, SOX4, SOX12, SOX13, SP3, SPI1, SPN, SPTA1, SRF, STAT3, STAT5A, STAT5B, STAT6, STK11, STOML2, STX11, SYK, TBX21, TCF3, TCF7, TCIRG1, TRA, TRB, TESPA1, TFRC, TGFB1, TGFBR2, THEMIS, THY1, TMEM98, TNFRSF1B, TNFRSF4, TNFRSF13C, TNFRSF14, TNFSF4, TNFSF8, TNFSF9, TNFSF11, TNFSF13B, TNFSF14, TNFSF18, TOX, TRAF3IP2, TRAF6, TRAJ18, TREML2, TREX1, TSC1, TXK, VAV1, VCAM1, VNN1, VSIR, VTCN1, WAS, WASHC1, WDFY4, WNT1, WNT4, WWP1, XBP1, XCL1, XRCC4, ZAP70, ZBTB1, ZBTB7B, ZBTB16, ZBTB32, ZEB1, ZFP36L1, ZFP36L2, ZMIZ1, ZNHIT1, ZP3
B cell activation	ABL1, ADA, ADAM17, ADGRG3, AHR, AICDA, ANKLE1, APLF, ATAD5, ATM, ATP11C, BAD, BANK1, BATF, BAX, BCL2, BCL3, BCL6, BCL11A, BLNK, BMI1, BST1, BTK, CARD11, CCR6, CD19, CD22, CD24, CD27, CD28, CD38, CD40, CD40LG, CD70, CD74, CD79A, CD79B, CD81, CD86, CD180, CD320, CDH17, CDKN1A, CHRNA4, CHRNA7, CHRNB2, CLCF1, CMTM7, CR2, CTPS1, CXCR5, CYLD, DCAF1, DCLRE1C, DLL1, DNAJB9, DOCK10, DOCK11, EP300, EPHB2, ERCC1, EXO1, EXOSC3, EXOSC6, EZH2, FCRL1, FLT3, FNIP1, FOXP1, FOXP3, FZD9, GAPT, GIMAP1, GON4L, GPR183, GPS2, HDAC5, HDAC7, HDAC9, NCKAP1L, HHEX, HMCES, HSPD1, ICOSLG, IFNA1, IFNA2, IFNA4, IFNA5, IFNA6, IFNA7, IFNA13, IFNA14, IFNA16, IFNB1, IFNE, IFNG, IFNK, IGBP1, IGHD, IGHE, IGHG1, IGHG3, IGHM, IGHV1-12, IGHV1-18, IGHV1-24, IGHV1-58, IGHV1-67, IGHV1-69, IGHV2-5, IGHV3-6, IGKC, IGLC1, IGLC2, IGLC3, IGLL1, IKZF1, IKZF3, IL2, IL2RG, IL4, IL4I1, IL5, IL6, IL7, IL7R, IL10, IL13, IL21, IL27RA, INPP5D, IRF2BP2, IRF8, IRS2, ITFG2, ITM2A, JAK3, KIT, KMT5B, KMT5C, LAT2, LAX1, LEF1, LFNG, LGALS1, LGALS8, LIG4, LYL1, MAD2L2, MALT1, MEF2C, MFNG, MIF, MLH1, MMP14, MS4A1, MSH2, MSH6, MYB, MZB1, NBN, NCKAP1L, NDFIP1, NFAM1, NFATC1, NFATC2, NHEJ1, NKX2-3, NOD2, NOTCH2,

	NSD2, NTRK1, ONECUT1, PAXIP1, PCID2, PELI1, PFDN1, PHB, PHB2, PIK3CD, PIK3R1, PLCG2, PLCL2, PMS2, PNP, POLM, POU1F1, POU2AF1, POU2F2, PPP2R3C, PRDM1, PRKCB, PRKCD, PRKDC, PTK2B, PTPN2, PTPN6, PTPRC, PTPRJ, RABL3, RAG1, RAG2, RASGRP1, RBPJ, RIF1, RNF8, RNF168, SASH3, SH3KBP1, SHB, SHLD1, SHLD2, SHLD3, SKAP2, SLAMF8, SLC15A4, SLC25A5, SLC39A10, SP3, SPI1, SPIB, ST3GAL1, STAT5A, STAT5B, STAT6, SWAP70, SYK, SYVN1, TBX21, TCF3, TCIRG1, TFRC, TGFB1, THEMIS2, TICAM1, TIRAP, TLR4, TLR9, TNFAIP3, TNFRSF4, TNFRSF13C, TNFSF4, TNFSF13, TNFSF13B, TNIP2, TPD52, TRAF3IP2, TRBC1, TRBC2, TRDC, TSHR, TXLNA, UNG, VAV3, VPREB1, WNT3A, XBP1, XRCC4, YY1, ZBTB1, ZBTB7A, ZFP36L1, ZFP36L2
Coagulation	ABAT, ADAMTS13, ADAMTS18, ADRA2A, ADRA2C, ADRB2, ADTRP, ALOX12, ANO6, ANXA2, ANXA5, ANXA8, AP3B1, APOE, APOH, AXL, BLK, BLOC1S3, BLOC1S4, BLOC1S6, C1GALT1C1, C1QTNF1, C3, C9, CAV1, CD9, CD34, CD36, CD40LG, CEACAM1, CELA2A, CFH, COMP, CPB2, CX3CL1, DMTN, DTNBP1, ENPP4, ENTPD1, ENTPD2, EPHB2, EVL, F2, F2R, F2RL1, F2RL2, F2RL3, F3, F5, F7, F8, F9, F10, F11, F11R, F12, F13A1, F13B, FBLN1, FCER1G, FERMT3, FGA, FGB, FGG, FN1, FOXA2, FZD6, GAS6, GATA1, GNA13, GNAS, GP1BA, GP1BB, GP5, GP6, GP9, HGFAC, HNF4A, HPS1, HPS4, HPS5, HPS6, HPSE, HRG, HS3ST5, ITGA2B, ITGB3, KLKB1, KNG1, LNPK, LYN, LYST, MERTK, MMRN1, MPIG6B, NBEAL2, NFE2L2, P2RX1, P2RY1, P2RY12, PAPSS2, PDGFA, PDGFB, PDGFRA, PDPN, PEAR1, PF4, PIK3CB, PIP5K1C, PLA2G4A, PLAU, PLEK, PLG, PPIA, PRDX2, PRKCA, PRKCD, PRKCQ, PRKG1, PROC, PROCR, PROS1, PROZ, PRSS56, PSEN1, PSEN2, PTGER3, PTPN6, PTPRJ, RAB27A, RAP2B, S100A9, SELP, SERPINA10, SERPINC1, SERPIND1, SERPINE1, SERPINE2, SERPINF2, SERPING1, SH2B3, SHH, SLC4A1, SLC7A11, SRF, ST3GAL4, STXBP1, STXBP3, SYK, TBXA2R, TEC, TFPI, TFPI2, THBD, THBS1, TLR4, TPSAB1, TREML1, TSPAN8, TSPAN32, TXK, TYRO3, UBASH3A, UBASH3B, VKORC1, VPS33B, VWF, WNT3A
Platelet activation	ABAT, ADAMTS18, ADRA2A, ADRA2C, ADRB2, ALOX12, APOE, AXL, BLK, BLOC1S3, BLOC1S4, C1GALT1C1, C1QTNF1, CD9, CD40LG, CEACAM1, CELA2A, CFH, COMP, CX3CL1, DMTN, ENTPD1, ENTPD2, EVL, F2, F2R, F11R, FCER1G, FERMT3, FGA, FGB, FGG, FN1, FZD6, GATA1, GNA13, GNAS, GP1BA, GP5, HRG, ITGA2B, ITGB3, LYN, MERTK, P2RX1, P2RY1, P2RY12, PDGFA, PDGFB, PDGFRA, PDPN, PEAR1, PF4, PIK3CB, PIP5K1C,

	PLA2G4A, PLEK, PPIA, PRKCA, PRKCD, PRKCQ, PRKG1, PTGER3, PTPN6, PTPRJ, RAP2B, SELP, SERPINE2, SH2B3, SLC7A11, SRF, STXBP1, STXBP3, SYK, TEC, TLR4, TREML1, TSPAN32, TXK, TYRO3, UBASH3A, UBASH3B, VPS33B, VWF, WNT3A
Positive regulation of hemostasis	ANO6, APOH, CD36, CPB2, DMTN, ENPP4, F2, F2R, F7, F12, HPSE, HRG, NFE2L2, PLG, PRDX2, S100A9, SERPINE1, SERPINF2, ST3GAL4, TBXA2R, THBS1
OXPHOS	ACTN3, AK2, ATP5F1C, ATP5F1D, ATP7A, CHCHD10, COQ7, COQ9, COX10, COX15, COX4I1, COX5A, COX8A, CYCS, DLD, DNAJC15, FXN, GADD45GIP1, LEXM, MECP2, MLXIPL, MSH2, NDUFA1, NDUFA10, NDUFA2, NDUFA3, NDUFA4, NDUFA5, NDUFA6, NDUFA7, NDUFA8, NDUFA9, NDUFAB1, NDUFAF1, NDUFB1, NDUFB10, NDUFB2, NDUFB3, NDUFB4, NDUFB5, NDUFB6, NDUFB7, NDUFB8, NDUFB9, NDUFC1, NDUFC2, NDUFS1, NDUFS2, NDUFS3, NDUFS4, NDUFS5, NDUFS6, NDUFS7, NDUFS8, NDUFV1, NDUFV2, NDUFV3, NIPSNAP2 , PARK7, PINK1, PMPCB, PPIF, SDHAF2, SDHC, SLC25A23, SLC25A33, SNCA, SURF1, TAZ, UQCR10, UQCRB, UQCRC1, UQCRC2, UQCRH, UQCRHL, VCP
Glycolysis	ALDOA, ALDOC, ARNT, ENO1, ENO2, ENO3, ENTPD5, GAPDH, GAPDHS, GCK, GPI, HIF1A, HK1, HK2, HK3, HTR2A, INSR, MYC, P2RX7, PFKFB2, PFKFB3, PFKFB4, PFKL, PFKM, PFKP, PGAM1, PGK1, PPP2R5D, PRKAA1, TPI1
MHC class I	HLA-A, HLA-B, HLA-C, HLA-E, HLA-F
MHC class II	CD74, HLA-DMA, HLA-DMB, HLA-DOA, HLA-DOB, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DRA, HLA-DRB1, HLA-DRB5
Translation initiation	ABCE1, ABCF1, AGO2, ALKBH1, BANK1, BOLL, CTIF, DAZL, DDX3X, DENR, DHX29, DHX33, DNAJC3, EIF1, EIF1AD, EIF1AX, EIF1B, EIF2A, EIF2AK1, EIF2AK3, EIF2AK4, EIF2B1, EIF2B2, EIF2B3, EIF2B4, EIF2B5, EIF2D, EIF2S1, EIF2S2, EIF3A, EIF3B, EIF3C, EIF3D, EIF3E, EIF3F, EIF3G, EIF3H, EIF3I, EIF3K, EIF3L, EIF3M, EIF4A1, EIF4A2, EIF4B, EIF4E, EIF4E1B, EIF4E2, EIF4E3, EIF4EBP1, EIF4EBP2, EIF4EBP3, EIF4G1, EIF4G2, EIF4G3, EIF4H, EIF5, EIF5B, EIF6, FECH, FMR1, GLE1, HABP4, IMPACT, KHDRBS1, KLHL25, LARP1, MCTS1, METTL3, MIF4GD, MKNK1, MTIF2, MTIF3, NCBP1, NCBP2, NCK1, NCK2, NPM1, PAIP1, PAIP2, PAIP2B, POLR2D, POLR2G, PPP1R15A, PPP1R15B, RARA, RBM4,

