

**Immune Response to Trauma: A Roadmap from Single-Cell Transcriptome and
Epigenome to Patient Classifications**

by

Tianmeng Chen

B.S. and M.S. (M.D. equivalent), Nankai University, 2012

M.S., Tufts University, 2016

Submitted to the Graduate Faculty of the
School of Medicine in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2021

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Tianmeng Chen

It was defended on

July 6, 2021

and approved by

Wendy Mars, Associate Professor, Department of Pathology

Wei Chen, Associate Professor, Department of Pediatrics

Yoram Vodovotz, Professor, Department of Surgery

Robert Lafyatis, Professor, Department of Medicine

Melanie Scott, Associate Professor, Department of Surgery

Dissertation Director: Timothy Billiar, Professor, Department of Surgery

Copyright © by Tianmeng Chen

2021

Immune Response to Trauma: A Roadmap from Single-Cell Transcriptome and Epigenome to Patient Classifications

Tianmeng Chen, PhD

University of Pittsburgh, 2021

Immune dysfunction is an important factor driving mortality and adverse outcomes after trauma but remains poorly understood, especially at cellular level. In this dissertation, we applied both single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq) to deconvolute the trauma-induced immune response.

First, we describe our work to apply scRNA-seq to circulating and bone marrow mononuclear cells in injured mice and peripheral blood mononuclear cells (PBMCs) in trauma patients. These are the first reported studies to characterize the mammalian response to systemic injury using scRNA-seq. In mice, the greatest changes in gene expression were seen in monocytes across both compartments. After systemic injury, the gene expression pattern of monocytes markedly deviated from steady state with corresponding changes in critical transcription factors (TFs), which can be traced back to myeloid progenitors. These changes involved up-regulation of inflammation and suppression of steady-state features, which were largely recapitulated in the human single-cell analysis. We generalized the major changes in human CD14⁺ monocytes into six signatures, which further defined two transcriptional subtypes (Signature Group: SG1 vs. SG2) identified in the whole-blood leukocyte transcriptome of trauma patients in the initial 12h after injury. Compared with SG2, SG1 patients exhibited delayed recovery, more severe organ dysfunction and a higher incidence of infection and non-infectious complications.

Next, we performed scATAC-seq on PBMCs isolated from a subset of trauma patients subjected to scRNA-seq, to determine if trauma-induced immune dysfunction was associated with

epigenomic changes. While corroborating previous transcriptomic changes, we uncovered global epigenetic alterations reflecting de-repression of polycomb targets, across multiple immune cell types. These included developmental loci, not normally expressed in hematopoietic lineages. Using whole-blood leukocyte transcriptomes of trauma patients, we validated the newly identified pathologic epigenomic signature in an independent dataset and defined the Epigenetic Groups (EG subtypes) associated with differential prognosis and distinct from SG subtypes.

Patient classifications based on either SG or EG subtypes including the independent prognostic value for each classification model were also recapitulated in burn and sepsis patients. These studies provide evidence for the broad impact of our analyses in the research field of critical illness.

Table of Contents

Preface	xvi
1.0 Introduction	1
1.1 Trauma	1
1.1.1 Definition and epidemiology	1
1.1.2 Transcriptomic studies and patient classifications	1
1.1.3 Remaining questions	2
1.2 Sepsis	3
1.2.1 Definition and epidemiology	3
1.2.2 Transcriptomic studies and patient classifications	3
1.2.3 Remaining questions	4
1.3 Human-mice comparison in critical illness	5
1.4 Rationale of this dissertation	6
2.0 A Roadmap from Single-Cell RNA sequencing to Patient Transcriptomic	
Subtypes for the Immune Response to Trauma	7
2.1 Introduction	7
2.2 Methods	8
2.2.1 Mouse polytrauma model	8
2.2.2 Patient and human volunteer enrollment	9
2.2.3 PBMCs and BMDCs isolation	9
2.2.4 Single-cell cDNA library preparation and sequencing	10
2.2.5 Single-cell sequencing data processing (mouse)	10

2.2.6 Single-cell sequencing data processing (human)	11
2.2.7 Doublet and low-quality cell removal.....	12
2.2.8 Cell staining for flow cytometry	12
2.2.9 Computation of PC associated genes and PC functional annotation	13
2.2.10 Pseudotime estimation	14
2.2.11 Regulon detection and PCA projection.....	14
2.2.12 Cell cycle phase assignment	15
2.2.13 RNA velocity computation	15
2.2.14 Enrichment analysis.....	16
2.2.15 Hierarchical clustering	16
2.2.16 Generation of customized signatures	17
2.2.17 Signature score calculation.....	18
2.2.18 Intrinsic signature score and intrinsic deviation score (IDS) calculation...19	
2.2.19 Clinical annotations of trauma dataset	19
2.2.20 Time-to-event analysis	20
2.2.21 Deconvolution of cell composition	20
2.2.22 Entrez ID – gene symbol and mouse – human homolog gene exchange	21
2.2.23 Study approval	21
2.2.24 Data and materials availability	22
2.3 Results.....	22
2.3.1 Dramatic transcriptomic changes in mouse circulating monocytes after systemic injury.....	22

2.3.2	Continuous changes in the myeloid cell transcriptome from the BM to the circulation after T/HS	27
2.3.3	Characterization of the transcriptomic changes in myeloid progenitors after T/HS.....	31
2.3.4	Characterization of the transcriptomic changes in the BM monocyte lineage after T/HS	33
2.3.5	Characterization of the transcriptomic changes in the BM neutrophil lineage after T/HS	40
2.3.6	Overview of the time-dependent transcriptomic changes in PBMCs from trauma patients	47
2.3.7	Characterization of the transcriptomic changes in human circulating monocytes.....	50
2.3.8	Generation of six CD14 ⁺ monocyte signatures	53
2.3.9	Validation of the six signatures in bulk RNA datasets	58
2.3.10	Two subtypes of trauma patients defined by the six signatures with differential prognostic value	59
2.3.11	Generation and validation of a classifier for subtype designation	66
2.4	Discussion	70
3.0	Longitudinal Analysis of Transcriptomic Subtypes in Trauma Patients	76
3.1	Introduction	76
3.2	Methods	77
3.3	Results.....	78
3.4	Discussion	85

4.0 Single-cell ATAC-seq Reveals Global Epigenetic Alterations in Immune Cells during Severe Trauma.....	87
4.1 Introduction	87
4.2 Methods	89
4.2.1 Patient and human volunteer enrollment	89
4.2.2 PBMC isolation, cryopreservation, and thawing	89
4.2.3 Nucleus isolation.....	90
4.2.4 Single-cell ATAC library preparation and sequencing.....	90
4.2.5 Single-cell ATAC data processing	91
4.2.6 Identification of DA peaks.....	92
4.2.7 Chromatin state enrichment analysis.....	92
4.2.8 Generation of state-barcode matrix and identification of state clusters.....	93
4.2.9 Enrichment analysis between gene sets.....	93
4.2.10 Hierarchical clustering	94
4.2.11 Calculation of signature scores	94
4.2.12 Time-to-event analysis	94
4.3 Results.....	95
4.3.1 Overview of trauma-induced open chromatin changes in PBMCs	95
4.3.2 Overview of trauma-induced open chromatin changes in circulating monocytes.....	97
4.3.3 Different regulatory mechanisms for the DA peaks at different CD14+ chromatin states	100

4.3.4 Characterization of CD14+ cell clustering based on global epigenetic alterations.	105
4.3.5 Generation and validation of gene signatures representing the global epigenetic alterations.	106
4.3.6 Global epigenetic alterations were the common changes across major immune cell types.	113
4.3.7 The independent prognostic value of epigenetic subtypes in trauma patients.	114
4.3.8 Recapitulation of epigenetic profile in scRNA data of trauma patients	122
4.3.9 The independent prognostic value of global epigenetic changes in burn and sepsis patients.	124
4.4 Discussion	136
5.0 Conclusions and Future Work.....	141
Appendix Spreadsheets	146
Bibliography	148

List of Tables

Table 1 Comparison of statistical methods between two papers.....	5
Table 2 Brief summary of gene modules identified from mouse bone marrow monocyte lineage.....	37
Table 3 Trauma Patient Clinical Characteristics.	48
Table 4 Six signatures derived from human CD14+ monocytes.	55
Table 5 Selected enriched GO terms for human CD14+ signatures.	57
Table 6 Basic characteristics and outcome parameters between two subtypes of trauma patients.	62
Table 7 C0C3C6_up signature.....	108
Table 8 The comparison between SG and EG classification.....	142

List of Figures

Figure 1 Overview of transcriptomic changes in mouse PBMCs at 6hrs after T/HS.....	23
Figure 2 Dramatic transcriptomic changes in mouse circulating monocytes after systemic injury.....	26
Figure 3 Generation of monocyte circulating markers.	27
Figure 4 Overview of paired mouse BMMCs + PBMCs at 6 hrs after T/HS.....	28
Figure 5 Expression of representative markers in paired BMMCs + PBMCs as shown in Figure 4B.....	29
Figure 6 Overview of transcriptomic changes in BMMCs at 6 hrs after T/HS.	30
Figure 7 Characterization of transcriptomic changes in the BM myeloid progenitors (mP) at 6 hrs after T/HS.....	32
Figure 8 Overview of transcriptomic changes in the BM monocyte lineages at 6 hrs after T/HS.	34
Figure 9 Characterization of transcriptomic changes in the BM monocyte lineages at 6 hrs after T/HS.	36
Figure 10 Transcriptomic changes in the BM monocyte lineages at 6 hrs after T/HS, Cont'd.	38
Figure 11 Validation of transcriptomic changes in the BM monocyte lineage at 6 hrs after T/HS by flow cytometry.....	39
Figure 12 Characterization of transcriptomic changes in the BM neutrophil lineage at 6 hrs after T/HS.	41

Figure 13 Simultaneous initiation of inflammatory and features of MDSC during emergency myelopoiesis.....	44
Figure 14 Characterization of transcriptomic changes in the BM monocyte and neutrophil lineages at 3 hrs and 24 hrs after T/HS.....	47
Figure 15 Overview of the transcriptomic changes in PBMCs from trauma patients over time.....	49
Figure 16 Characterization of the transcriptomic changes in human circulating monocytes after trauma.....	52
Figure 17 Generation and validation of six CD14+ monocyte signatures.....	56
Figure 18 The human and mouse data are generally consistent in the monocyte compartment before and after trauma.....	58
Figure 19 Six signatures define two patient subtypes associated with different prognosis..	61
Figure 20 Systemic characterization of the prognostic value for the six signatures.....	64
Figure 21 Generation of intrinsic signature score (IDS) to give a potential biological explanation for the patient heterogeneity with critical illness.....	67
Figure 22 Correlation between IDS and the PC highlighted in red blocks shown in Figure 20.....	67
Figure 23 Generation and validation of the classifier for SG subtype designation.....	69
Figure 24 Similarity between the new trajectory of monocytes after T/HS and the G pathway characterized by <i>Yanez et al.</i>.....	72
Figure 25 Interpretation of sepsis subtypes using our findings.....	74
Figure 26 Graphic summary of single-cell RNA-seq analyses.....	75
Figure 27 Prognostic value of SG subtypes at 1d, 4d and 7d post injury.....	79