	RPL13A, RPS2, RPS6KB1, RPS6KB2, RXRA, TMED2, TNF, TPR, UHMK1, YTHDF1, YTHDF2, YTHDF3
Response to type I IFN	ADAR, CACTIN, CDC37, CNOT7, DCST1, FADD, IFITM1, IFITM2, IFITM3, IFNAR1, IFNAR2, IKBKE, IRAK1, IRF3, IRF7, ISG15, LSM14A, MAVS, METTL3, MUL1, MYD88, NLRC5, OAS2, PTPN2, SAMHD1, SETD2, SHMT2, SMPD1, STAT1, STAT2, TBK1, TRIM6, TRIM56, TTLL12, UBE2K, WNT5A, YTHDF2, YTHDF3, ZBP1
Response to IFNγ	ACTG1, ACTR2, ACTR3, ARG1, BST2, CAPG, CASP1, CCL2, CCL3, CCL4, CCL5, CCL7, CCL20, CCL22, CCL26, CD40, CD47, CD74, CDC37, CDC42, CDC42EP2, CDC42EP4, CIITA, CITED1, CXCL16, CYP27B1, DAPK1, DAPK3, DNAJA3, EPRS, EVL, FLNB, GAPDH, GBP2, GBP3, GBP4, GBP5, GBP7, GCH1, GSN, IFITM1, IFITM2, IFITM3, IFNG, IL12RB1, IL23R, IRF1, IRF8, IRGM, JAK2, KIF5B, KIF16B, KYNU, MED1, MEFV, MRC1, MYO1C, MYO18A, NLRC5, NMI, PARP9, PARP14, PDE12, PPARG, PTPN2, RAB11FIP5, RAB12, RAB20, RAB43, RPL13A, RPS6KB1, SIRPA, SLC11A1, SLC26A6, SNCA, SOCS1, STAT1, STX4, STX8, STX11, STXBP1, STXBP2, STXBP3, STXBP4, SYNCRIP, TLR2, TLR4, TRIM21, TXK, VAMP3, VAMP4, VAMP8, VIM, VPS26B, WAS, XCL1 , ZYX
Response to IFN-β	AIM2, BST2, CDC34, DDX41, GBP2, GBP3, HTRA2, IFIT1, IFIT3, IFITM1, IFITM2, IFITM3, IFNAR2, IKBKE, IRF1, IRGM2, NDUFA13, PLSCR1, PNPT1, STAT1, TRIM6, UBE2G2, UBE2K, XAF1
Coronavirus disease - COVID-19	ACE, ACE2, ADAM17, ADAR, AGTR1, C1QA, C1QB, C1QC, C1R, C1S, C2, C3, C3AR1, C4A, C4B, C5, C5AR1, C6, C7, C8A, C8B, C8G, C9, CASP1, CCL2, CFB, CFD, CGAS, CHUK, CSF2, CSF3, CXCL10, CXCL8, CYBB, DDX58, EGFR, EIF2AK2, F13A1, F13B, F2, FAU, FCGR2A, FGA, FGB, FGG, FOS, HBEGF, IFIH1, IFNA1, IFNA10, IFNA13, IFNA14, IFNA16, IFNA17, IFNA2, IFNA21, IFNA4, IFNA5, IFNA6, IFNA7, IFNA8, IFNAR1, IFNAR2, IFNB1, IKBKB, IKBKE, IKBKG, IL12A, IL12B, IL1B, IL2, IL6, IL6R, IL6ST, IRAK1, IRAK4, IRF3, IRF9, ISG15, JAK1, JUN, MAP3K7, MAPK1, MAPK10, MAPK11, MAPK12, MAPK13, MAPK14, MAPK3, MAPK8, MAPK9, MAS1, MASP1, MASP2, MAVS, MBL2, MMP1, MMP3, MX1, MX2, MYD88, NFKB1, NFKBIA, NFKBIB, NLRP3, NRP1, OAS1, OAS2, OAS3, PIK3CA, PIK3CB, PIK3CD, PIK3R1, PIK3R2, PIK3R3, PLCG1, PLCG2, PRKCA, PRKCB, PRKCG, RELA, RPL10, RPL10A, RPL10L, RPL11, RPL12, RPL13, RPL13A, RPL14, RPL15, RPL17,

	RPL17-C18orf32, RPL18, RPL18A, RPL19, RPL21, RPL22, RPL22L1, RPL23, RPL23A, RPL24, RPL26, RPL26L1, RPL27, RPL27A, RPL28, RPL29, RPL3, RPL30, RPL31, RPL32, RPL34, RPL35, RPL35A, RPL36, RPL36A, RPL36A-HNRNPH2, RPL36AL, RPL37, RPL37A, RPL38, RPL39, RPL3L, RPL4, RPL41, RPL5, RPL6, RPL7, RPL7A, RPL8, RPL9, RPLP0, RPLP1, RPLP2, RPS10, RPS10-NUDT3, RPS11, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS17, RPS18, RPS19, RPS2, RPS20, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27, RPS27A, RPS27L, RPS28, RPS29, RPS3, RPS3A, RPS4X, RPS4Y1, RPS4Y2, RPS5, RPS6, RPS7, RPS8, RPS9, RPSA, RSL24D1, SELP, STAT1, STAT2, STAT3, SYK, TAB2, TBK1, TLR2, TLR3, TLR4, TLR7, TLR8, TMPRSS2, TNF, TNFRSF1A, TRAF3, TRAF6, TYK2, UBA52, VWF
HLA-DR	HLA-DRA, HLA-DRB1, HLA-DRB5

Table B.3 Count and frequency of each cluster of platelets composed of each disease severity.

Clusters	Severity	Count	Frequency
C0	CV	1107	0.082
C0	FT	2876	0.214
C0	HC	743	0.055
C0	MD	1027	0.076
C0	ML	1959	0.146
C0	SLE	53	0.004
C0	SV	5688	0.423
C1	CV	879	0.083
C1	FT	2101	0.2
C1	HC	608	0.058
C1	MD	949	0.09
C1	ML	1839	0.175
C1	SLE	51	0.005
C1	SV	4102	0.39
C2	CV	561	0.073
C2	FT	1377	0.179
C2	HC	305	0.04
C2	MD	722	0.094

C2	ML	1338	0.174
C2	SLE	17	0.002
C2	SV	3373	0.438
C3	CV	81	0.021
C3	FT	202	0.052
C3	HC	636	0.165
C3	MD	712	0.184
C3	ML	962	0.249
C3	SLE	4	0.001
C3	SV	1267	0.328
C4	CV	137	0.043
C4	FT	1224	0.385
C4	HC	139	0.044
C4	MD	127	0.04
C4	ML	264	0.083
C4	SLE	4	0.001
C4	SV	1284	0.404
C5	CV	155	0.055
C5	FT	520	0.185
C5	HC	472	0.167
C5	MD	310	0.11

C5	ML	263	0.093
C5	SLE	10	0.004
C5	SV	1088	0.386
C6	CV	168	0.072
C6	FT	262	0.112
C6	HC	18	0.008
C6	MD	293	0.125
C6	ML	375	0.16
C6	SLE	11	0.005
C6	SV	1211	0.518
C7	CV	125	0.081
C7	FT	211	0.137
C7	HC	65	0.042
C7	MD	150	0.097
C7	ML	212	0.138
C7	SLE	13	0.008
C7	SV	764	0.496
C8	CV	370	0.36
C8	FT	86	0.084
C8	HC	57	0.055
C8	MD	15	0.015

C8	ML	38	0.037
C8	SLE	0	0
C8	SV	462	0.449
C9	CV	28	0.047
C9	FT	263	0.442
C9	HC	102	0.171
C9	MD	9	0.015
C9	ML	37	0.062
C9	SLE	0	0
C9	SV	156	0.262
C10	CV	51	0.105
C10	FT	62	0.127
C10	HC	36	0.074
C10	MD	8	0.016
C10	ML	42	0.086
C10	SLE	6	0.012
C10	SV	282	0.579
C11	CV	6	0.021
C11	FT	221	0.784
C11	HC	3	0.011
C11	MD	0	0

C11	ML	10	0.035
C11	SLE	0	0
C11	SV	42	0.149
C12	CV	27	0.158
C12	FT	9	0.053
C12	HC	21	0.123
C12	MD	8	0.047
C12	ML	20	0.117
C12	SLE	0	0
C12	SV	86	0.503

Table B.4 Count and frequency of each cluster of platelets composed of each outcome.

Clusters	Outcomes	Count	Frequency
C0	HC	743	0.055
C0	FT	2876	0.214
C0	S	7135	0.53
C0	Unknown	2699	0.201
C1	HC	608	0.058
C1	FT	2101	0.2
C1	S	5527	0.525
C1	Unknown	2293	0.218
C2	HC	305	0.04
C2	FT	1377	0.179
C2	S	4325	0.562
C2	Unknown	1686	0.219
C3	HC	636	0.165
C3	FT	202	0.052
C3	S	2394	0.62
C3	Unknown	632	0.164
C4	HC	139	0.044
C4	FT	1224	0.385
C4	S	1106	0.348

C4	Unknown	710	0.223
C5	HC	472	0.167
C5	FT	520	0.185
C5	S	1190	0.422
C5	Unknown	636	0.226
C6	HC	18	0.008
C6	FT	262	0.112
C6	S	1675	0.716
C6	Unknown	383	0.164
C7	HC	65	0.042
C7	FT	211	0.137
C7	S	1020	0.662
C7	Unknown	244	0.158
C8	HC	57	0.055
C8	FT	86	0.084
C8	S	774	0.753
C8	Unknown	111	0.108
C9	HC	102	0.171
C9	FT	263	0.442
C9	S	156	0.262
C9	Unknown	74	0.124

C10	HC	36	0.074
C10	FT	62	0.127
C10	S	350	0.719
C10	Unknown	39	0.08
C11	HC	3	0.011
C11	FT	221	0.784
C11	S	45	0.16
C11	Unknown	13	0.046
C12	HC	21	0.123
C12	FT	9	0.053
C12	S	53	0.31
C12	Unknown	88	0.515

Table B.5 Up-regulated genes in each platelet cluster compared to other clusters of platelets.