Figure 28 K-M analysis in subgroup analysis.	81
Figure 29 Differential outcomes across four the SG1-SG2 combinations for 12h and 1d. ..	84
Figure 30 Overview of patient clinical information for scATAC-seq	96
Figure 31 Overview of open chromatin profile in PBMCs isolated from trauma patients in a time-series manner and paired with healthy controls.	97
Figure 32 Characterization of the trauma-induced open chromatin profiles in myeloid cells.	99
Figure 33 Different regulatory mechanisms for the DA peaks at different chromatin states.	102
Figure 34 ChromHMM 15-state model enrichment.	104
Figure 35 Characterization of global epigenetic changes across ChromHMM 15 states in CD14+ monocytes.	106
Figure 36 Validation of the C0C3C6_up signature in trauma bulk microarray data.....	112
Figure 37 The changes in the accessibility of C0C3C6_up peaks in other immune cell types besides CD14+ monocytes.	114
Figure 38 Epigenetic subtypes and prognostic value in trauma patients.	116
Figure 39 Expression of epigenetic signatures or genes among trauma Epigenetic Groups (EGs) along with healthy controls.	117
Figure 40 Gene expression profile among trauma Epigenetic Groups (EGs) (related to Figure 38).	119
Figure 41 Gene expression profile of polycomb targets across trauma patients shown in (Figure 38A).....	121
Figure 42 Recapitulate epigenetic profile in scRNA data.	123

Figure 43 Epigenetic subtypes and prognostic value in burn patients.	126
Figure 44 Expression of epigenetic signatures and genes among burn Epigenetic Groups (EGs) along with healthy controls (related to Figure 43).....	127
Figure 45 Different prognostic patterns of EGs in burn patients and impact of age on the patterns.	129
Figure 46 Epigenetic subtypes and prognostic value in sepsis patients.	131
Figure 47 Expression of epigenetic signatures and genes among sepsis Epigenetic Groups (EGs) along with healthy controls (related to Figure 46).....	132
Figure 48 Gene expression profile of polycomb targets across sepsis patients.	135
Figure 49 Relationship between Mars and EG classifications.....	139
Figure 50 New diagram of patient classifications in critical illness: Two different methods of classifications with independent prognostic value beyond each other.....	145

Preface

The past five years were tremendously valuable period in my life, bringing me from new to proficient to omics data analysis and trauma immunology, from a student to gradually becoming a more independent researcher. As the start of this dissertation, I would like to thank everyone who has kindly helped me along this journey.

First, I would like to thank my principal investigator, Dr. Timothy Billiar. Thank him for his advices along this long journey and giving me the chance to take over this interesting and important project. He created as many chances as he can to help me to move the project forward. He also gave me a lot of freedom to think independently and select the direction of my interest. He always encouraged me to attend conferences to broaden my view and created many opportunities for me to present and get feedback from the experts in the corresponding research field. I am really grateful that I can work in such an academic environment with sufficient freedom to gain the skills I need and to finish the project efficiently.

I would also like to thank my co-mentor, Dr. Wei Chen, for his guidance in the analysis of single-cell omics data, survival and longitudinal data. I also thank Tao Sun and Yue Wei, two lab members from Dr. Chen's lab, for the support in the development of random forest classifier and longitudinal analysis, respectively.

I am deeply appreciated all the extremely valuable guidance, suggestions and comments from Dr. Harinder Singh in the analysis of scATAC-seq data and I am so thankful for all his efforts in helping us to prepare the manuscript associated with Chapter 4.

I would like to thank all the lab members in Dr. Billiar's lab. Especially, thank Julia Conroy in preparing high-quality libraries for single-cell RNA and ATAC sequencing. Thank Zachary

Secunda in building the polytrauma mouse models. Thank Guang Fu and Patricia Loughran for teaching me how to do flow cytometry. Thank Debra Williams, Susan Schoanover and Kathy DiGiacoma for their administrative support. Thank Dr. Meihong Deng for helping me setting up this project at the very beginning.

All our collaborators were also essential for this project. Dr. Kong Chen and his lab member Xiaojing An kindly helped us to prepare single-cell cDNA libraries initially and also helped us to set up the platform of single-cell cDNA library preparation in our lab. Dr. Matthew Delano, Dr. Lyle Moldawer and Dr. Philip Efron provided very helpful interpretation for the clinical data of the trauma bulk microarray data generated by the Inflammation and the Host Response to Injury Large-Scale Collaborative. Dr. Christopher Seymour and Dr. Derek Angus provided insightful comments on the current sepsis classification. Dr. Rami Namas, Dr. Jason Sperry and Dr. Yoram Vodovotz contributed to the collection of clinical samples and clinical records for the patients subjected to single-cell sequencing. Moreover, all the patients who agreed to participate in the research must be counted.

Additionally, the research was supported in part by the University of Pittsburgh Center for Research Computing through the resources provided. We specially acknowledge the assistance of Dr. Fangping Mu. We also thank Dr. Yinghong Pan from the UPMC Genome Center in the assistance of sequencing our single-cell libraries.

Besides Dr. Billiar and Dr. Chen, I would like to thank all the other committee members, Dr. Yoram Vodovotz, Dr. Wendy Mars, Dr. Robert Lafyatis and Dr. Melanie Scott, for their valuable comments and feedback in each committee meeting. I am really grateful for the tremendous help and support from our program directors Dr. John Horn and Dr. Wendy Mars during my most difficult time.

Finally, I would like to thank my family and family relations. I am so appreciative that my aunt and uncle for their help with my life abroad and their advices on my career. Most importantly, I want to express my thanks to my parents from the bottom of my heart, for their deeply love and unconditional support. You raise me up to more than I can be.

Thank all of you who have kindly helped me walking through this long journey and thank all the ones who love me and who I love.

1.0 Introduction

1.1 Trauma

1.1.1 Definition and epidemiology

Traumatic injury is one of the major etiologies leading to critical illness. Severe injury is among top 3 causes of death and morbidity in individuals under 54 years old [1, 2]. Traumatic injury is an acute insult and causes an abrupt transition from a healthy state (i.e. homeostasis) to a state best described as a system-wide physiologic crisis, involving both innate and adaptive immune response, and an increased sensitivity to secondary infections [3]. Advances in clinical management have reduced early death substantially; however persistent organ dysfunction and delayed infections, both associated with immune dysfunction, remain poorly understood and difficult to prevent [4].

1.1.2 Transcriptomic studies and patient classifications

Three dynamic patterns of multiple organ dysfunction syndrome (MODS) after trauma that cannot be fully explained by injury severity have been described: (1) early recovery (median Time-to-Recovery [TTR]: 4 days, mortality: 14.4%), (2) delayed recovery (median TTR: 13 days, mortality: 35%) and (3) prolonged recovery (median TTR: 25 days, mortality: 46%) [5]. The diversity in clinical trajectories of trauma patients points to the heterogeneity in immune response among trauma patients.

Up till now, the largest scale transcriptomic analysis of human immune responses to trauma was a study of unseparated circulating leukocytes from severely injured patients published in 2011 by Xiao *et al.*, that included 167 patients sampled longitudinally up to 28 days after injury [6]. This study grouped the patients into uncomplicated, intermediate and complicated courses based on the outcome parameter (TTR) and discuss how transcriptomic changes differ from these outcome-based patient groups. This study introduced a novel paradigm to describe the immune-inflammatory response to trauma: an early induction of excessive pro-inflammatory pathways and simultaneous suppression of adaptive immune responses. Patients suffering complicated courses manifested leukocyte transcriptional patterns consistent with prolonged immune dysregulation [6]. Cabrera *et al.* demonstrated subsequently that differential transcriptomic changes could be identified within whole-blood leukocytes within 2 hours in severely injured patients who subsequently developed multiple organ dysfunction syndrome [7]. The work based on analysis of MODS cluster or clustering based on Luminex (cytokines) rather than transcriptomics largely supported the above results of trauma patients [8, 9].

1.1.3 Remaining questions

Even though there are several transcriptomic studies, none of these characterized the potential transcriptomic heterogeneity and differential prognostic value among trauma patients. Besides, all the available transcriptomic studies in trauma are bulk transcriptomes from unseparated leukocyte populations of cells. Studies at the single-cell level have been limited to the identification of the appearance of Th17 cells by Mass cytometry (CyTOF) in the circulation of severely injured patients [10]. Thus, little is known about the cell-specific pathways behind the pathogenic inflammation and immunosuppression that follows trauma.

1.2 Sepsis

1.2.1 Definition and epidemiology

Sepsis is another common etiology leading to critical illness. Different from trauma (non-infectious inflammatory studies), sepsis involves the host immune response to life-threatening infection and has a higher mortality. Compared with trauma patients, the average age of sepsis patients is much older, 65-68 years in most developed countries, which may also involve deficiency in immune response due to immuno-senescence [11, 12]. Despite of these differences, sepsis shares a global gene expression pattern in circulating compartment similar to trauma (Spearman correlation coefficient = 0.77, $p < 0.0001$) [13], indicating different etiologies of critical illness share some transcriptomic features.

1.2.2 Transcriptomic studies and patient classifications

Transcriptomic subtypes of whole-blood leukocytes have been characterized in sepsis patients, yielding two major patient classifications: SRS (1 and 2) [14] and Mars (1, 2, 3 and 4) [15], both were published in the journal *Lancet Respir Med*. For SRS subtypes, SRS1 patients exhibited an immunosuppressed phenotype including T-cell exhaustion and suppressed antigen presentation, compared with SRS2 patients. For Mars subtypes, Mars2 was associated with a gene expression pattern aligned with high levels of inflammation, cytokine signaling and activation of pattern recognition pathways. Mars3 was more enriched in adaptive immune signaling pathways, and Mars4 more specifically expressed genes involved in interferon signaling. Mars1 was a more

distinct subtype from the other 3 Mars subtypes, showing a high level of genes involved in heme metabolism along with a marked decrease in both innate and adaptive immune associated genes.

Among these subtypes, SRS1 and Mars1 were documented to have the worse prognosis in the original publications [14, 15]. However, when the authors of Mars paper tried to map the patients from the SRS paper using Mars designations, 60-70% of the patients associated with Mars1 turned out to be SRS2 patients (Figure S7 from the original paper [15]). Thus, the transcriptomic subtypes in sepsis are still not-well resolved and need to be refined. Furthermore, the underlying biological and cellular changes among different subtypes are not fully understand. Especially, it is difficult to interpret what is happening in the Mars1 patients at the transcriptional level. Thus, it is still largely unknown how to differentially treat each subset of patients based on their transcriptomic patterns.

1.2.3 Remaining questions

Hundreds of clinical trials have been carried out to attempt to improve the survival of sepsis patients based on targeting the immune response. However, all have failed to bring new treatments into the care of sepsis patients. These failures have been largely attributed to the unresolved patient heterogeneity due to non-genetic (different sources of pathogens, sites and timeliness) and genetic factors [16], which have not been well-defined as we mentioned above.

1.3 Human-mice comparison in critical illness

The lack of representative preclinical models (esp. rodent models) is another major factor, that may underlie numerous failed clinical trials in trauma and sepsis [16]. In 2013, Seok *et al* [17] compared dynamic changes in gene expression and the major signaling pathways activated between mouse models of burn, trauma and endotoxemia with the corresponding human subjects subjected to similar insults. They concluded that the transcriptomic responses in mouse models poorly represent for human responses to trauma, burns and endotoxemia (close to random correlation). This report challenged the use of mice in inflammatory diseases. In 2015, Takao *et al* [13] pointed out the issues with statistical analysis in that paper and re-analyzed the same datasets. This led to some improvement in consistency between humans and mice. The statistical methods used by the two papers are compared side by side below (Table 1):

Table 1 Comparison of statistical methods between two papers.

	Poorly mimic, Seok <i>et al</i>	Greatly mimic, Takao <i>et al</i>
DEGs included for comparison	Mouse orthologs corresponding to responsive human genes, and <i>vice versa</i> .	Genes significantly changed in both humans and mice.
Correlation	Pearson's coefficient of determination, r^2	Spearman's correlation coefficient, ρ (Rational: No reason to assume linearity and normal distribution)
Datasets	Some datasets include multiple subsets (Such as different mouse strains). Didn't analyze separately.	Compare each condition with the human reference condition independently.

DEGs: Differentially Expressed Genes.

However, there was one confounder which was not addressed in either paper, namely the difference in the composition of circulating leukocytes across two species. Human circulating

leukocytes are dominated by myeloid cells, while lymphocytes are the major population in mice. In this case, the comparison within specific cell populations is necessary and has not been done.

1.4 Rationale of this dissertation

Based on the above knowledge gaps in trauma, we decided to use the state-of-art technology, single-cell sequencing (scRNA-seq and scATAC-seq), to address the potential confounders in prior transcriptomic analyses and provide a full landscape of the immune response to trauma. Thus, this dissertation is mainly composed of data-driven hypotheses and validation. We used published bulk whole-blood transcriptomic data from trauma patients to validate the observations derived from single-cell data, and further explore the patient heterogeneity and potential prognostic value of our single-cell findings. Due to the controversy of the consistency between human and mice, we performed scRNA-seq on immune cells from both trauma patients and mice to provide a cross-species transcriptomic comparison at cellular level. Considering the association between trauma, burn and sepsis, we also used the signatures derived from single-cell trauma data to query the published bulk whole-blood transcriptomic data from burn and sepsis patients, aiming to provide broad new insights shared across different etiologies of critical illness.

2.0 A Roadmap from Single-Cell RNA sequencing to Patient Transcriptomic Subtypes for the Immune Response to Trauma

This chapter has been published and adapted from the below publication:

Chen, T., et al., A road map from single-cell transcriptome to patient classification for the immune response to trauma. JCI Insight, 2021. 6(2) [18].

2.1 Introduction

Immune dysfunction is an important factor driving mortality and adverse outcomes after trauma but remains poorly understood, especially at cellular level. To deconvolute the trauma-induced immune response, we applied single-cell RNA sequencing to circulating and bone marrow mononuclear cells in injured mice and circulating mononuclear cells in trauma patients.

To provide the landscape of transcriptomic changes at the single-cell level after systemic injury, we carried out scRNA-seq on bone marrow and circulating mononuclear cells (BMNCs and PBMCs) from injured mice and PBMCs from trauma patients. Studies in both a well-controlled mouse model of trauma and a detailed time course study in 10 severely injured humans identified the greatest changes in Ly6C⁺/CD14⁺ monocytes. This led us to characterize the major regulatory features in myeloid cells after systemic injury. To correlate these features with outcomes, we analyzed databases representing global gene expression changes in circulating leukocytes in large patient studies of trauma, burns and sepsis. In addition to providing a comprehensive landscape of the dynamic changes in transcriptomic patterns in myeloid

mononuclear cells after severe injury, we identify patient subtypes with potential prognostic value along with the critical regulatory networks (transcription factors) at the cellular level.

2.2 Methods

2.2.1 Mouse polytrauma model

We utilized a previously described mouse model of polytrauma that combines features commonly observed in critically ill trauma patients, including severe hemorrhagic shock and tissue trauma [19, 20]. Briefly, anesthetized 8-12 week old male C57BL/6 (Jackson Laboratory, Cat#000664) mice were subjected to bilateral lower extremity crush injury + injection of bone homogenate (a surrogate for long bone fracture). This was immediately followed by hemorrhagic shock for 1.5hrs at a mean arterial pressure of 28-32mmHg and then resuscitation with Lactated Ringers solution at 3x the volume of the shed blood. We harvested the blood and bone marrow samples at 3 different time points: 3hrs (escalation phase), 6hrs (peak systemic inflammation), and 24hrs (recovery phase) post-injury. Uninjured male littermates were used as controls. Peripheral blood was collected by cardiac puncture for PBMCs isolation. Tibias and femurs were collected for BMMCs isolation. Young male mice were used for this study because the greatest percentage of severely injured trauma patients are young males [21].

2.2.2 Patient and human volunteer enrollment

Patients suffering blunt or penetrating trauma that were admitted to the intensive care unit and suffering hypotension (systolic blood pressure <90 mmHg) or tachycardia (heart rate > 108) on admission were eligible for enrollment. Ten patients covering wide range of age and injury severity were selected for analysis. Blood samples for PBMCs isolation were obtained within 4hrs of injury and at 24hrs and 72hrs after injury. Blood drawn from a healthy age and sex matched uninjured volunteer was used to establish the baseline for each patient. The 72hr samples from 2 patients are not available (1 early death and 1 subject refused the final blood draw), for a total of 38 samples.

2.2.3 PBMCs and BMBCs isolation

Mouse PBMCs were isolated by standard procedure of Ficoll gradient centrifugation (Ficoll-Paque Premium 1.084, Cat#17-5446-02, GE Healthcare). Residual red blood cells were removed using red blood cell lysing buffer (R7757, Sigma-Aldrich). Bone marrow was obtained from tibias and femurs by flushing the marrow cavities with 3ml PBS (containing 2% FBS and 2mM EDTA). BMBCs were isolated by Ficoll gradient centrifugation using the same protocol described for mouse PBMCs. Human PBMCs were isolated by Ficoll-Paque PLUS (Cat#17-1440-03, GE Healthcare) using SepMate tubes (Stemcell) to accelerate the procedure. Other steps were largely the same as the mouse protocol.

The isolated cells were cryopreserved and thawed for analysis according to the 10x Genomics protocol [22]. Briefly, PBMCs were resuspended in 0.5ml resuspension medium (40% FBS in DMEM) and 0.5ml freezing medium (40% FBS + 30% DMSO in DMEM) in a 1:1 ratio.

BMMCs were resuspended in 1ml freezing media (90% FBS + 10% DMSO). Cells were chilled in CoolCell (Corning) in -80 overnight and transferred to liquid nitrogen. Cryopreserved vials were thawed in the water bath at 37°C, removed from water bath when a tiny ice crystal remained and then transferred to a 50ml conical tube after thawing was complete. A milliliter of thawing medium (PBMCs: 10% FBS in DMEM; BMMCs: 20% FBS in PBS) was added dropwise (5 sec/drop), followed by 2ml, 4ml, 8ml, 16ml thawing medium at ~ 1-min intervals. After this, the cells were washed and resuspended in calcium and magnesium free buffer (PBMCs: PBS with 0.04% BSA; BMMCs: PBS with 10% FBS) for cDNA library preparation.

2.2.4 Single-cell cDNA library preparation and sequencing

PBMCs and BMMCs were isolated by standard Ficoll centrifugation. Single-cell 3' cDNA libraries were prepared following 10x Genomics protocol [22] (mouse: v2; human: v3). Cases with corresponding controls were processed in parallel within the same batch. Libraries were pair-end sequenced on an Illumina HiSeq platform, with a read length of 150bp at each end. On average 180 million reads were sequenced for each sample.

2.2.5 Single-cell sequencing data processing (mouse)

To minimize the potential batch effect, we analyzed each batch of mouse data separately. For each batch, raw sequencing data was processed using the 10x Cell Ranger pipeline, *cellranger count* followed by *cellranger aggr* (mouse: v2.1.0, mapped to mm10) to generate a UMI count matrix and then further processed using the Seurat (v2.3.4) [23]. Genes expressed in ≥ 3 cells were retained. Cells with gene counts less than 200 or more than 5000 were filtered out. The number of

detected molecules per cell (nUMI), an unwanted source of variation, were regressed out by the `ScaleData()` function. Principle Component Analysis (PCA) was performed upon variable genes. Significant principal components (PCs) were defined by a clear elbow in the plot of the PC standard deviation. t-SNE, UMAP, and clustering analyses were performed based on the significant PCs. The resolution for graph-based clustering was tuned back and forth until the identified clusters were biologically meaningful. DEGs between clusters were computed by using the `FindMarkers()` function with default methods based on the Wilcoxon rank sum test. For specific cell populations of interest, we extracted the UMI count sub matrix and re-did the secondary analysis mentioned above in order to analyze the differences at a higher resolution. By this analytic workflow, we demonstrated that (1) duplicates largely overlap and (2) the major conclusions can be easily confirmed using different batches.

2.2.6 Single-cell sequencing data processing (human)

Based on the mouse experiments, biological replicates prepared in parallel were highly reproducible. Thus, similar to the mouse analysis, different time points sampled from the same patients with the matched healthy control were processed by *cellranger count/aggr* (v3.0.0, mapped to GRCh38) and then by Seurat (v3.0.2) for quality control and pre-processing in order to largely preserve the differences along timeline. Genes expressed in ≥ 3 cells were retained. Cells with gene counts less than 200 or more than 5000 or $\geq 20\%$ mitochondria genes were filtered out. To overcome human heterogeneity and to identify the same cell type or functional state in population, the data from different individuals were integrated by Seurat integration standard workflow [24]. To clarify, the integrated data were only used for dimension reduction (e.g. PCA, UMAP, t-SNE) and the downstream analysis taking the results of dimension reduction as input

(e.g. clustering). Other analyses were performed based on the original data matrix. For example, DEGs were identified by logistic regression using uncorrected and log-normalized expression data with batch as a potential variable, and then corrected by Bonferroni method for multiple testing (default method by Seurat).

2.2.7 Doublet and low-quality cell removal

Different cell types have different number of expressed genes. E.g. monocytes have more expressed genes compared with lymphocytes. One hard cutoff was not be applicable for all cell types. Thus, in initial quality control, we used a relatively low threshold to include more potential high-quality cells. In the analysis of a specific population, doublets and low-quality cells were more distinguishable at the higher resolution, usually forming small isolated clusters. Doublets were identified based on the biological knowledge, e.g. co-expressing both T and B cell markers. Low-quality cell clusters were identified by: (1) Most top genes were still mitochondria genes after initial quality control or (2) The number of expressed genes was extremely lower than other clusters and in the absence of cluster-specific genes. Thus, doublet and low-quality cell removal was an iterative process.

2.2.8 Cell staining for flow cytometry

Fresh mouse BMDCs isolated by Ficoll centrifugation from control and T/HS (time point: 6hrs) were stained for flow cytometry, 2 mice/group. Transcription Factor Staining Buffer Set from eBioscience (Cat# 00-5523-00) was used for intracellular staining of IRF8. Experiment was repeated for three times. Analysis was performed on by flow cytometer (LSR-15) and analyzed

with FlowJo software. The CD45⁺ live cells in the monocyte lineage were defined by 2 gating strategies: (1) Lin⁻CD115⁺Ly6G⁻ and (2) Lin⁻Ly6C⁺⁺Ly6G⁻. Multi-dimensional protein data of Lin⁻Ly6C⁺⁺Ly6G⁻ gated monocytes were also visualized by Matlab/Cyt3 [25]. The compensated data (fcs files) of gated populations were taken as the input and Arcsinh transformed. The cells from each mouse were down-sampled to 3000 cells for visualization.