Cluster	Gene	avg_log2FC	pct.1	pct.2	p_val	p_val_adj
C0	MYL9	1.4497984	0.93	0.68	0	0
C0	FCER1G	1.3785031	0.7	0.43	0	0
C0	PARVB	1.3686138	0.66	0.43	0	0
C1	TMEM140	1.0820013	0.63	0.38	0	0
C1	PNMA1	1.0578438	0.36	0.22	0	0
C1	CDKN1A	1.044135	0.61	0.41	0	0
C2	DEPP1	2.4646081	0.23	0.01	0	0
C2	INKA1	2.3932049	0.64	0.1	0	0
C2	FAM110A	2.2112608	0.81	0.23	0	0
C3	JUNB	3.2852118	0.81	0.14	0	0
C3	JUN	3.2788035	0.79	0.12	0	0
C3	RPL34	3.173778	0.98	0.32	0	0
C4	HPSE	1.464646	0.6	0.15	0	0
C4	WFDC1	1.4141081	0.35	0.06	0	0
C4	PF4	1.3198794	1	0.93	0	0
C5	JCHAIN	1.4655803	0.26	0.11	7.35E-164	9.60E-160
C5	TMSB10	0.3179517	0.65	0.49	7.48E-66	9.77E-62

C5	PSME2	0.2810193	0.08	0.1	1.21E-64	1.58E-60
C6	CRBN	2.5758489	0.72	0.23	0	0
C6	CD58	2.5490962	0.33	0.05	0	0
C6	SRSF3	2.5017011	0.64	0.2	0	0
C7	S100A8	2.772694	0.73	0.43	0	0
C7	MALAT1	2.7540619	0.97	0.66	0	0
C7	NEAT1	2.3456765	0.65	0.25	0	0
C8	RPLP0	3.0268134	0.97	0.26	0	0
C8	RPS6	2.8436011	0.97	0.27	0	0
C8	RPS23	2.8411861	0.99	0.32	0	0
C9	HBB	6.9482949	0.58	0.21	0	0
C9	HBA2	6.5330751	0.45	0.08	0	0
C9	HBA1	5.6765527	0.5	0.1	0	0
C10	CD74	3.8920131	0.99	0.3	0	0
C10	HLA-DRA	3.4150658	0.9	0.09	0	0
C10	CD79A	2.7783787	0.67	0.02	0	0
C11	TPT1	2.2400798	1	0.64	1.13E-210	1.48E-206
C11	POLR2L	2.3414138	0.9	0.14	7.60E-195	9.93E-191
C11	CSRP1	1.9426621	0.8	0.09	3.92E-171	5.12E-167

C12	MALAT1	2.7481118	1	0.66	1.09E-122	1.42E-118
C12	RPL3	2.4830703	0.97	0.29	1.54E-117	2.02E-113
C12	RPS16	2.2162231	0.98	0.26	7.54E-106	9.85E-102

Table B.6 Consistently up/down-regulated genes in COVID-19, similar symptom hospitalized (SSH), sepsis, and SLE patients compared to healthy controls (HC).

Overlapped up-regulated genes	AP2B1, ATP5F1E, ATP5MPL, BEX3, BRK1, C9orf16, CASP4, CD226, CYTOR, DNAJC15, DSTN, DYNLRB1, EIF1, FCER1G, GLA, GNG5, GPX4, GTF2A2, HSPB1, IFI27L2, IFITM2, IFITM3, ISCA1, KIFAP3, LCN2, MAPRE1, NAA38, NAP1L1, NDUFS5, OST4, RHEB, RHOC, S100A8, S100A9, SERF2, SMIM27, STMP1, TBC1D15, TIMP1, TMEM219, UBL5, UQCR11, XK, YWHAE, ZFAND3
Overlapped down-regulated genes	ANXA1, BTG1, BTG2, CABP5, CD3E, CD48, CD52, CD7, COMMD6, CORO1A, CRIP1, CXCR4, DDX5, DUSP1, EEF1A1, EEF1B2, EEF1D, EEF2, EIF2S3, EIF3E, EIF3F, EIF3H, EIF3K, EIF3L, ETS1, FAU, FOS, GIMAP7, GRHL1, GTPBP2, HNRNPA1, HNRNPDL, HSPA8, ICAM3, IDS, IER2, IL32, IL7R, JUN, JUNB, KLF2, LCK, LDHB, LEF1, LEPROTL1, LIMD2, LINC00861, LITAF, LSP1, LTB, MAL, MALAT1, MRFAP1, NACA, NOSIP, NPM1, ODC1, PABPC1, PFDN5, PTPRC, RACK1, RGCC, RPL10, RPL10A, RPL11, RPL12, RPL13, RPL14, RPL17, RPL18, RPL18A, RPL19, RPL22, RPL23A, RPL24, RPL26, RPL28, RPL29, RPL3, RPL30, RPL32, RPL34, RPL35, RPL36, RPL37, RPL39, RPL4, RPL5, RPL6, RPL7, RPL7A, RPL8, RPLP0, RPLP2, RPS12, RPS13, RPS14, RPS15, RPS15A, RPS16, RPS18, RPS19, RPS2, RPS21, RPS23, RPS24, RPS25, RPS26, RPS27A, RPS28, RPS3, RPS3A, RPS4X, RPS5, RPS6, RPS7, RPS8, RPSA, SARAF, SLC25A6, SNRPD2, SSR2, TCF7, TNS1, TOMM20, TOMM7, TSPAN18, TUBA1C, ZFP36L2

REFERENCES

1. Murphy D. Northern blotting. Methods Mol Biol. 1993;18:337-40.

2. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science. 1995;270(5235):467-70.

3. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57-63.

4. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008;5(7):621-8.

5. Bacher R, Kendziorski C. Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol. 2016;17:63.

6. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6(5):377-82.

7. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell. 2015;58(4):610-20.

8. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012;489(7416):391-9.

9. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017;65(4):631-43 e4.

10. Mereu E, Lafzi A, Moutinho C, Ziegenhain C, McCarthy DJ, Álvarez-Varela A, et al. Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. Nat Biotechnol. 2020;38(6):747-55.

11. Aliaghaei M, Haun JB. Optimization of Mechanical Tissue Dissociation Using an Integrated Microfluidic Device for Improved Generation of Single Cells Following Digestion. Front Bioeng Biotechnol. 2022;10:841046.

12. Reichard A, Asosingh K. Best Practices for Preparing a Single Cell Suspension from Solid Tissues for Flow Cytometry. Cytometry A. 2019;95(2):219-26.

13. Benck CJ, Martinov T, Fife BT, Chatterjea D. Isolation of Infiltrating Leukocytes from Mouse Skin Using Enzymatic Digest and Gradient Separation. J Vis Exp. 2016(107):e53638.

14. Perfetto SP, Chattopadhyay PK, Lamoreaux L, Nguyen R, Ambrozak D, Koup RA, et al. Amine-reactive dyes for dead cell discrimination in fixed samples. Curr Protoc Cytom. 2010;Chapter 9:Unit 9 34.

15. Menon V, Thomas R, Ghale AR, Reinhard C, Pruszak J. Flow cytometry protocols for surface and intracellular antigen analyses of neural cell types. J Vis Exp. 2014(94).

16. Krishnaswami SR, Grindberg RV, Novotny M, Venepally P, Lacar B, Bhutani K, et al. Using single nuclei for RNA-seq to capture the transcriptome of postmortem neurons. Nat Protoc. 2016;11(3):499-524.

17. McKinnon KM. Flow Cytometry: An Overview. Curr Protoc Immunol. 2018;120:5 1 -5 1 11.

18. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161(5):1202-14.

19. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187-201.

20. Rettig JR, Folch A. Large-scale single-cell trapping and imaging using microwell arrays. Anal Chem. 2005;77(17):5628-34.

21. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell. 2018;172(5):1091-107 e17.

22. Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, Perez Alea

M, et al. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. Proc Natl Acad Sci U S A. 2008;105(36):13427-32.

23. Yoshimoto N, Kida A, Jie X, Kurokawa M, Iijima M, Niimi T, et al. An automated system for high-throughput single cell-based breeding. Sci Rep. 2013;3:1191.

24. Kuppers R, Zhao M, Hansmann ML, Rajewsky K. Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. EMBO J. 1993;12(13):4955-67.

25. Suarez-Quian CA, Goldstein SR, Pohida T, Smith PD, Peterson JI, Wellner E, et al. Laser capture microdissection of single cells from complex tissues. Biotechniques. 1999;26(2):328-35.

26. Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012;30(8):777-82.

27. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012;2(3):666-73.

28. Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T, et al. Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol. 2013;14(4):R31.

29. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014;9(1):171-81.

30. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014;343(6172):776-9.

31. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.

32. Sheng K, Cao W, Niu Y, Deng Q, Zong C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. Nat Methods. 2017;14(3):267-70.

33. Trombetta JJ, Gennert D, Lu D, Satija R, Shalek AK, Regev A. Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. Curr Protoc Mol Biol. 2014;107:4 22 1-4 17.

34. Gierahn TM, Wadsworth MH, 2nd, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods. 2017;14(4):395-8.

35. Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science. 2018;360(6385):176-82.

36. Huang XT, Li X, Qin PZ, Zhu Y, Xu SN, Chen JP. Technical Advances in Single-Cell RNA Sequencing and Applications in Normal and Malignant Hematopoiesis. Front Oncol. 2018;8:582.

37. Aigrain L, Gu Y, Quail MA. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. BMC Genomics. 2016;17:458.

38. Cock PJ, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 2010;38(6):1767-71.

39. Zhou Q, Su X, Jing G, Chen S, Ning K. RNA-QC-chain: comprehensive and fast quality control for RNA-Seq data. BMC Genomics. 2018;19(1):144.

40. Perez-Rubio P, Lottaz C, Engelmann JC. FastqPuri: high-performance preprocessing of RNA-seq data. BMC Bioinformatics. 2019;20(1):226.

41. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

42. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37(8):907-15.

43. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525-7.

44. Shainer I, Stemmer M. Choice of pre-processing pipeline influences clustering quality of scRNA-seq datasets. BMC Genomics. 2021;22(1):661.

45. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30(7):923-30.

46. Anders S, Pyl PT, Huber W. HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166-9.

47. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017;27(3):491-9.

48. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC, et al. Classification of low quality cells from single-cell RNA-seq data. Genome Biol. 2016;17:29.

49. Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. Biostatistics. 2018;19(4):562-78.

50. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2014;11(2):163-6.

51. Chen W, Li Y, Easton J, Finkelstein D, Wu G, Chen X. UMI-count modeling and differential expression analysis for single-cell RNA sequencing. Genome Biol. 2018;19(1):70.

52. Altman N, Krzywinski M. Points of significance: Sources of variation. Nat Methods. 2015;12(1):5-6.

53. Zhao Y, Li MC, Konate MM, Chen L, Das B, Karlovich C, et al. TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository. J Transl Med. 2021;19(1):269.

54. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

55. Bacher R, Chu LF, Leng N, Gasch AP, Thomson JA, Stewart RM, et al. SCnorm: robust normalization of single-cell RNA-seq data. Nat Methods. 2017;14(6):584-6.

56. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018;36(5):411-20.

57. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20(1):296.

58. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell. 2018;174(3):716-29 e27.

59. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods. 2018;15(7):539-42.

60. Li WV, Li JJ. An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat Commun. 2018;9(1):997.

61. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet. 2019;10:317.

62. Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, Goh M, et al. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. 2020;21(1):12.

63. Ryu Y, Han GH, Jung E, Hwang D. Integration of Single-Cell RNA-Seq Datasets: A Review of Computational Methods. Mol Cells. 2023;46(2):106-19.

64. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019;177(7):1888-902 e21.

65. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast,

sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289-96.

66. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat Biotechnol. 2019;37(6):685-91.

67. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 2018;36(5):421-7.

68. Xiang R, Wang W, Yang L, Wang S, Xu C, Chen X. A Comparison for Dimensionality Reduction Methods of Single-Cell RNA-seq Data. Front Genet. 2021;12:646936.

69. Townes FW, Hicks SC, Aryee MJ, Irizarry RA. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. Genome Biol. 2019;20(1):295.

70. Kobak D, Berens P. The art of using t-SNE for single-cell transcriptomics. Nat Commun. 2019;10(1):5416.

71. Van der Maaten L HG. Visualizing data using t-SNE. Journal of machine learning research. 2008;9(11).

72. McInnes L HJ, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426. 2018.

73. Oyelade J, Isewon I, Oladipupo F, Aromolaran O, Uwoghiren E, Ameh F, et al. Clustering Algorithms: Their Application to Gene Expression Data. Bioinform Biol Insights. 2016;10:237-53.

74. Petegrosso R, Li Z, Kuang R. Machine learning and statistical methods for clustering single-cell RNA-sequencing data. Brief Bioinform. 2020;21(4):1209-23.

75. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163-72.

76. Shao X, Liao J, Lu X, Xue R, Ai N, Fan X. scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data. iScience. 2020;23(3):100882.

77. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-87.e29.

78. Zhang AW, O'Flanagan C, Chavez EA, Lim JLP, Ceglia N, McPherson A, et al. Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. Nat Methods. 2019;16(10):1007-15.

79. Liu J, Gao C, Sodicoff J, Kozareva V, Macosko EZ, Welch JD. Jointly defining cell types from multiple single-cell datasets using LIGER. Nat Protoc. 2020;15(11):3632-62.

80. Dominguez Conde C, Xu C, Jarvis LB, Rainbow DB, Wells SB, Gomes T, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022;376(6594):eabl5197.

81. Das S, Rai A, Rai SN. Differential Expression Analysis of Single-Cell RNA-Seq Data: Current Statistical Approaches and Outstanding Challenges. Entropy (Basel). 2022;24(7).

82. Wang T, Li B, Nelson CE, Nabavi S. Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data. BMC Bioinformatics. 2019;20(1):40.

83. Mooney MA, Wilmot B. Gene set analysis: A step-by-step guide. Am J Med Genet B Neuropsychiatr Genet. 2015;168(7):517-27.

84. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

85. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7.

86. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g:Profiler-a

web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016;44(W1):W83-9.

87. Franchini M, Pellecchia S, Viscido G, Gambardella G. Single-cell gene set enrichment analysis and transfer learning for functional annotation of scRNA-seq data. NAR Genom Bioinform. 2023;5(1):lqad024.

88. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021;2(3):100141.

89. Kumar MP, Du J, Lagoudas G, Jiao Y, Sawyer A, Drummond DC, et al. Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. Cell Rep. 2018;25(6):1458-68.e4.

90. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12(1):1088.

91. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020;15(4):1484-506.

92. Wang Y, Wang R, Zhang S, Song S, Jiang C, Han G, et al. iTALK: an R Package to characterize and illustrate intercellular communication. bioRxiv. 2019:507871.

93. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. 2020.

94. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159-62.

95. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381-6.

96. Bendall SC, Davis KL, Amir el AD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in

human B cell development. Cell. 2014;157(3):714-25.

97. Ranek JS, Stanley N, Purvis JE. Integrating temporal single-cell gene expression modalities for trajectory inference and disease prediction. Genome Biol. 2022;23(1):186.

98. Bar-Joseph Z, Gitter A, Simon I. Studying and modelling dynamic biological processes using time-series gene expression data. Nat Rev Genet. 2012;13(8):552-64.

99. Qiu X, Li J, Bonenfant J, Jaroszewski L, Mittal A, Klein W, et al. Dynamic changes in human single-cell transcriptional signatures during fatal sepsis. J Leukoc Biol. 2021;110(6):1253-68.

100. Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Brauninger M, et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. Proc Natl Acad Sci U S A. 2015;112(51):15672-7.

101. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979-82.

102. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci U S A. 2014;111(52):E5643-50.

103. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Gottgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019;20(1):59.

104. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;38(12):1408-14.

105. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. Science. 2006;311(5762):796-800.

106. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010;468(7326):911-20.

107. Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering.

Nat Methods. 2017;14(11):1083-6.

108. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS One. 2010;5(9).

109. Skok Gibbs C, Jackson CA, Saldi GA, Tjarnberg A, Shah A, Watters A, et al. High-performance single-cell gene regulatory network inference at scale: the Inferelator 3.0. Bioinformatics. 2022;38(9):2519-28.

110. Moerman T, Aibar Santos S, Bravo Gonzalez-Blas C, Simm J, Moreau Y, Aerts J, et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. Bioinformatics. 2019;35(12):2159-61.

111. Chan TE, Stumpf MPH, Babtie AC. Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst. 2017;5(3):251-67 e3.

112. Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics. 2017;33(15):2314-21.

113. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science. 2006;313(5786):504-7.

114. Kingma DP, Welling M. Auto-encoding variational bayes. arXiv preprint arXiv:13126114. 2013.

115. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun. 2019;10(1):390.

116. Zhao J, Wang N, Wang H, Zheng C, Su Y. SCDRHA: A scRNA-Seq Data Dimensionality Reduction Algorithm Based on Hierarchical Autoencoder. Front Genet. 2021;12:733906.

117. Svensson V, Gayoso A, Yosef N, Pachter L. Interpretable factor models of single-cell RNA-seq via variational autoencoders. Bioinformatics. 2020;36(11):3418-21.

118. Li X, Zhang S, Wong KC. Deep embedded clustering with multiple objectives on

scRNA-seq data. Brief Bioinform. 2021;22(5).

119. Tian T, Zhang J, Lin X, Wei Z, Hakonarson H. Model-based deep embedding for constrained clustering analysis of single cell RNA-seq data. Nat Commun. 2021;12(1):1873.

120. Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw. 2009;20(1):61-80.

121. Wang J, Ma A, Chang Y, Gong J, Jiang Y, Qi R, et al. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. Nat Commun. 2021;12(1):1882.

122. Xu C, Lopez R, Mehlman E, Regier J, Jordan MI, Yosef N. Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. Mol Syst Biol. 2021;17(1):e9620.

123. Shao X, Yang H, Zhuang X, Liao J, Yang P, Cheng J, et al. scDeepSort: a pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. Nucleic Acids Res. 2021;49(21):e122.

124. Zhou Y, Peng M, Yang B, Tong T, Zhang B, Tang N. scDLC: a deep learning framework to classify large sample single-cell RNA-seq data. BMC Genomics. 2022;23(1):504.

125. Yang F WW, Wang F, Fang Y, Tang D, Huang J, Lu H, Yao J. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nature Machine Intelligence. 2022;4(10):852-66.

126. Yang L HS, Carbonell J. A theory of transfer learning with applications to active learning. . Machine learning. 2013;90:161-89.

127. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Buttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol. 2022;40(1):121-30.

128. Chen J, Wang X, Ma A, Wang QE, Liu B, Li L, et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. Nat Commun.

2022;13(1):6494.

129. Seninge L, Anastopoulos I, Ding H, Stuart J. VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. Nat Commun. 2021;12(1):5684.

130. Chawla S, Rockstroh A, Lehman M, Ratther E, Jain A, Anand A, et al. Gene expression based inference of cancer drug sensitivity. Nat Commun. 2022;13(1):5680.

131. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 2019;37(7):773-82.

132. Pruszak J, Sonntag KC, Aung MH, Sanchez-Pernaute R, Isacson O. Markers and methods for cell sorting of human embryonic stem cell-derived neural cell populations. Stem Cells. 2007;25(9):2257-68.

133. Feyaerts D, Hedou J, Gillard J, Chen H, Tsai ES, Peterson LS, et al. Integrated plasma proteomic and single-cell immune signaling network signatures demarcate mild, moderate, and severe COVID-19. Cell Rep Med. 2022;3(7):100680.

134. You M, Chen L, Zhang D, Zhao P, Chen Z, Qin EQ, et al. Single-cell epigenomic landscape of peripheral immune cells reveals establishment of trained immunity in individuals convalescing from COVID-19. Nat Cell Biol. 2021;23(6):620-30.

135. Rao A, Barkley D, Franca GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596(7871):211-20.

136. Heimberger AB. Functional imaging of immune cell subpopulations in the tumor microenvironment: clinical implications. J Clin Invest. 2022;132(16).

137. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014;343(6166):84-7.

138. Pham T, Tyagi A, Wang YS, Guo J. Single-cell proteomic analysis. WIREs Mech Dis. 2021;13(1):e1503.

139. Petrosius V, Schoof EM. Recent advances in the field of single-cell proteomics. Transl Oncol. 2023;27:101556.

140. Bandura DR, Baranov VI, Ornatsky OI, Antonov A, Kinach R, Lou X, et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem. 2009;81(16):6813-22.

141. Hughes AJ, Spelke DP, Xu Z, Kang CC, Schaffer DV, Herr AE. Single-cell western blotting. Nat Methods. 2014;11(7):749-55.

142. Budnik B, Levy E, Harmange G, Slavov N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 2018;19(1):161.

143. Wimmers F, Donato M, Kuo A, Ashuach T, Gupta S, Li C, et al. The single-cell epigenomic and transcriptional landscape of immunity to influenza vaccination. Cell. 2021;184(15):3915-35 e21.

144. Muto Y, Wilson PC, Ledru N, Wu H, Dimke H, Waikar SS, et al. Single cell transcriptional and chromatin accessibility profiling redefine cellular heterogeneity in the adult human kidney. Nat Commun. 2021;12(1):2190.

145. Rotem A, Ram O, Shoresh N, Sperling RA, Goren A, Weitz DA, et al. Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol. 2015;33(11):1165-72.

146. Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, et al. Massively multiplex single-cell Hi-C. Nat Methods. 2017;14(3):263-6.

147. Xia C, Babcock HP, Moffitt JR, Zhuang X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. Sci Rep. 2019;9(1):7721.

148. Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat Neurosci. 2021;24(3):425-36.

149. Hernandez S, Lazcano R, Serrano A, Powell S, Kostousov L, Mehta J, et al.

Challenges and Opportunities for Immunoprofiling Using a Spatial High-Plex Technology: The NanoString GeoMx((R)) Digital Spatial Profiler. Front Oncol. 2022;12:890410.

150. Wen L, Fan Z, Mikulski Z, Ley K. Imaging of the immune system - towards a subcellular and molecular understanding. J Cell Sci. 2020;133(5).

151. Holcomb EA, Pearson AN, Jungles KM, Tate A, James J, Jiang L, et al. High-content CRISPR screening in tumor immunology. Front Immunol. 2022;13:1041451.

152. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018;50(8):1-14.

153. Lahnemann D, Koster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biol. 2020;21(1):31.

154. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas. Elife. 2017;6.

155. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol. 2018;18(1):35-45.

156. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science. 2017;356(6335).

157. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single-cell transcriptomics to explore the immune system in health and disease. Science. 2017;358(6359):58-63.