Fluorophore-conjugated antibodies against myeloid lineage markers (CD11b, Ly6G, Ly6C, CD115), a transcription factor IRF8 with IgG1 kappa Isotype Control, other lineage markers (CD3ε, TCRγδ, B220, NK1.1, Ter119, CD19) and Leukocyte common antigen CD45 were listed as below: LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit (Invitrogen, Cat# L34965), anti-CD3ε FITC (145-2C11) (eBioscience, Cat# 11-0031-85), anti-TCRγδ FITC (GL3) (eBioscience, Cat# 11-5711-82), anti-B220 FITC (RA3-6B2) (eBioscience, Cat# 11-0452-82), anti-NK1.1 FITC (PK136) (BD, Cat# 553164), anti-Ter119 FITC (TER-119) (BioLegend, Cat# 116205), anti-CD19 FITC (1D3) (BD, Cat# 553785), anti-CD45 BUV395 (30-F11) (Bioscience, Cat# 564279), anti-CD11b PE-Cy7 (M1/70) (BioLegend, Cat# 101216), anti-Ly6G APC-Cy7 (1A8) (BD, Cat# 560600), anti-Ly6C PerCP-Cy5.5 (HK1.4) (eBioscience, Cat# 45-5932-82), anti-CD115 PE (AFS98) (eBioscience, Cat# 12-1152-81), anti-IRF8 APC (V3GYWCH) (eBioscience, Cat# 17-9852-80), anti-IgG1 kappa Isotype Control APC (P3.6.2.8.1) (eBioscience, Cat# 17-4714-82).

2.2.9 Computation of PC associated genes and PC functional annotation

For the PC of interest, we computed Pearson's correlation between the scaled expression value by Seurat [23] and PC coordinates for each gene. A Benjamini-Hochberg adjusted p-value ≤ 0.05 was used as the cutoff to define PC-associated genes. With the correlation coefficient as the rank, GSEA was performed using the fgsea R package (v1.6.0). The top enriched gene sets with

positive NES (normalized enrichment scores) largely indicated the biological functions on the PC positive side, vice versa.

2.2.10 Pseudotime estimation

Pseudotime was computed by the Monocle [26] R package (v2.8.0) using the default parameters taking UMI matrix as input. To avoid the influence of cell cycle phases, pseudotime was computed after removing all cell cycle genes based on GO term annotation (GO:0007049). We forced the myeloid progenitor-enriched state as pseudotime zero. Among the variable genes identified from Seurat [23], the genes differentially expressed along pseudotime were identified by the differential GeneTest() function. The genes with q-value <0.001 (q-value was provided by Monocle) were used to build up the transcriptomic profile and cluster into gene modules.

2.2.11 Regulon detection and PCA projection

Myeloid regulons were computed with the SCENIC [27] R package (v1.0.1.1) using the UMI count matrix of myeloid cells with the default parameters. The computed regulons were further used as the gene sets for enrichment analysis or to calculate signature scores.

For motif references (from <https://resources.aertslab.org/cistarget/>), “mm10_refseq-r80_500bp_up_and_100bp_down_tss.mc9nr.feather” and “mm10_refseq-r80_10kb_up_and_down_tss.mc9nr.feather” RcisTarget databases were downloaded for mouse regulon detection. “hg38_refseq-r80_10kb_up_and_down_tss.mc9nr.feather” and “hg38_refseq-r80_500bp_up_and_100bp_down_tss.mc9nr.feather” were downloaded for human regulon detection.

Regulons were projected onto PCA 2D space largely following the published methods [28, 29]. Each single cell had a coordinate for each PC and was assigned an AUC (area under curve) value for each regulon by SCENIC. Pearson's correlation was calculated between AUC values and PC coordinates. Finally, each regulon was plotted on PCA 2D space based on the correlation coefficients with corresponding two PCs, respectively.

2.2.12 Cell cycle phase assignment

Cell cycle phases were computed by the `cyclone()` function from `scran` R package (v1.8.2). The UMI count matrix was taken as the input. The pre-defined classifiers provided with the `scran` package were used to assign cell cycle phases.

2.2.13 RNA velocity computation

RNA velocity of myeloid progenitors was computed by `velocity` [30] (v0.17). The mouse specific reference, "mm10_rmsk.gtf", was downloaded from [UCSC genome browser](#) as expressed repeats annotation. "genes.gtf" was generated by `cellranger mkref` as a genome annotation file. The `cellranger count` output was taken as the input to generate a loom file for each sample. Loom files from different samples were merged using the `loompy` python package. Finally, RNA velocity was estimated using the `velocity` R package (v0.6).

2.2.14 Enrichment analysis

(1) Enrichment between two gene sets: For a pre-ranked gene list, GSEA [31] was performed using the fgsea R package (v1.6.0). For the gene lists without rank, “a” represents the number of shared genes between gene set 1 and gene set 2; “b” represents the number of genes only in gene set 2; “c” represents the number of genes only in gene set 1. Universe genes (N) were defined as the genes expressed in $\geq 0.5\%$ of the cells used to compute corresponding gene modules. Fold enrichment (FE) was computed as below:

$$FE = \frac{a}{a+b} \bigg/ \frac{a+c}{N}$$

The hypergeometric p value for enrichment was computed and adjusted by the Benjamini-Hochberg method for multiple testing. MSigDB gene sets v5.2 were used.

(2) Enrichment or deletion between human monocyte clusters and time points: the two-sided p values of the χ^2 test and odds ratio were computed. The p value was adjusted by the Benjamini-Hochberg method.

2.2.15 Hierarchical clustering

Ward clustering was performed using hclust() function in R with the agglomeration method set as “ward.D2” [32]. The distance matrix was 1 minus Pearson’s correlation. Hierarchical clustering was used to identify gene or patient clusters.

2.2.16 Generation of customized signatures

(1) Circulating monocyte markers: To roughly estimate the developmental relationship among mouse circulating monocytes, GSE95702 [33] was used to generate this signature. The monocyte development referenced in this paper is MDP (monocyte-macrophage DC progenitors) → cMoP (common monocyte progenitors) → BM Ly6C⁺ → blood Ly6C⁺ → blood Ly6C^{int} → blood Ly6C^{low}. We identified the genes that were up-regulated in circulating monocytes compared with the monocyte progenitors and Ly6C⁺ monocytes in BM. DEGs were computed using the limma R package (v3.36.3) on the log₂ transformed expression values with Benjamini-Hochberg adjusted p-value = 10⁻⁴ as the cutoff.

(2) Monocyte differentiation associated genes: To validate the pseudotime of the monocyte lineage, we extracted the genes which are positively or negatively correlated with monocyte differentiation in bone marrow from the published dataset GSE95702 [33]. First, PCA was performed on the samples of MDP, cMop and BM Ly6C⁺ monocytes. PC1 largely represented the differentiation process of monocytes in bone marrow, with the above 3 populations aligning from the left, middle and to the right. For each gene, we computed Pearson's correlation between the log₂ transformed expression value and PC1 coordinates. Benjamini-Hochberg adjusted p-value ≤ 0.05 was used as the cutoff to define the genes which are significantly associated with PC1. Among these genes, the ones with positive Pearson's correlation coefficient are the genes which are gradually down regulated along monocyte differentiation (signature name: Diff down), and those with negative Pearson's correlation coefficient were gradually up regulated along monocyte differentiation (signature name: Diff up).

(3) MDSC signatures: To evaluate immunosuppressive potential, signatures for MDSC from 5 different sources were extracted from the published dataset GSE21927 [34]. Generally,

different sources of MDSC were compared with the CD11b⁺ counterparts isolated from healthy spleen from the same species. DEGs were computed using the limma R package (v3.36.3) on the expression values (already log₂ transformed). The genes that were significantly up-regulated in MDSC were used to identify MDSC signatures. The cutoff was set as Benjamini-Hochberg adjusted p-value = 10⁻³ for BM-derived MDSC and 10⁻⁵ for tumor-derived MDSC.

2.2.17 Signature score calculation

Signature scores were calculated as the average expression of the signature genes (or probes for microarray data) after z-score transformation across the patients, as described by Guo *et al* [35]. The “signature score matrix” based on the six human monocyte signatures was calculated in this way.

(1) For mouse single-cell RNA sequencing data, the average of Seurat (v2.3.4) [23] scaled values (stored in @scale.data slot which are normalized, scaled, log and z-score transformed) was the signature score for each cell. (2) For GEO datasets, the expression values in the matrix were log₂ transformed (if this had not already been performed) and z-score transformed. Then, the average of the log₂ and z-score transformed expression values was defined as the signature score for each subject.

We noticed that MDSC signatures contains multiple cell cycle genes. In this case, cell cycle phases would become a significant cofounder, especially in the BM. Thus, MDSC signatures were calculated after removal of all the cell cycle genes based on the GO term annotation (<http://www.informatics.jax.org/go/term/GO:0007049>).

2.2.18 Intrinsic signature score and intrinsic deviation score (IDS) calculation

We defined intrinsic signature scores (annotated as IC1-IC6, corresponding to the six signatures C1-C6) in order to reflect relative expression of the six signatures the within an individual. For each patient, six intrinsic signature scores were calculated as follows: (1) Extraction of the log2 transformed expression values of 129 signature genes from the full transcriptomic data; (2) Z-score transformation across all signature genes; (3) For each signature, the corresponding z-score transformed values were averaged to get the intrinsic signature score. Thus, each patient was assigned 6 values.

Random forest was used to establish the classifier taking the six intrinsic signature scores as input to predict which subtype the patient should belong to. We used all the first sampled data points from 167 trauma patients as the training data set with 5-fold cross validation. Subtypes were obtained from clustering analysis. Burn and Sepsis datasets with available survival data were used as two independent validation cohorts.

C1-C3 were the signatures induced after trauma, so their signs were “+1”. C4-C6 were the signatures suppressed after trauma, so their signs were “-1”. IDS was calculated by the equation shown below:

$$IDS = \frac{\sum_{i=1}^6 IC_i \times Sign_i}{6}$$

2.2.19 Clinical annotations of trauma dataset

In the trauma dataset, samples annotated as either “low RNA quality” or “incomplete time points” in metadata were excluded. Trauma patient outcomes were originally classified as an

uncomplicated when time to recovery (TTR) was <5 days, intermediate when TTR was ≥ 5 and ≤ 14 days and complicated when TTR was >14 days or the patient died within 28 days. TTR was defined as the number of days following the trauma event until organ dysfunction had resolved [6]. For the analysis of this data set, we pooled intermediate and uncomplicated cases together and classified the patients into either complicated (TTR >14 days or death) or non-complicated (TTR ≤ 14 days) categories.

2.2.20 Time-to-event analysis

For the gene array datasets that have multiple time points for each patient (trauma and burn), only the 1st sampled gene array data points were included for the survival analysis. For trauma patients, event was set as recovery status, due to the very few cases of in-hospital deaths (160 alive vs. 7 dead, mortality rate: 4.2%). The definition of recovery was based on the annotation from the original dataset [6]. For non-survivors, “hospital length of stay” was used as the time and recovery status was annotated as “No”. For burn and sepsis patients, the event was set as in-hospital death. A Kaplan-Meier curve was plotted by survival R package (v2.43.3) for visualization 28-day prognosis and the Log-rank p value was computed. Cox proportional hazards model was performed by `coxph()` function in R adjusting for covariates, including age, sex, serum lactate within 6hr and severity (if available).

2.2.21 Deconvolution of cell composition

To prepare the input of the array data, CEL files were downloaded from GSE36809 and then processed by `CEL_to_mixture.R` provided by CIBERSORT. For the input of signature matrix,

we used the signature matrix of LM22, which is CIBERSORT provided and contains 22 functionally defined human immune subsets. We used the deconvoluted results for the 1st time points of 167 trauma patient and added the neutrophil + monocyte percentage into the cox regression model.

2.2.22 Entrez ID – gene symbol and mouse – human homolog gene exchange

Genome wide annotation for mouse (org.Mm.eg.db, v3.6.0) and for human (org.Hs.eg.db, v3.6.0) were installed. Entrez ID – gene symbol exchange was performed by AnnotationDbi R package (v1.42.1). The HomoloGene data file was obtained by getHomologene() function from homologene R package. Mouse entrez IDs were mapped to human homolog entrez IDs using annotation Tools R package (v1.58.0)

2.2.23 Study approval

Mouse experimental protocols were approved by the Institutional Animal Use and Care Committee of the University of Pittsburgh. Experimental procedures were carried out in accordance with all regulations regarding the care and use of experimental animals (National Institutes of Health).

Trauma patients and healthy volunteers were enrolled in an observational study approved by the University of Pittsburgh Institutional Review Board. Informed consent was obtained from all the subjects (or next of kin).

2.2.24 Data and materials availability

The raw scRNA-seq datasets (both human and mouse) in the FASTQ format with filtered gene/barcode matrix have been uploaded to the Gene Expression Omnibus (GEO) (GSE162806). The analyzed published datasets can be accessed via GEO (Trauma bulk data: GSE36809, Burn bulk data: GSE37069, Mars sepsis bulk data: GSE65682, SRS sepsis bulk data: E-MTAB-4421). All clinical annotations for trauma and burn bulk data were obtained from the lead author of the original paper [6].

2.3 Results

2.3.1 Dramatic transcriptomic changes in mouse circulating monocytes after systemic injury

To deconvolute the immune response to trauma, scRNA-seq was performed on PBMCs isolated from mice subjected to tissue trauma with hemorrhagic shock (T/HS) [19, 20] and their uninjured littermates (2 mice/group) (Figure 1A). As shown in the t-Distributed Stochastic Neighbor Embedding [36] (t-SNE) plots (Figure 1B), biological duplicates overlapped very well. Thus, the shift between experimental groups can be expected reflect the trauma-induced differences. At 6h after injury, the peak of systemic inflammation [37], the greatest changes were observed in monocytes represented by an obvious transcriptomic shift in the t-SNE plot and the largest number of DEGs (Figure 1B-1H).

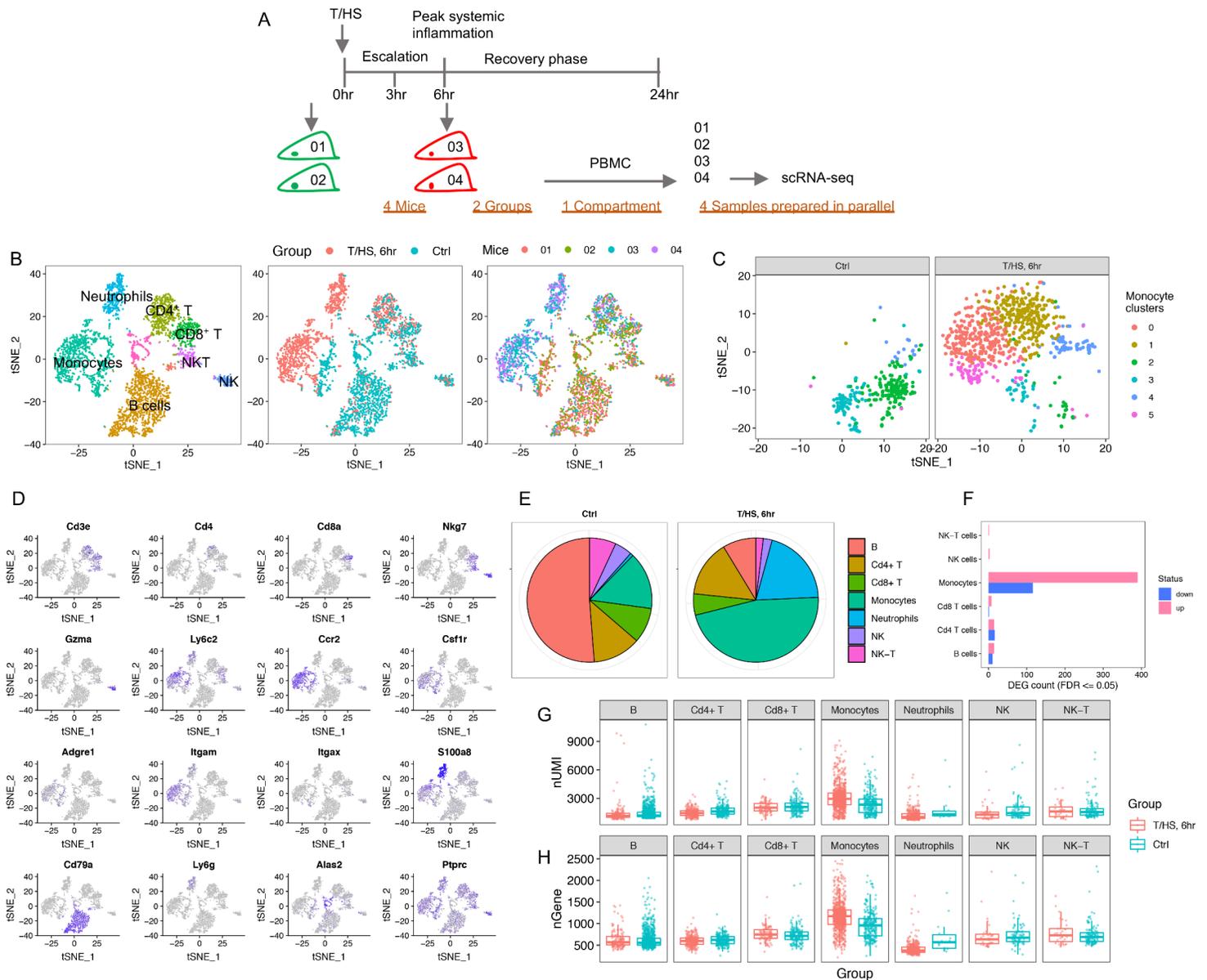


Figure 1 Overview of transcriptomic changes in mouse PBMCs at 6hrs after T/HS.

(A) Experimental design of the data shown in Figure 1-2. (B) t-SNE plot shows PBMCs from four mice color coded by major cell types, by animal groups or by individual mice. (C) Identified six clusters in circulating monocytes. (D) t-SNE plot of PBMCs as shown in Figure 1B. The expression of representative markers is shown. (E) Changes in the fractions of major cell types in PBMCs after T/HS. (F) The number of DEGs detected in each cell type (adjusted p-value < 0.05). (G-H) The number of molecules (UMI: unique molecular identifier) (G) and detected genes (H) are shown by cell types + groups. Each dot represents a cell.

To characterize monocytes at a higher resolution, we extracted and re-analyzed monocytes separately. The designation of mouse circulating monocytes is based on the surface marker Ly6C (coded by *Ly6c2*) [38], and circulating Ly6C⁺ monocytes give rise to Ly6C⁻ monocytes [33]. We recapitulated two steady state clusters corresponding to classical (cluster 2: *Ly6c2*⁺) and patrolling monocytes (cluster 3: *Ly6c2*⁻) [38] in the uninjured mice. We also identified three new monocyte clusters that were distinct from steady-state monocytes, which showed a gradient in *Ly6c2* expression after T/HS (clusters 1, 0 and 5: *Ly6c2*^{high}, *Ly6c2*^{int}, *Ly6c2*^{low}). Cluster 4 was comprised of monocyte-platelet aggregates highly expressing platelet markers (*Pf4* and *Pbpp*) (Figure 1C and Figure 2A). Principal component analysis (PCA) revealed a right shift of monocytes after T/HS on PC1. Based on gene set enrichment analysis (GSEA) [31], the right side of PC1 associates with inflammation and the left side with lymphocyte activation (Figure 2B-2C), suggesting that T/HS-induced monocytes are more inflammatory but deficient in the capacity for lymphocyte activation compared with steady-state monocytes.

The monocyte clusters that appeared after T/HS could be derived from cells already present in the circulation that underwent transcriptional changes or from bone marrow (BM). To address these possibilities, we generated a customized gene signature representing the upregulated genes in circulating monocytes compared with BM monocytes under steady state [33] (Figure 3). Monocyte-platelet aggregates (cluster 4) were excluded from the analyses of developmental status, due to the confounding effects of multiplets in single-cell analysis. The newly identified monocytes after T/HS displayed lower signature scores than the steady-state monocytes (Figure 2D), indicating that the monocytes after T/HS were more immature. Furthermore, the DEG were largely preserved in the newly identified clusters after T/HS (1→0→5, “→” followed a decreasing gradient in *Ly6c2* expression) and in the steady-state clusters (2→3), but were minimally shared

between the two experimental conditions (Figure 2E). Both of these observations suggested that the new monocyte clusters observed after trauma were derived from BM.

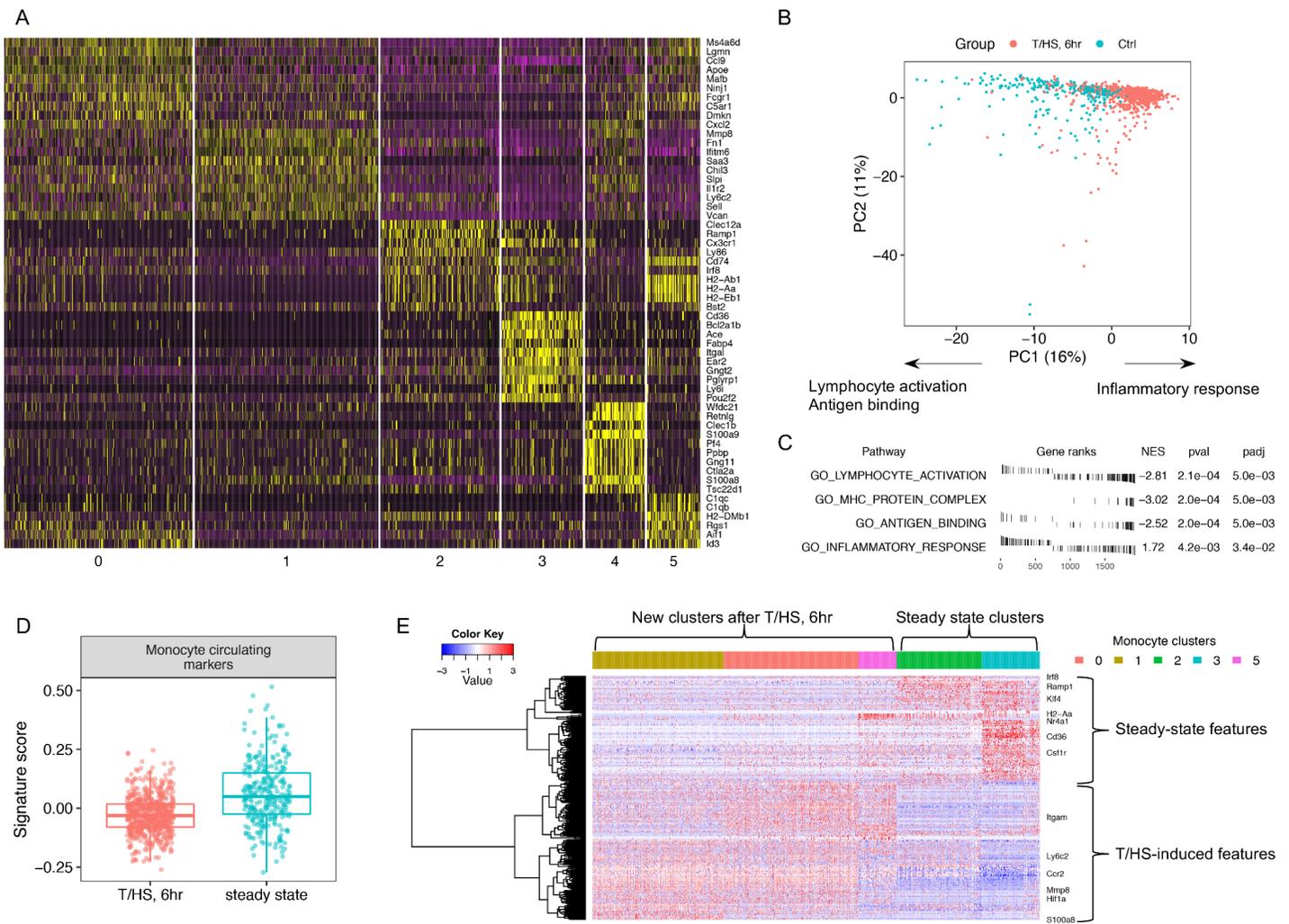


Figure 2 Dramatic transcriptomic changes in mouse circulating monocytes after systemic injury.