158. Gasteiger G, D'Osualdo A, Schubert DA, Weber A, Bruscia EM, Hartl D. Cellular Innate Immunity: An Old Game with New Players. J Innate Immun. 2017;9(2):111-25.

159. Dunkelberger JR, Song WC. Complement and its role in innate and adaptive immune responses. Cell Res. 2010;20(1):34-50.

160. Uthaisangsook S, Day NK, Bahna SL, Good RA, Haraguchi S. Innate immunity and its role against infections. Ann Allergy Asthma Immunol. 2002;88(3):253-64; quiz

65-6, 318.

161. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol. 2010;11(5):373-84.

162. Marshall JS, Warrington R, Watson W, Kim HL. An introduction to immunology and immunopathology. Allergy Asthma Clin Immunol. 2018;14(Suppl 2):49.

163. Jung D, Giallourakis C, Mostoslavsky R, Alt FW. Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. Annu Rev Immunol. 2006;24:541-70.

164. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302(5909):575-81.

165. Horwitz DA, Fahmy TM, Piccirillo CA, La Cava A. Rebalancing Immune Homeostasis to Treat Autoimmune Diseases. Trends Immunol. 2019;40(10):888-908.

166. Nagata S, Tanaka M. Programmed cell death and the immune system. Nat Rev Immunol. 2017;17(5):333-40.

167. Rosenblum MD, Remedios KA, Abbas AK. Mechanisms of human autoimmunity. J Clin Invest. 2015;125(6):2228-33.

168. Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. Mol Cell. 2014;54(5):716-27.

169. Chaplin DD. Overview of the immune response. J Allergy Clin Immunol. 2010;125(2 Suppl 2):S3-23.

170. Luckheeram RV, Zhou R, Verma AD, Xia B. CD4(+)T cells: differentiation and functions. Clin Dev Immunol. 2012;2012:925135.

171. Bjorklund AK, Forkel M, Picelli S, Konya V, Theorell J, Friberg D, et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat Immunol. 2016;17(4):451-60.

172. Yeo AT, Rawal S, Delcuze B, Christofides A, Atayde A, Strauss L, et al. Single-

cell RNA sequencing reveals evolution of immune landscape during glioblastoma progression. Nat Immunol. 2022;23(6):971-84.

173. Miragaia RJ, Gomes T, Chomka A, Jardine L, Riedel A, Hegazy AN, et al. Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. Immunity. 2019;50(2):493-504 e7.

174. Oh DY, Kwek SS, Raju SS, Li T, McCarthy E, Chow E, et al. Intratumoral CD4(+) T Cells Mediate Anti-tumor Cytotoxicity in Human Bladder Cancer. Cell. 2020;181(7):1612-25 e13.

175. Holla P, Dizon B, Ambegaonkar AA, Rogel N, Goldschmidt E, Boddapati AK, et al. Shared transcriptional profiles of atypical B cells suggest common drivers of expansion and function in malaria, HIV, and autoimmunity. Sci Adv. 2021;7(22).

176. Li Q, Cheng Z, Zhou L, Darmanis S, Neff NF, Okamoto J, et al. Developmental Heterogeneity of Microglia and Brain Myeloid Cells Revealed by Deep Single-Cell RNA Sequencing. Neuron. 2019;101(2):207-23 e10.

177. Villar J, Segura E. Decoding the Heterogeneity of Human Dendritic Cell Subsets. Trends Immunol. 2020;41(12):1062-71.

178. Qi F, Xu G, Liao X, Wang F, Yuan J, Wang H, et al. ScRNA-seq revealed the kinetic of nasopharyngeal immune responses in asymptomatic COVID-19 carriers. Cell Discov. 2021;7(1):56.

179. Zhang Y, Wang D, Peng M, Tang L, Ouyang J, Xiong F, et al. Single-cell RNA sequencing in cancer research. J Exp Clin Cancer Res. 2021;40(1):81.

180. Lee J, Geng S, Li S, Li L. Single Cell RNA-Seq and Machine Learning Reveal Novel Subpopulations in Low-Grade Inflammatory Monocytes With Unique Regulatory Circuits. Front Immunol. 2021;12:627036.

181. Sebastian A, Hum NR, McCool JL, Wilson SP, Murugesh DK, Martin KA, et al. Single-cell RNA-Seq reveals changes in immune landscape in post-traumatic osteoarthritis. Front Immunol. 2022;13:938075.

182. Fernandez-Garcia J, Franco F, Parik S, Altea-Manzano P, Pane AA, Broekaert D,

et al. CD8(+) T cell metabolic rewiring defined by scRNA-seq identifies a critical role of ASNS expression dynamics in T cell differentiation. Cell Rep. 2022;41(7):111639.

183. Scharer CD, Patterson DG, Mi T, Price MJ, Hicks SL, Boss JM. Antibody-secreting cell destiny emerges during the initial stages of B-cell activation. Nat Commun. 2020;11(1):3989.

184. O'Neill MB, Quach H, Pothlichet J, Aquino Y, Bisiaux A, Zidane N, et al. Single-Cell and Bulk RNA-Sequencing Reveal Differences in Monocyte Susceptibility to Influenza A Virus Infection Between Africans and Europeans. Front Immunol. 2021;12:768189.

185. Zhang JY, Wang XM, Xing X, Xu Z, Zhang C, Song JW, et al. Single-cell landscape of immunological responses in patients with COVID-19. Nat Immunol. 2020;21(9):1107-18.

186. Zander R, Khatun A, Kasmani MY, Chen Y, Cui W. Delineating the transcriptional landscape and clonal diversity of virus-specific CD4(+) T cells during chronic viral infection. Elife. 2022;11.

187. Vick SC, Frutoso M, Mair F, Konecny AJ, Greene E, Wolf CR, et al. A regulatory T cell signature distinguishes the immune landscape of COVID-19 patients from those with other respiratory infections. Sci Adv. 2021;7(46):eabj0274.

188. Perez RK, Gordon MG, Subramaniam M, Kim MC, Hartoularos GC, Targ S, et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. Science. 2022;376(6589):eabf1970.

189. Schafflick D, Xu CA, Hartlehnert M, Cole M, Schulte-Mecklenbeck A, Lautwein T, et al. Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. Nat Commun. 2020;11(1):247.

190. Wu X, Liu Y, Jin S, Wang M, Jiao Y, Yang B, et al. Single-cell sequencing of immune cells from anticitrullinated peptide antibody positive and negative rheumatoid arthritis. Nat Commun. 2021;12(1):4977.

191. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-

seq. Science. 2016;352(6282):189-96.

192. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018;174(5):1293-308 e36.

193. Ma RY, Black A, Qian BZ. Macrophage diversity in cancer revisited in the era of single-cell omics. Trends Immunol. 2022;43(7):546-63.

194. Sheih A, Voillet V, Hanafi LA, DeBerg HA, Yajima M, Hawkins R, et al. Clonal kinetics and single-cell transcriptional profiling of CAR-T cells in patients undergoing CD19 CAR-T immunotherapy. Nat Commun. 2020;11(1):219.

195. Guruprasad P, Lee YG, Kim KH, Ruella M. The current landscape of single-cell transcriptomics for cancer immunotherapy. J Exp Med. 2021;218(1).

196. Leader AM, Grout JA, Maier BB, Nabet BY, Park MD, Tabachnikova A, et al. Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. Cancer Cell. 2021;39(12):1594-609 e12.

197. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med. 2019;25(8):1251-9.

198. Hui Z, Zhang J, Ren Y, Li X, Yan C, Yu W, et al. Single-cell profiling of immune cells after neoadjuvant pembrolizumab and chemotherapy in IIIA non-small cell lung cancer (NSCLC). Cell Death Dis. 2022;13(7):607.

199. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell. 2018;175(4):984-97 e24.

200. Chen L, Deng H, Cui H, Fang J, Zuo Z, Deng J, et al. Inflammatory responses and inflammation-associated diseases in organs. Oncotarget. 2018;9(6):7204-18.

201. Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat Med. 2020;26(6):842-4.

202. Stephenson E, Reynolds G, Botting RA, Calero-Nieto FJ, Morgan MD, Tuong ZK, et al. Single-cell multi-omics analysis of the immune response in COVID-19. Nat Med. 2021;27(5):904-16.

203. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat Med. 2020;26(7):1070-6.

204. Kazer SW, Aicher TP, Muema DM, Carroll SL, Ordovas-Montanes J, Miao VN, et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. Nat Med. 2020;26(4):511-8.

205. Aldrich AL, Horn CM, Heim CE, Korshoj LE, Kielian T. Transcriptional Diversity and Niche-Specific Distribution of Leukocyte Populations during Staphylococcus aureus Craniotomy-Associated Biofilm Infection. J Immunol. 2021;206(4):751-65.

206. Patir A, Gossner A, Ramachandran P, Alves J, Freeman TC, Henderson NC, et al. Single-cell RNA-seq reveals CD16(-) monocytes as key regulators of human monocyte transcriptional response to Toxoplasma. Sci Rep. 2020;10(1):21047.

207. Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. Cell Discov. 2020;6:31.

208. Bhamidipati K, Silberstein JL, Chaichian Y, Baker MC, Lanz TV, Zia A, et al. CD52 Is Elevated on B cells of SLE Patients and Regulates B Cell Function. Front Immunol. 2020;11:626820.

209. Ramesh A, Schubert RD, Greenfield AL, Dandekar R, Loudermilk R, Sabatino JJ, Jr., et al. A pathogenic and clonally expanded B cell transcriptome in active multiple sclerosis. Proc Natl Acad Sci U S A. 2020;117(37):22932-43.

210. Stephenson W, Donlin LT, Butler A, Rozo C, Bracken B, Rashidfarrokhi A, et al. Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. Nat Commun. 2018;9(1):791.

211. Alivernini S, MacDonald L, Elmesmari A, Finlay S, Tolusso B, Gigante MR, et

al. Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. Nat Med. 2020;26(8):1295-306.

212. Jakel S, Agirre E, Mendanha Falcao A, van Bruggen D, Lee KW, Knuesel I, et al. Altered human oligodendrocyte heterogeneity in multiple sclerosis. Nature. 2019;566(7745):543-7.

213. Jaeger N, Gamini R, Cella M, Schettini JL, Bugatti M, Zhao S, et al. Single-cell analyses of Crohn's disease tissues reveal intestinal intraepithelial T cells heterogeneity and altered subset distributions. Nat Commun. 2021;12(1):1921.

214. Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. Science. 2022;376(6589):eabf3041.

215. Mathys H, Davila-Velderrain J, Peng Z, Gao F, Mohammadi S, Young JZ, et al. Single-cell transcriptomic analysis of Alzheimer's disease. Nature. 2019;570(7761):332-7.

216. Tiklova K, Nolbrant S, Fiorenzano A, Bjorklund AK, Sharma Y, Heuer A, et al. Single cell transcriptomics identifies stem cell-derived graft composition in a model of Parkinson's disease. Nat Commun. 2020;11(1):2434.

217. Fernandez DM, Rahman AH, Fernandez NF, Chudnovskiy A, Amir ED, Amadori L, et al. Single-cell immune landscape of human atherosclerotic plaques. Nat Med. 2019;25(10):1576-88.

218. Sauler M, McDonough JE, Adams TS, Kothapalli N, Barnthaler T, Werder RB, et al. Characterization of the COPD alveolar niche using single-cell RNA sequencing. Nat Commun. 2022;13(1):494.

219. Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med. 2018;24(7):986-93.

220. Hara T, Chanoch-Myers R, Mathewson ND, Myskiw C, Atta L, Bussema L, et al. Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. Cancer Cell. 2021;39(6):779-92 e11.

221. Xiao Y, Wang Z, Zhao M, Deng Y, Yang M, Su G, et al. Single-Cell Transcriptomics Revealed Subtype-Specific Tumor Immune Microenvironments in Human Glioblastomas. Front Immunol. 2022;13:914236.

222. Wu F, Fan J, He Y, Xiong A, Yu J, Li Y, et al. Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. Nat Commun. 2021;12(1):2540.

223. Zavidij O, Haradhvala NJ, Mouhieddine TH, Sklavenitis-Pistofidis R, Cai S, Reidy M, et al. Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. Nat Cancer. 2020;1(5):493-506.

224. Zhang L, Li Z, Skrzypczynska KM, Fang Q, Zhang W, O'Brien SA, et al. Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. Cell. 2020;181(2):442-59 e29.

225. He D, Wang D, Lu P, Yang N, Xue Z, Zhu X, et al. Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. Oncogene. 2021;40(2):355-68.

226. Chung W, Eum HH, Lee HO, Lee KM, Lee HB, Kim KT, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. Nat Commun. 2017;8:15081.

227. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018;24(7):978-85.

228. Zhang L, Yu X, Zheng L, Zhang Y, Li Y, Fang Q, et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. Nature. 2018;564(7735):268-72.

229. Sehgal K, Portell A, Ivanova EV, Lizotte PH, Mahadevan NR, Greene JR, et al. Dynamic single-cell RNA sequencing identifies immunotherapy persister cells following PD-1 blockade. J Clin Invest. 2021;131(2).

230. Liu B, Hu X, Feng K, Gao R, Xue Z, Zhang S, et al. Temporal single-cell tracing

reveals clonal revival and expansion of precursor exhausted T cells during anti-PD-1 therapy in lung cancer. Nat Cancer. 2022;3(1):108-21.

231. Zhang Y, Chen H, Mo H, Hu X, Gao R, Zhao Y, et al. Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. Cancer Cell. 2021;39(12):1578-93 e8.

232. He J, Xiong X, Yang H, Li D, Liu X, Li S, et al. Defined tumor antigen-specific T cells potentiate personalized TCR-T cell therapy and prediction of immunotherapy response. Cell Res. 2022;32(6):530-42.

233. Noe A, Cargill TN, Nielsen CM, Russell AJC, Barnes E. The Application of Single-Cell RNA Sequencing in Vaccinology. J Immunol Res. 2020;2020:8624963.

234. Garraud O, Cognasse F. Are Platelets Cells? And if Yes, are They Immune Cells? Front Immunol. 2015;6:70.

235. Wool GD, Miller JL. The Impact of COVID-19 Disease on Platelets and Coagulation. Pathobiology. 2021;88(1):15-27.

236. Vardon-Bounes F, Ruiz S, Gratacap MP, Garcia C, Payrastre B, Minville V. Platelets Are Critical Key Players in Sepsis. Int J Mol Sci. 2019;20(14).

237. Tiru B, DiNino EK, Orenstein A, Mailloux PT, Pesaturo A, Gupta A, et al. The Economic and Humanistic Burden of Severe Sepsis. Pharmacoeconomics. 2015;33(9):925-37.

238. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet. 2020;395(10219):200-11.

239. Aziz M, Jacob A, Yang WL, Matsuda A, Wang P. Current trends in inflammatory and immunomodulatory mediators in sepsis. J Leukoc Biol. 2013;93(3):329-42.

240. van der Poll T, van de Veerdonk FL, Scicluna BP, Netea MG. The immunopathology of sepsis and potential therapeutic targets. Nat Rev Immunol. 2017;17(7):407-20.

241. Pierrakos C, Velissaris D, Bisdorff M, Marshall JC, Vincent JL. Biomarkers of sepsis: time for a reappraisal. Crit Care. 2020;24(1):287.

242. Davenport EE, Burnham KL, Radhakrishnan J, Humburg P, Hutton P, Mills TC, et al. Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study. Lancet Respir Med. 2016;4(4):259-71.

243. Sweeney TE, Perumal TM, Henao R, Nichols M, Howrylak JA, Choi AM, et al. A community approach to mortality prediction in sepsis via gene expression analysis. Nat Commun. 2018;9(1):694.

244. Reyes M, Filbin MR, Bhattacharyya RP, Billman K, Eisenhaure T, Hung DT, et al. An immune-cell signature of bacterial sepsis. Nat Med. 2020;26(3):333-40.

245. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med. 2001;345(19):1368-77.

246. Kumar A, Roberts D, Wood KE, Light B, Parrillo JE, Sharma S, et al. Duration of hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock. Crit Care Med. 2006;34(6):1589-96.

247. Mouncey PR, Osborn TM, Power GS, Harrison DA, Sadique MZ, Grieve RD, et al. Trial of early, goal-directed resuscitation for septic shock. N Engl J Med. 2015;372(14):1301-11.

248. Peake SL, Delaney A, Bailey M, Bellomo R, Cameron PA, Cooper DJ, et al. Goal-directed resuscitation for patients with early septic shock. N Engl J Med. 2014;371(16):1496-506.

249. Pro CI, Yealy DM, Kellum JA, Huang DT, Barnato AE, Weissfeld LA, et al. A randomized trial of protocol-based care for early septic shock. N Engl J Med. 2014;370(18):1683-93.

250. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. 2019;47(D1):D721-D8.

251. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, et al. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell. 2018;173(5):1307.

252. Yuan H, Yan M, Zhang G, Liu W, Deng C, Liao G, et al. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res. 2019;47(D1):D900-D8.

253. Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 2015;16:278.

254. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16(5):284-7.

255. Kox WJ, Volk T, Kox SN, Volk HD. Immunomodulatory therapies in sepsis. Intensive Care Med. 2000;26 Suppl 1:S124-8.

256. Marimuthu R, Francis H, Dervish S, Li SCH, Medbury H, Williams H. Characterization of Human Monocyte Subsets by Whole Blood Flow Cytometry Analysis. J Vis Exp. 2018(140).

257. Slyker JA, Lohman-Payne B, John-Stewart GC, Dong T, Mbori-Ngacha D, Tapia K, et al. The impact of HIV-1 infection and exposure on natural killer (NK) cell phenotype in Kenyan infants during the first year of life. Front Immunol. 2012;3:399.

258. Medina F, Segundo C, Campos-Caro A, González-García I, Brieva JA. The heterogeneity shown by human plasma cells from tonsil, blood, and bone marrow reveals graded stages of increasing maturity, but local profiles of adhesion molecule expression. Blood. 2002;99(6):2154-61.

259. Middleton EA, Rowley JW, Campbell RA, Grissom CK, Brown SM, Beesley SJ, et al. Sepsis alters the transcriptional and translational landscape of human and murine platelets. Blood. 2019;134(12):911-23.

260. Frederick M, Skinner HD, Kazi SA, Sikora AG, Sandulache VC. High expression of oxidative phosphorylation genes predicts improved survival in squamous cell carcinomas of the head and neck and lung. Sci Rep. 2020;10(1):6380.

261. George MJ, Bynum J, Nair P, Cap AP, Wade CE, Cox CS, et al. Platelet

biomechanics, platelet bioenergetics, and applications to clinical practice and translational research. Platelets. 2018;29(5):431-9.

262. Chapman LM, Aggrey AA, Field DJ, Srivastava K, Ture S, Yui K, et al. Platelets present antigen in the context of MHC class I. J Immunol. 2012;189(2):916-23.

263. Jiang H, Lin JJ, Tao J, Fisher PB. Suppression of human ribosomal protein L23A expression during cell growth inhibition by interferon-beta. Oncogene. 1997;14(4):473-80.

264. Bernardes JP, Mishra N, Tran F, Bahmer T, Best L, Blase JI, et al. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. Immunity. 2020;53(6):1296-314.e9.

265. Manne BK, Denorme F, Middleton EA, Portier I, Rowley JW, Stubben C, et al. Platelet gene expression and function in patients with COVID-19. Blood. 2020;136(11):1317-29.

266. Paulson RF, Shi L, Wu DC. Stress erythropoiesis: new signals and new stress progenitor cells. Curr Opin Hematol. 2011;18(3):139-45.

267. Bapat A, Schippel N, Shi X, Jasbi P, Gu H, Kala M, et al. Hypoxia promotes erythroid differentiation through the development of progenitors and proerythroblasts. Exp Hematol. 2021;97:32-46.e35.

268. Rogers HM, Yu X, Wen J, Smith R, Fibach E, Noguchi CT. Hypoxia alters progression of the erythroid program. Exp Hematol. 2008;36(1):17-27.

269. Hotchkiss RS, Monneret G, Payen D. Immunosuppression in sepsis: a novel understanding of the disorder and a new therapeutic approach. Lancet Infect Dis. 2013;13(3):260-8.

270. Gomez HG, Gonzalez SM, Londono JM, Hoyos NA, Nino CD, Leon AL, et al. Immunological characterization of compensatory anti-inflammatory response syndrome in patients with severe sepsis: a longitudinal study*. Crit Care Med. 2014;42(4):771-80.

271. Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. Nat Rev Immunol. 2011;11(11):762-74.

272. Zhou Z, Ren L, Zhang L, Zhong J, Xiao Y, Jia Z, et al. Heightened Innate Immune Responses in the Respiratory Tract of COVID-19 Patients. Cell Host Microbe. 2020;27(6):883-90.e2.

273. Camp SM, Ceco E, Evenoski CL, Danilov SM, Zhou T, Chiang ET, et al. Unique Toll-Like Receptor 4 Activation by NAMPT/PBEF Induces NFκB Signaling and Inflammatory Lung Injury. Sci Rep. 2015;5:13135.

274. Ye SQ, Simon BA, Maloney JP, Zambelli-Weiner A, Gao L, Grant A, et al. Pre-B-cell colony-enhancing factor as a potential novel biomarker in acute lung injury. Am J Respir Crit Care Med. 2005;171(4):361-70.

275. Gesing J, Scheuermann K, Wagner IV, Löffler D, Friebe D, Kiess W, et al. NAMPT serum levels are selectively elevated in acute infectious disease and in acute relapse of chronic inflammatory diseases in children. PLoS One. 2017;12(8):e0183027.

276. Karampela I, Christodoulatos GS, Kandri E, Antonakos G, Vogiatzakis E, Dimopoulos G, et al. Circulating eNampt and resistin as a proinflammatory duet predicting independently mortality in critically ill patients with sepsis: A prospective observational study. Cytokine. 2019;119:62-70.

277. Quijada H, Bermudez T, Kempf CL, Valera DG, Garcia AN, Camp SM, et al. Endothelial eNAMPT amplifies pre-clinical acute lung injury: efficacy of an eNAMPT-neutralising monoclonal antibody. Eur Respir J. 2021;57(5).