(A) Top markers for each monocyte cluster as shown in Figure 1C. Cells (columns) are ordered by clusters. (B) PCA plot of circulating monocytes color coded by groups. Biological interpretations are annotated based on the results shown in (B). (C) Selected enriched GO terms of PC1-associated genes by GSEA. (D) The monocytes after T/HS express less monocyte circulating markers than monocytes from control mice. (E) RNA profile of circulating monocytes built upon pairwise DEGs between each two-cluster combination. Cells (columns) are ordered by clusters. Genes (rows) are clustered into two large clusters generally representing either steady-state or T/HS-induced features.

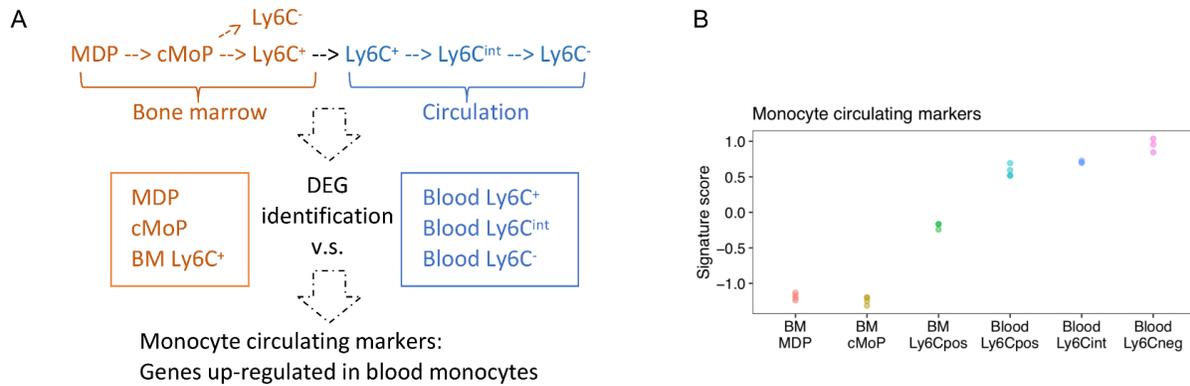


Figure 3 Generation of monocyte circulating markers.

(A) Schema describing the workflow for how monocyte circulating markers were extracted from GSE95702 dataset.

(B) Visualization of the expression of extracted monocyte circulating markers in monocyte subsets characterized in GSE95702 dataset, which confirms the extracted signatures are indeed the genes up-regulated in circulating monocytes.

2.3.2 Continuous changes in the myeloid cell transcriptome from the BM to the circulation after T/HS

We next carried out scRNA-seq on paired PBMCs and BMMCs from additional control and T/HS mice at 6h (2 mice/group) (Figure 4A). t-SNE across the circulating and BM compartments displays the large differences in myeloid cells after T/HS, especially in the monocyte lineage. The changes initiated in the BM were continuous to the circulating compartment (Figure 4B and Figure 5). PCA indicates that BM neutrophil and monocyte lineages underwent similar changes after systemic injury demonstrated by the positive side of PC3 and represented by inflammation, response to stress, and apoptosis (Figure 6A-6C).

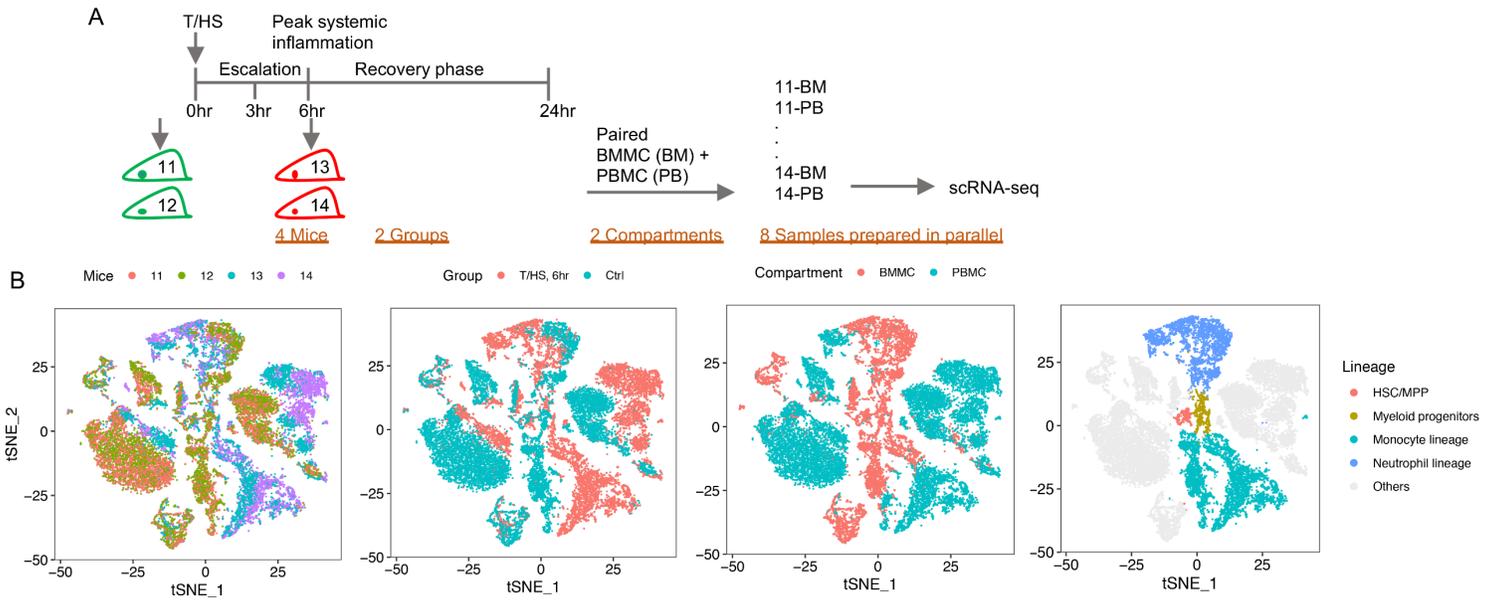


Figure 4 Overview of paired mouse BMMCs + PBMCs at 6 hrs after T/HS.

(A) Experimental design for the data shown in Figure 4-13, 2 mice/group. (B) t-SNE plot of BMMCs + PBMCs from the four mice are color coded by individual mice, by groups, by cell compartments or by cell lineages. HSC: hematopoietic stem cells; MPP: multipotent progenitors.

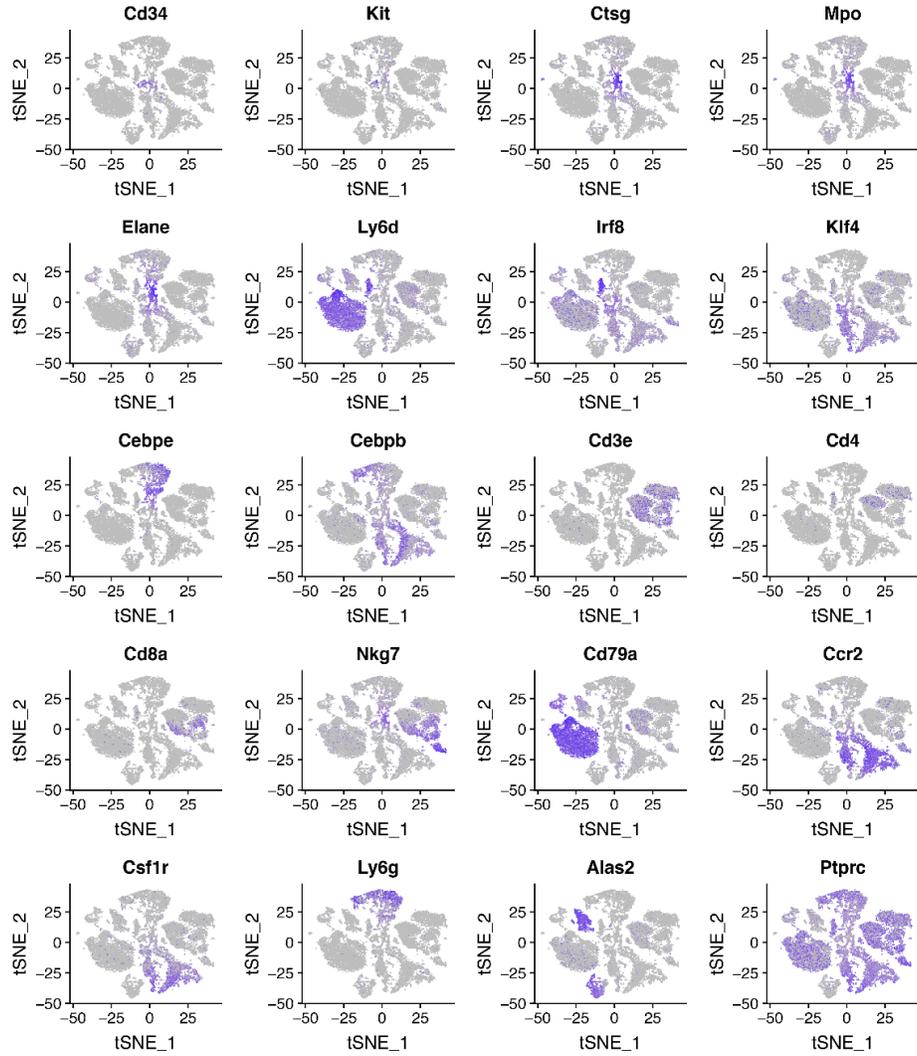


Figure 5 Expression of representative markers in paired BMBCs + PBMCs as shown in Figure 4B.