278. Schenck EJ, Ma KC, Price DR, Nicholson T, Oromendia C, Gentzler ER, et al. Circulating cell death biomarker TRAIL is associated with increased organ dysfunction in sepsis. JCI Insight. 2019;4(9).

279. Cziupka K, Busemann A, Partecke LI, Pötschke C, Rath M, Traeger T, et al. Tumor necrosis factor-related apoptosis-inducing ligand (TRAIL) improves the innate immune response and enhances survival in murine polymicrobial sepsis. Crit Care Med. 2010;38(11):2169-74.

280. Yan M, Marsters SA, Grewal IS, Wang H, Ashkenazi A, Dixit VM. Identification of a receptor for BLyS demonstrates a crucial role in humoral immunity. Nat Immunol. 2000;1(1):37-41.

281. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, et al. Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science. 2020;369(6504):718-24.

282. Cheron A, Floccard B, Allaouchiche B, Guignant C, Poitevin F, Malcus C, et al. Lack of recovery in monocyte human leukocyte antigen-DR expression is independently associated with the development of sepsis after major trauma. Crit Care. 2010;14(6):R208.

283. McDunn JE, Turnbull IR, Polpitiya AD, Tong A, MacMillan SK, Osborne DF, et al. Splenic CD4+ T cells have a distinct transcriptional response six hours after the onset of sepsis. J Am Coll Surg. 2006;203(3):365-75.

284. Hotchkiss RS, Moldawer LL, Opal SM, Reinhart K, Turnbull IR, Vincent JL. Sepsis and septic shock. Nat Rev Dis Primers. 2016;2:16045.

285. Svensson V, da Veiga Beltrame E, Pachter L. A curated database reveals trends in single-cell transcriptomics. Database (Oxford). 2020;2020.

286. Jiang Y, Rosborough BR, Chen J, Das S, Kitsios GD, McVerry BJ, et al. Single cell RNA sequencing identifies an early monocyte gene signature in acute respiratory distress syndrome. JCI Insight. 2020;5(13).

287. Darden DB, Bacher R, Brusko MA, Knight P, Hawkins RB, Cox MC, et al. Single-Cell RNA-seq of Human Myeloid-Derived Suppressor Cells in Late Sepsis Reveals Multiple Subsets With Unique Transcriptional Responses: A Pilot Study. Shock. 2021;55(5):587-95.

288. Shen XF, Cao K, Jiang JP, Guan WX, Du JF. Neutrophil dysregulation during sepsis: an overview and update. J Cell Mol Med. 2017;21(9):1687-97.

289. Demaret J, Venet F, Friggeri A, Cazalis MA, Plassais J, Jallades L, et al. Marked alterations of neutrophil functions during sepsis-induced immunosuppression. J Leukoc Biol. 2015;98(6):1081-90.

290. Holmsen H, Kaplan KL, Dangelmaier CA. Differential energy requirements for platelet responses. A simultaneous study of aggregation, three secretory processes,

arachidonate liberation, phosphatidylinositol breakdown and phosphatidate production. Biochem J. 1982;208(1):9-18.

291. Tsai MJ, Shih CJ, Chen YT. Association of prior antiplatelet agents with mortality in sepsis patients. Intensive Care Med. 2016;42(4):605-7.

292. Akinosoglou K, Alexopoulos D. Use of antiplatelet agents in sepsis: a glimpse into the future. Thromb Res. 2014;133(2):131-8.

293. Osborne TF, Veigulis ZP, Arreola DM, Mahajan SM, Röösli E, Curtin CM. Association of mortality and aspirin prescription for COVID-19 patients at the Veterans Health Administration. PLoS One. 2021;16(2):e0246825.

294. Dubois C, Marcé D, Faivre V, Lukaszewicz AC, Junot C, Fenaille F, et al. High plasma level of S100A8/S100A9 and S100A12 at admission indicates a higher risk of death in septic shock patients. Sci Rep. 2019;9(1):15660.

295. Mello FV, Land MGP, Costa ES, Teodósio C, Sanchez ML, Bárcena P, et al. Maturation-associated gene expression profiles during normal human bone marrow erythropoiesis. Cell Death Discov. 2019;5:69.

296. Treumann A, Lifely MR, Schneider P, Ferguson MA. Primary structure of CD52. J Biol Chem. 1995;270(11):6088-99.

297. Evan JR, Bozkurt SB, Thomas NC, Bagnato F. Alemtuzumab for the treatment of multiple sclerosis. Expert Opin Biol Ther. 2018;18(3):323-34.

298. Zhang X, Tao Y, Chopra M, Ahn M, Marcus KL, Choudhary N, et al. Differential reconstitution of T cell subsets following immunodepleting treatment with alemtuzumab (anti-CD52 monoclonal antibody) in patients with relapsing-remitting multiple sclerosis. J Immunol. 2013;191(12):5867-74.

299. Wang J, Zhang G, Sui Y, Yang Z, Chu Y, Tang H, et al. CD52 Is a Prognostic Biomarker and Associated With Tumor Microenvironment in Breast Cancer. Front Genet. 2020;11:578002.

300. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One. 2011;6(7):e21800.

301. Blighe K RS, Lewis M EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1100, https://githubcom/kevinblighe/EnhancedVolcano. 2021.

302. Koupenova M, Clancy L, Corkrey HA, Freedman JE. Circulating Platelets as Mediators of Immunity, Inflammation, and Thrombosis. Circ Res. 2018;122(2):337-51.

303. Maouia A, Rebetz J, Kapur R, Semple JW. The Immune Nature of Platelets Revisited. Transfus Med Rev. 2020;34(4):209-20.

304. Jenne CN, Urrutia R, Kubes P. Platelets: bridging hemostasis, inflammation, and immunity. Int J Lab Hematol. 2013;35(3):254-61.

305. Jin R, Yu S, Song Z, Zhu X, Wang C, Yan J, et al. Soluble CD40 ligand stimulates CD40-dependent activation of the β2 integrin Mac-1 and protein kinase C zeda (PKCζ) in neutrophils: implications for neutrophil-platelet interactions and neutrophil oxidative burst. PLoS One. 2013;8(6):e64631.

306. Gudbrandsdottir S, Hasselbalch HC, Nielsen CH. Activated platelets enhance IL-10 secretion and reduce TNF-α secretion by monocytes. J Immunol. 2013;191(8):4059-67.

307. Ali RA, Wuescher LM, Worth RG. Platelets: essential components of the immune system. Curr Trends Immunol. 2015;16:65-78.

308. Ammon C, Kreutz M, Rehli M, Krause SW, Andreesen R. Platelets induce monocyte differentiation in serum-free coculture. J Leukoc Biol. 1998;63(4):469-76.

309. Chatterjee M, von Ungern-Sternberg SN, Seizer P, Schlegel F, Büttcher M, Sindhu NA, et al. Platelet-derived CXCL12 regulates monocyte function, survival, differentiation into macrophages and foam cells through differential involvement of CXCR4-CXCR7. Cell Death Dis. 2015;6:e1989.

310. Portier I, Campbell RA. Role of Platelets in Detection and Regulation of Infection. Arterioscler Thromb Vasc Biol. 2021;41(1):70-8.

311. Xiang B, Zhang G, Guo L, Li XA, Morris AJ, Daugherty A, et al. Platelets protect

from septic shock by inhibiting macrophage-dependent inflammation via the cyclooxygenase 1 signalling pathway. Nat Commun. 2013;4:2657.

312. Duffau P, Seneschal J, Nicco C, Richez C, Lazaro E, Douchet I, et al. Platelet CD154 potentiates interferon-alpha secretion by plasmacytoid dendritic cells in systemic lupus erythematosus. Sci Transl Med. 2010;2(47):47ra63.

313. Gorog DA, Storey RF, Gurbel PA, Tantry US, Berger JS, Chan MY, et al. Current and novel biomarkers of thrombotic risk in COVID-19: a Consensus Statement from the International COVID-19 Thrombosis Biomarkers Colloquium. Nat Rev Cardiol. 2022;19(7):475-95.

314. Assinger A, Schrottmaier WC, Salzmann M, Rayes J. Platelets in Sepsis: An Update on Experimental Models and Clinical Data. Front Immunol. 2019;10:1687.

315. Linge P, Fortin PR, Lood C, Bengtsson AA, Boilard E. The non-haemostatic role of platelets in systemic lupus erythematosus. Nat Rev Rheumatol. 2018;14(4):195-213.

316. Wang C, Deng R, Gou L, Fu Z, Zhang X, Shao F, et al. Preliminary study to identify severe from moderate cases of COVID-19 using combined hematology parameters. Ann Transl Med. 2020;8(9):593.

317. Shen Y, Huang X, Zhang W. Platelet-to-lymphocyte ratio as a prognostic predictor of mortality for sepsis: interaction effect with disease severity-a retrospective study. BMJ Open. 2019;9(1):e022896.

318. Wu Y, Chen Y, Yang X, Chen L, Yang Y. Neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) were associated with disease activity in patients with systemic lupus erythematosus. Int Immunopharmacol. 2016;36:94-9.

319. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, et al. Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. Sci Immunol. 2020;5(49).

320. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, et al. Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell. 2020;182(6):1419-40.e23.

321. Arunachalam PS, Wimmers F, Mok CKP, Perera R, Scott M, Hagan T, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science. 2020;369(6508):1210-20.

322. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. Cell. 2021;184(7):1895-913.e19.

323. Liu C, Martins AJ, Lau WW, Rachmaninoff N, Chen J, Imberti L, et al. Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. Cell. 2021;184(7):1836-57.e22.

324. Combes AJ, Courau T, Kuhn NF, Hu KH, Ray A, Chen WS, et al. Global absence and targeting of protective immune states in severe COVID-19. Nature. 2021;591(7848):124-30.

325. Mistry P, Nakabo S, O'Neil L, Goel RR, Jiang K, Carmona-Rivera C, et al. Transcriptomic, epigenetic, and functional analyses implicate neutrophil diversity in the pathogenesis of systemic lupus erythematosus. Proc Natl Acad Sci U S A. 2019;116(50):25222-8.

326. Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nature immunology. 2019;20(2):163-72.

327. Oh S AJ, Al-Dulaimi R et al. HGNChelper: identification and correction of invalid gene symbols for human and mouse. F1000Research 2020; 2020.

328. Carlson M. GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 382. 2019.

329. Yu G. Using meshes for MeSH term enrichment and semantic analyses. Bioinformatics. 2018;34(21):3766-7.

330. Yu G, Wang LG, Yan GR, He QY. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. Bioinformatics. 2015;31(4):608-9.

331. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst.

2015;1(6):417-25.

332. Nakazato T, Bono H, Matsuda H, Takagi T. Gendoo: functional profiling of gene and disease features using MeSH vocabulary. Nucleic Acids Res. 2009;37(Web Server issue):W166-9.

333. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nature methods. 2017;14(10):979.

334. Horton LE, Cross RW, Hartnett JN, Engel EJ, Sakabe S, Goba A, et al. Endotheliopathy and Platelet Dysfunction as Hallmarks of Fatal Lassa Fever. Emerg Infect Dis. 2020;26(11):2625-37.

335. Laursen MA, Larsen JB, Hvas AM. Platelet function in disseminated intravascular coagulation: A systematic review. Platelets. 2018;29(3):238-48.

336. Popescu NI, Lupu C, Lupu F. Disseminated intravascular coagulation and its immune mechanisms. Blood. 2022;139(13):1973-86.

337. Coenen DM, Mastenbroek TG, Cosemans JMEM. Platelet interaction with activated endothelium: mechanistic insights from microfluidics. Blood. 2017;130(26):2819-28.

338. Boral BM, Williams DJ, Boral LI. Disseminated Intravascular Coagulation. Am J Clin Pathol. 2016;146(6):670-80.

339. Kim J, Lee JH, Iyer VR. Global identification of Myc target genes reveals its direct role in mitochondrial biogenesis and its E-box usage in vivo. PLoS One. 2008;3(3):e1798.

340. Gu SX, Tyagi T, Jain K, Gu VW, Lee SH, Hwa JM, et al. Thrombocytopathy and endotheliopathy: crucial contributors to COVID-19 thromboinflammation. Nat Rev Cardiol. 2021;18(3):194-209.

341. Akkerman JW, Holmsen H. Interrelationships among platelet responses: studies on the burst in proton liberation, lactate production, and oxygen uptake during platelet aggregation and Ca2+ secretion. Blood. 1981;57(5):956-66.

342. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019;37(5):547-54.

343. Silva-Filho JL, Caruso-Neves C, Pinheiro AAS. IL-4: an important cytokine in determining the fate of T cells. Biophys Rev. 2014;6(1):111-8.

344. Xu E, Xie Y, Al-Aly Z. Long-term neurologic outcomes of COVID-19. Nat Med. 2022;28(11):2406-15.

345. Banerjee M, Huang Y, Joshi S, Popa GJ, Mendenhall MD, Wang QJ, et al. Platelets Endocytose Viral Particles and Are Activated via TLR (Toll-Like Receptor) Signaling. Arterioscler Thromb Vasc Biol. 2020;40(7):1635-50.

346. Guo L, Shen S, Rowley JW, Tolley ND, Jia W, Manne BK, et al. Platelet MHC class I mediates CD8+ T-cell suppression during sepsis. Blood. 2021;138(5):401-16.

347. Castro F, Cardoso AP, Gonçalves RM, Serre K, Oliveira MJ. Interferon-Gamma at the Crossroads of Tumor Immune Surveillance or Evasion. Front Immunol. 2018;9:847.

348. Germann T, Gately MK, Schoenhaut DS, Lohoff M, Mattner F, Fischer S, et al. Interleukin-12/T cell stimulating factor, a cytokine with multiple effects on T helper type 1 (Th1) but not on Th2 cells. Eur J Immunol. 1993;23(8):1762-70.

349. Liang C, Bencurova E, Psota E, Neurgaonkar P, Prelog M, Scheller C, et al. Population-Predicted MHC Class II Epitope Presentation of SARS-CoV-2 Structural Proteins Correlates to the Case Fatality Rates of COVID-19 in Different Countries. Int J Mol Sci. 2021;22(5).

350. Litvinov RI, Evtugina NG, Peshkova AD, Safiullina SI, Andrianova IA, Khabirova AI, et al. Altered platelet and coagulation function in moderate-to-severe COVID-19. Sci Rep. 2021;11(1):16290.

351. Taylor CT, Doherty G, Fallon PG, Cummins EP. Hypoxia-dependent regulation of inflammatory pathways in immune cells. J Clin Invest. 2016;126(10):3716-24.

352. Galbraith MD, Kinning KT, Sullivan KD, Araya P, Smith KP, Granrath RE, et al. Specialized interferon action in COVID-19. Proc Natl Acad Sci U S A. 2022;119(11).

353. de Lusignan S, Hoang U, Liyanage H, Tripathy M, Yonova I, Byford R, et al. Integrating molecular point-of-care testing for influenza into primary care: a mixed-methods feasibility study. Br J Gen Pract. 2020;70(697):e555-e62.

354. Galani IE, Rovina N, Lampropoulou V, Triantafyllia V, Manioudaki M, Pavlos E, et al. Untuned antiviral immunity in COVID-19 revealed by temporal type I/III interferon patterns and flu comparison. Nat Immunol. 2021;22(1):32-40.

355. Fantuzzi L, Tagliamonte M, Gauzzi MC, Lopalco L. Dual CCR5/CCR2 targeting: opportunities for the cure of complex disorders. Cell Mol Life Sci. 2019;76(24):4869-86.

356. Murphy PM. Viral exploitation and subversion of the immune system through chemokine mimicry. Nat Immunol. 2001;2(2):116-22.

357. Turner JE, Paust HJ, Bennstein SB, Bramke P, Krebs C, Steinmetz OM, et al. Protective role for CCR5 in murine lupus nephritis. Am J Physiol Renal Physiol. 2012;302(11):F1503-15.

358. Spriggs MK, Armitage RJ, Strockbine L, Clifford KN, Macduff BM, Sato TA, et al. Recombinant human CD40 ligand stimulates B cell proliferation and immunoglobulin E secretion. J Exp Med. 1992;176(6):1543-50.

359. Rahman M, Roller J, Zhang S, Syk I, Menger MD, Jeppsson B, et al. Metalloproteinases regulate CD40L shedding from platelets and pulmonary recruitment of neutrophils in abdominal sepsis. Inflamm Res. 2012;61(6):571-9.

360. Songdej N, Rao AK. Hematopoietic transcription factor mutations: important players in inherited platelet defects. Blood. 2017;129(21):2873-81.

361. Eustes AS, Campbell RA, Middleton EA, Tolley ND, Manne BK, Montenont E, et al. Heparanase expression and activity are increased in platelets during clinical sepsis. J Thromb Haemost. 2021;19(5):1319-30.

362. Buijsers B, Yanginlar C, de Nooijer A, Grondman I, Maciej-Hulme ML, Jonkman I, et al. Increased Plasma Heparanase Activity in COVID-19 Patients. Front Immunol. 2020;11:575047.

363. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA. 2020;323(11):1061-9.

364. McAbee GN, Brosgol Y, Pavlakis S, Agha R, Gaffoor M. Encephalitis Associated with COVID-19 Infection in an 11-Year-Old Child. Pediatr Neurol. 2020;109:94.

365. Ferini-Strambi L, Salsone M. COVID-19 and neurological disorders: are neurodegenerative or neuroimmunological diseases more vulnerable? J Neurol. 2021;268(2):409-19.

366. Al-Sarraj S, Troakes C, Hanley B, Osborn M, Richardson MP, Hotopf M, et al. Invited Review: The spectrum of neuropathology in COVID-19. Neuropathol Appl Neurobiol. 2021;47(1):3-16.

367. Widmann CN, Heneka MT. Long-term cerebral consequences of sepsis. Lancet Neurol. 2014;13(6):630-6.

368. De Oliveira DMP, Bohlmann L, Conroy T, Jen FE, Everest-Dass A, Hansford KA, et al. Repurposing a neurodegenerative disease drug to treat Gram-negative antibiotic-resistant bacterial sepsis. Sci Transl Med. 2020;12(570).

369. McGlasson S, Wiseman S, Wardlaw J, Dhaun N, Hunt DPJ. Neurological Disease in Lupus: Toward a Personalized Medicine Approach. Front Immunol. 2018;9:1146.

370. Kopeikina E, Ponomarev ED. The Role of Platelets in the Stimulation of Neuronal Synaptic Plasticity, Electric Activity, and Oxidative Phosphorylation: Possibilities for New Therapy of Neurodegenerative Diseases. Front Cell Neurosci. 2021;15:680126.

371. Ferrer-Raventos P, Beyer K. Alternative platelet activation pathways and their role in neurodegenerative diseases. Neurobiol Dis. 2021;159:105512.

372. Espinosa-Parrilla Y, Gonzalez-Billault C, Fuentes E, Palomo I, Alarcon M. Decoding the Role of Platelets and Related MicroRNAs in Aging and Neurodegenerative Disorders. Front Aging Neurosci. 2019;11:151.

373. Wojsiat J, Laskowska-Kaszub K, Mietelska-Porowska A, Wojda U. Search for

Alzheimer's disease biomarkers in blood cells: hypotheses-driven approach. Biomark Med. 2017;11(10):917-31.

374. Goshua G, Pine AB, Meizlish ML, Chang CH, Zhang H, Bahel P, et al. Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study. Lancet Haematol. 2020;7(8):e575-e82.

375. Barrett TJ, Cornwell M, Myndzar K, Rolling CC, Xia Y, Drenkova K, et al. Platelets amplify endotheliopathy in COVID-19. Sci Adv. 2021;7(37):eabh2434.

376. Chang JC. Sepsis and septic shock: endothelial molecular pathogenesis associated with vascular microthrombotic disease. Thromb J. 2019;17:10.

377. Rondina MT, Carlisle M, Fraughton T, Brown SM, Miller RR, Harris ES, et al. Platelet-monocyte aggregate formation and mortality risk in older patients with severe sepsis and septic shock. J Gerontol A Biol Sci Med Sci. 2015;70(2):225-31.

378. DeMarini DM. Genotoxicity of tobacco smoke and tobacco smoke condensate: a review. Mutat Res. 2004;567(2-3):447-74.

379. Yamaguchi NH. Smoking, immunity, and DNA damage. Transl Lung Cancer Res. 2019;8(Suppl 1):S3-S6.

380. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst. 1999;91(14):1194-210.

381. Qiu F, Liang CL, Liu H, Zeng YQ, Hou S, Huang S, et al. Impacts of cigarette smoking on immune responsiveness: Up and down or upside down? Oncotarget. 2017;8(1):268-84.

382. Papoutsopoulou S, Satsangi J, Campbell BJ, Probert CS. Review article: impact of cigarette smoking on intestinal inflammation-direct and indirect mechanisms. Aliment Pharmacol Ther. 2020;51(12):1268-85.

383. Zuo L, He F, Sergakis GG, Koozehchian MS, Stimpfl JN, Rong Y, et al. Interrelated role of cigarette smoking, oxidative stress, and immune response in COPD and corresponding treatments. Am J Physiol Lung Cell Mol Physiol. 2014;307(3):L205-18.

384. Strzelak A, Ratajczak A, Adamiec A, Feleszko W. Tobacco Smoke Induces and Alters Immune Responses in the Lung Triggering Inflammation, Allergy, Asthma and Other Lung Diseases: A Mechanistic Review. Int J Environ Res Public Health. 2018;15(5).

385. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014;14(8):535-46.

386. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. Methods Mol Biol. 2018;1711:243-59.

387. Koks G, Uudelepp ML, Limbach M, Peterson P, Reimann E, Koks S. Smoking-induced expression of the GPR15 gene indicates its potential role in chronic inflammatory pathologies. Am J Pathol. 2015;185(11):2898-906.

388. Wang C, Yu Q, Song T, Wang Z, Song L, Yang Y, et al. The heterogeneous immune landscape between lung adenocarcinoma and squamous carcinoma revealed by single-cell RNA sequencing. Signal Transduct Target Ther. 2022;7(1):289.

389. Alisoltani A, Qiu X, Jaroszewski L, Sedova M, Iyer M, Godzik A. Gender differences in smoking-induced changes in the tumor immune microenvironment. Arch Biochem Biophys. 2023;739:109579.