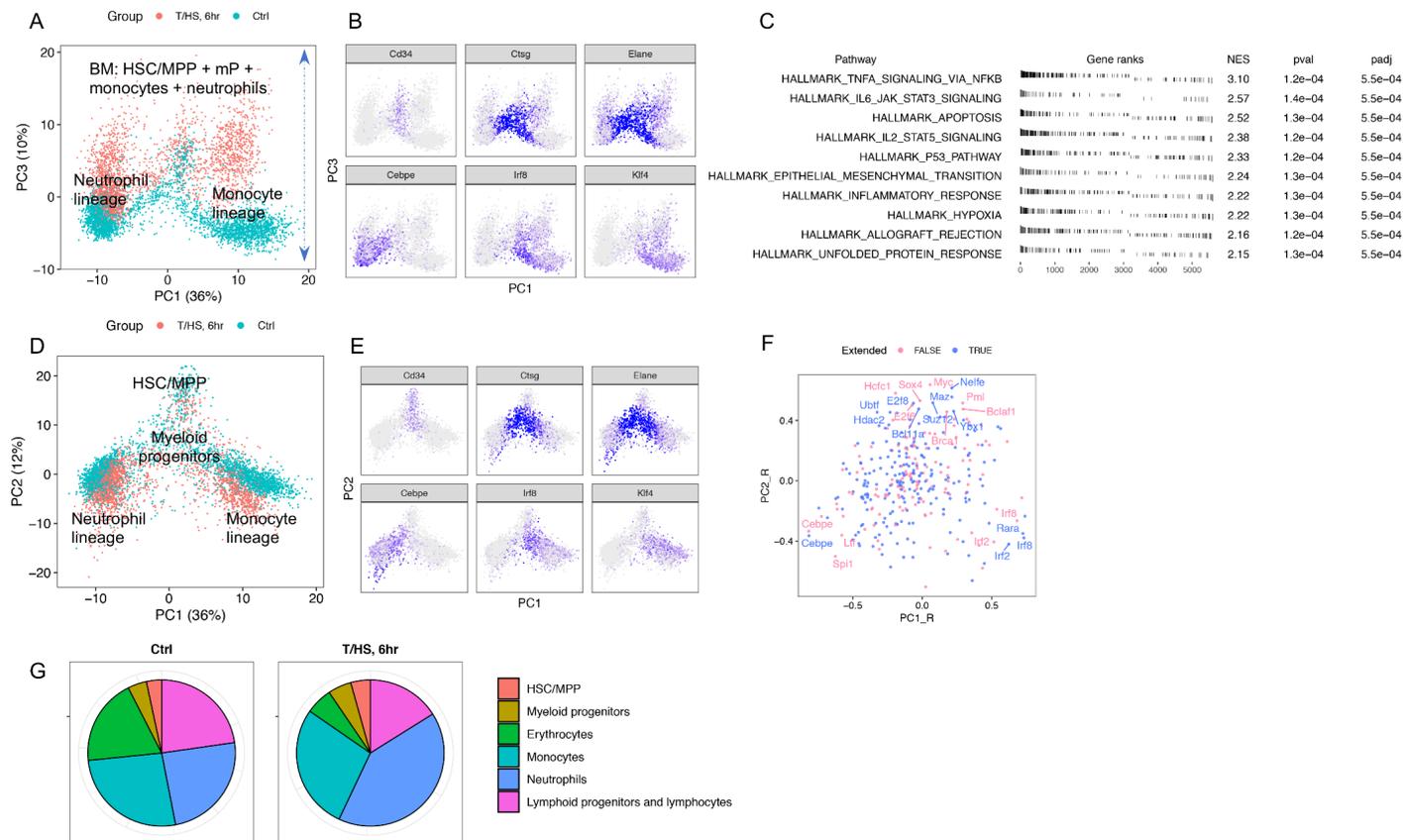


Figure 6 Overview of transcriptomic changes in BMDCs at 6 hrs after T/HS.

(A) PCA plot (PC1 vs. PC3) of BMDCs including HSC/MPP (hematopoietic stem cells/multipotent progenitors) and mP (myeloid progenitors) as shown in the 4th panel of Figure 4B color coded by groups. (B) Expression of lineage markers. (C) Top ten hallmark pathways enriched on the positive side of PC3. (D) PCA (PC1 vs. PC2) plot of BM myeloid cells, including HSC/MPP + myeloid progenitors, color coded by groups. Expression of lineage markers are shown in (E). (F) Projection of the regulons on the PCA 2D space as shown in (D-E). Well-established TFs (e.g. *Cebpe* ~ neutrophils [39], *Irf8* ~ monocytes [39], *Sox4* ~ stem cells [40]) largely overlay with the known corresponding lineages, supporting the reliability of the computed regulons. (G) Changes in the fractions of major cell types in BMDCs after T/HS.

The regulatory pathways associated with the myeloid trajectories were further explored by computing gene regulatory networks (regulons) using SCENIC [27]. A regulon represents the co-

expressed set of genes detected within scRNA-seq data, including a core transcription factor (TF) and the TF regulated genes containing the TF binding motif depicted as “TF (number of genes)”. We followed published methods [28, 29] and projected the regulons on PCA 2D space. Well-established TFs largely overlay with the known corresponding lineages (Figure 6D-6F), supporting the reliability of the computed regulons. SCENIC provides two kinds of regulons: (1) Main regulons (non-extended) only using the high confidence annotations; and (2) extended regulons also including lower confidence annotations. Both types of regulons yielded similar results (Figure 6F). In the following analysis, we only used the main regulons to establish the regulatory landscape.

2.3.3 Characterization of the transcriptomic changes in myeloid progenitors after T/HS

We next characterized the transcriptomic changes in myelopoiesis at a proximal branching point by analyzing BM myeloid progenitors (mP) from control and T/HS mice (Figure 4B, the 4th panel). These cells co-expressed myeloid progenitor mRNA markers (*Ctsg*, *Mpo* and *Elane*) (Figure 5) and largely corresponded to common myeloid progenitors (CMP) and granulocyte-monocyte progenitors (GMP) as previously characterized [39]. We identified five distinct progenitor clusters (Figure 7A) that included: multi-Lin (clusters 3 and 4), monocyte-skewed (clusters 0 and 1) and neutrophil-skewed (cluster 2) mP as shown in UMAP (Uniform Manifold Approximation and Projection) [41]. RNA velocity [30] is an algorithm estimating the future state of single cells. The opposite directions in the RNA velocity (shown by the arrows) observed for clusters 3 and 4 further suggests two potential cell fates and supports their identity as multi-Lin progenitors (Figure 7B). Monocle [26] is an algorithm to reconstruct the trajectory of

differentiation using scRNA-seq data. The trajectories of mP established by Monocle (Figure 7C) were largely consistent with the identified clusters.

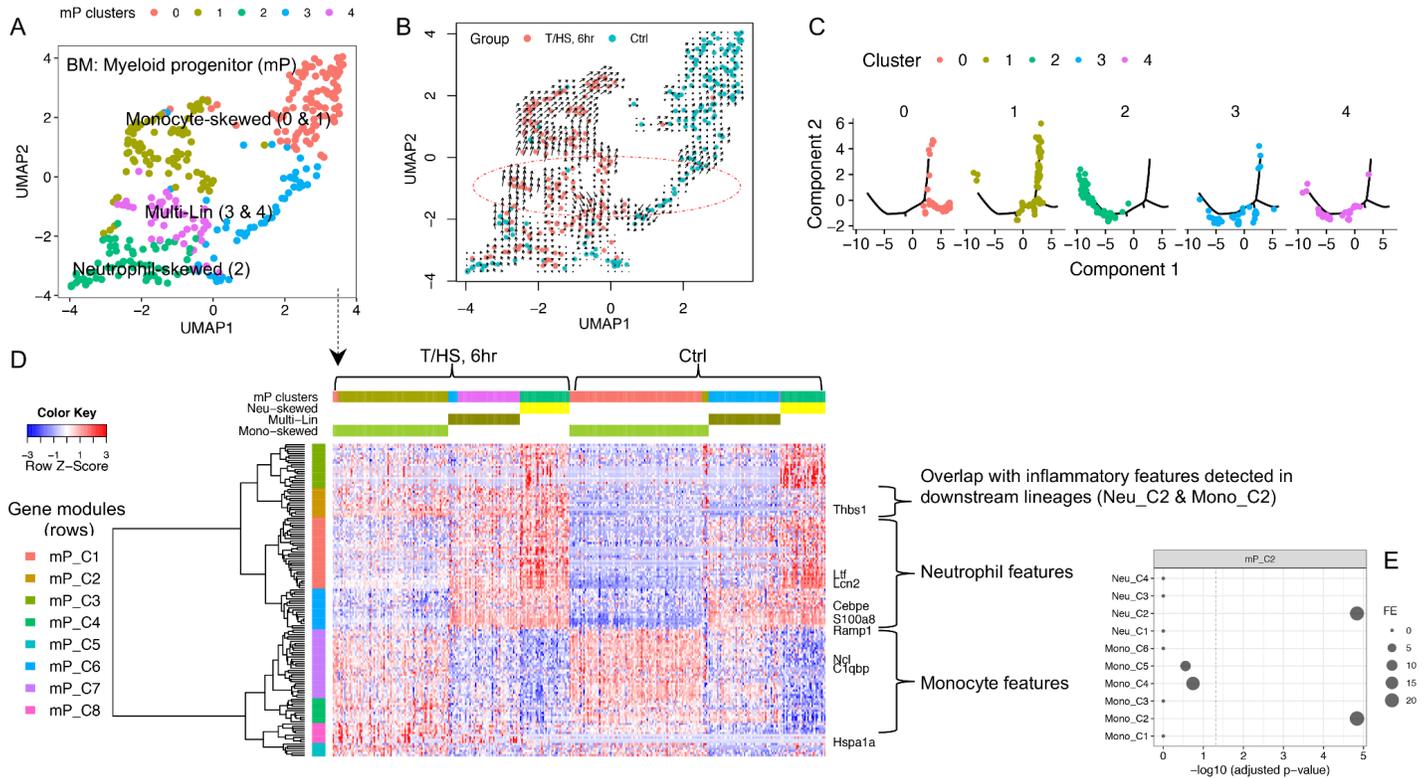


Figure 7 Characterization of transcriptomic changes in the BM myeloid progenitors (mP) at 6 hrs after T/HS.

(A) UMAP plot color coded by five mP clusters. (B) UMAP plot color coded by groups (RNA velocity shown as arrows). (C) The developmental trajectories constructed by Monocle 2. Cells are color coded by mP clusters. (D) RNA profile built upon pairwise DEG between mP clusters (fold change ≥ 2 & Bonferroni adjusted p-value < 0.05). Cells (columns) are ordered first by groups then by clusters. Genes (rows) are clustered into eight gene modules (mP_C1~C8). Single-cell transcriptomic data were collected from n=2 mice/group as shown in Figure 4A. (E) Enrichment analysis is performed between mP_C2 and each gene module identified from the BM monocyte and neutrophil lineages, demonstrating that the features of mP_C2 are preserved in the downstream lineages. Hypergeometric p-value was computed and then adjusted by Benjamini-Hochberg methods. Black vertical dash line annotated where adjusted p-value is equal to 0.05.

The transcriptomic profiles between the five mP clusters were established by pairwise DEGs comparisons. Ward hierarchical clustering [32] yielded 8 gene modules (mP_C1-C8) (Figure 7D). All mouse gene modules in this study are annotated as “CellType_Cluster (C)#”. The steady state multi-Lin mP expressed neutrophil and monocyte lineage modules at low levels, while clusters representing skewed mP highly expressed single lineage-specific modules. These patterns are consistent with the binary cell fate choice in myeloid progenitors described under baseline conditions by Olsson *et al* [39]. Trauma induced two major changes that include: (1) A monocyte-to-neutrophil shift in multi-Lin mP and (2) an up-regulation of gene module mP_C2 across all T/HS mP clusters. The features of mP_C2 were preserved in the analysis of the downstream monocytes and neutrophils discussed below (Figure 7E).

2.3.4 Characterization of the transcriptomic changes in the BM monocyte lineage after T/HS

To characterize the full developmental trajectories during monopoiesis, we analyzed BM monocytes along with BM mP together (Scheme of the analysis depicted in Figure 8A). PCA indicates that T/HS induced dramatic changes in the monocyte lineage represented by PC1 (Figure 8B and 8F). To characterize the changes along monocyte differentiation, we first computed pseudotime using Monocle 2 [26]. The pseudotime analysis was validated using genes extracted from an independent dataset that correlate with normal BM monocyte differentiation [33] (Figure 8C). In control mice, the changes of these genes were consistent with our computed pseudotime confirming that our pseudotime is biologically relevant (Figure 8D).

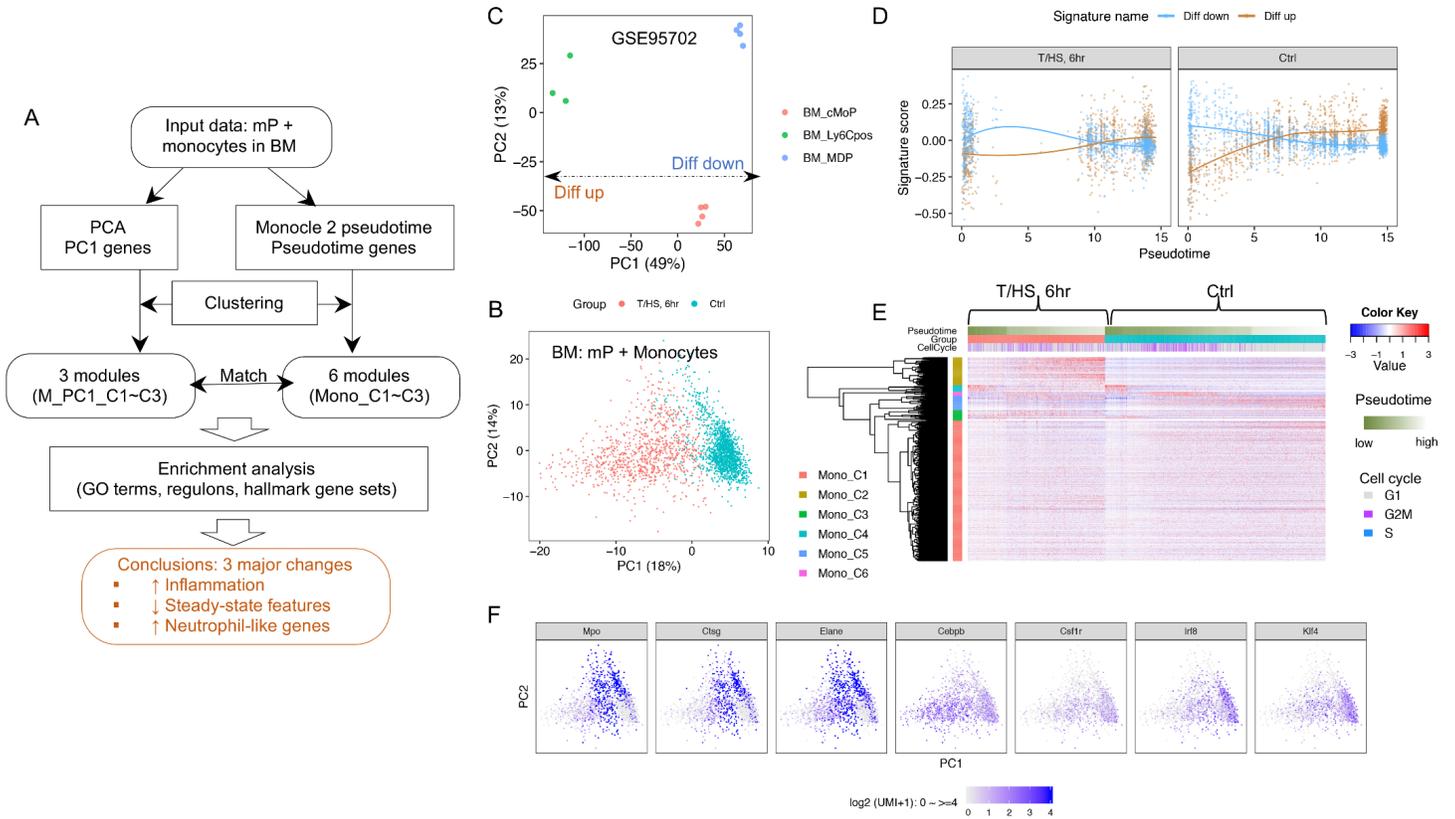


Figure 8 Overview of transcriptomic changes in the BM monocyte lineages at 6 hrs after T/HS.

(A) Schema describing the workflow for Figure 8-10. (B) PCA plot of BM monocyte lineage (including all mP, as shown in the 4th panel of Figure 4B) color coded by groups. (C) To validate the computed pseudotime, we extracted genes positively and negatively associated with BM monocyte differentiation from the GSE95702 dataset (Supplemental Methods). (D) Signature scores were calculated for each single cell shown in PCA plot in (B) and plotted along pseudotime. Smoothing lines were fitted by Loess regression. (E) RNA profile of the BM monocyte lineage built upon pseudotime-associated genes identified by Monocle 2. Cells (columns) are ordered first by groups then by pseudotime. Genes (rows) are clustered into six gene modules (Mono_C1~C6). (F) Expression of representative markers are shown in the PCA plot as shown in (B).

We next established the RNA profile of pseudotime-associated genes and identified six gene modules (Mono_C1-C6) by Ward clustering (Figure 8E). To interpret biological functions and reveal critical regulatory networks, we performed an enrichment analysis using the MSigDB

[42] gene ontology and hallmark gene sets and computed regulons using SCENIC [27] (Table 2 and Figure 9). Mono_C2, which was *Cebpb* (C/EBPb protein coding gene)-regulon associated was up-regulated after T/HS and increased continuously along differentiation (Figure 9A). This is consistent with the known roles for *Cebpb* in emergency myelopoiesis [43, 44]. The CD11b coding gene, *Itgam*, is also included in Mono_C2 (Table 2), and this corresponded to changes observed in circulating cells (Figure 2E). *Klf4* and *Irf8* are TFs critical for steady state monopoiesis [45]. *Irf8* is also responsible for monocyte lineage commitment [39]. Mono_C1 and Mono_C5, were associated with *Klf4* and *Irf8* regulons, respectively (Figure 9D). The monocyte lineage marker *Csf1r* (gene coding CD115), was included in Mono_C1 (Table 2). Unexpectedly, both Mono_C1 and Mono_C5 were down regulated after T/HS (Figure 9A). The changes in the gene expression of these critical TFs after T/HS were consistent with the corresponding changes in regulon expression (Figure 9B-9C), further supporting the results of the regulon enrichment analysis. We also identified a cell-cycle module (Mono_C6), a progenitor module (Mono_C4) and a stress-responsive module (Mono_C3) (Table 2 and Figure 9).

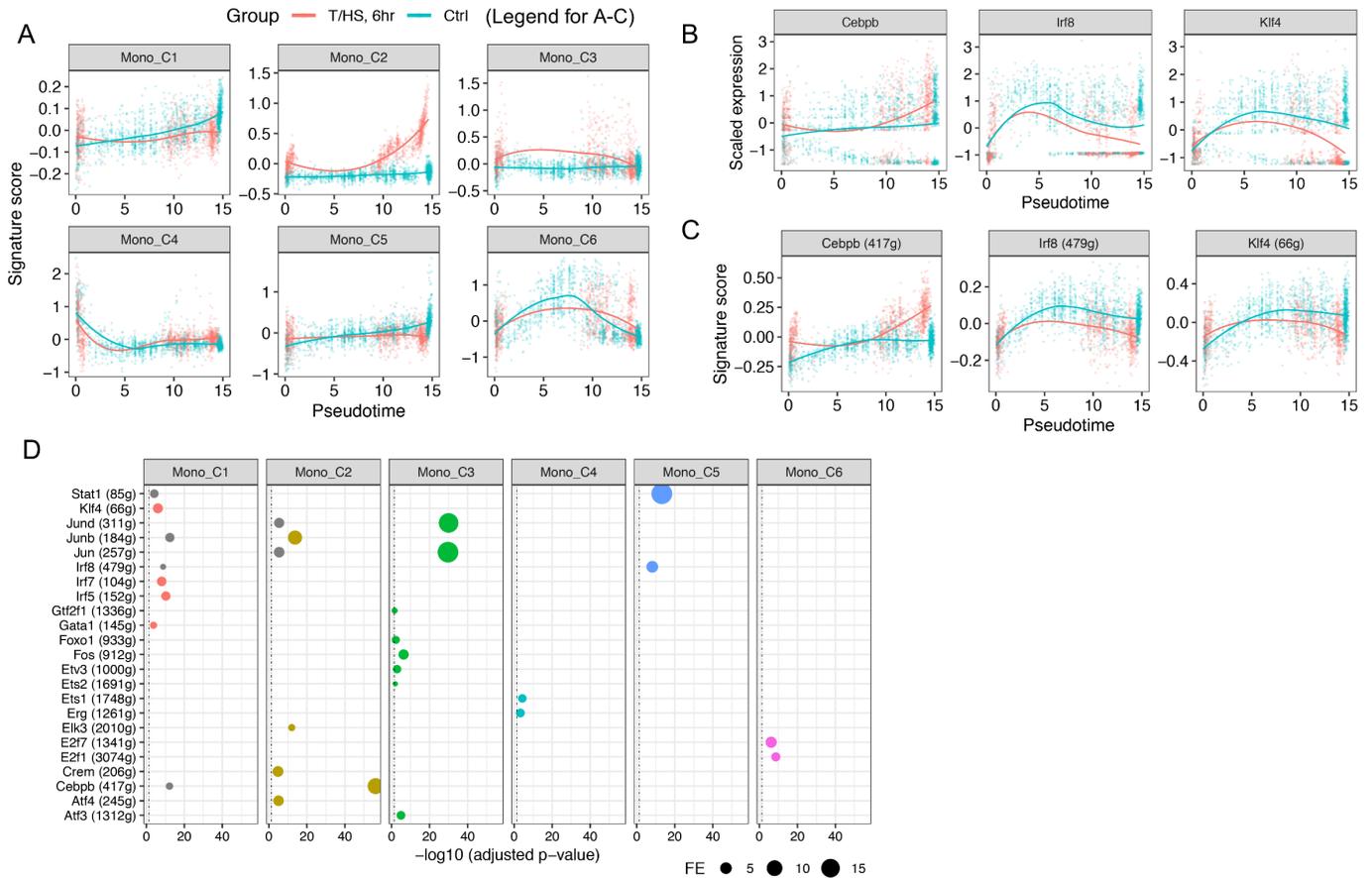


Figure 9 Characterization of transcriptomic changes in the BM monocyte lineages at 6 hrs after T/HS.

(A-C) Expression of each gene module (A), critical TF (B) and corresponding regulon (C) along pseudotime. Smoothing lines were fitted by Loess regression. (D) Enrichment analysis between gene modules and regulons. Hypergeometric p-value was computed. Only the relationships with Benjamini-Hochberg adjusted p-value < 0.05 (labeled as black vertical dashed line) with fold enrichment (FE) ≥ 2 and the number of overlapping genes ≥ 15 are shown. Relationships are color coded by top enriched gene modules (with highest FE).

Table 2 Brief summary of gene modules identified from mouse bone marrow monocyte lineage.

Modules	Representative genes	Regulons	Biological functions
Mono_C1	<i>Csf1r, Cd74, H2-Aa, Ear2, Ly6i</i>	<i>Klf4, Irf5</i>	Lymphocyte activation, IL-12 production
Mono_C2	<i>Itgam, Cd14, Thbs1, Cebpb, Mmp8, Sell, Hif1a, Mafb</i>	<i>Cebpb, Junb</i>	Inflammation, chemotaxis
Mono_C3	<i>Junb, Fos, Fosb, Hsp90aa1</i>	<i>Jund, Jun</i>	Response to stress
Mono_C4	<i>Mpo, Ctsg, Elane, Lcn2, Ltf</i>	<i>Erg, Ets1</i>	Progenitor or neutrophil associated genes
Mono_C5	<i>Irf8, Klf4, Stat1, Irf7, Mx1</i>	<i>Irf8, Stat1</i>	Steady-state monopoiesis
Mono_C6	<i>Tubb4b, Top2a, Cenpa</i>	<i>E2f7, E2f1</i>	cell cycle

To characterize the major differences in BM monocytes between control and T/HS, we extracted PC1-associated genes from the PCA shown in Figure 8B (Pearson's correlation: adjusted p-value < 0.05 and $|r| \geq 0.3$) and identified three gene modules (MonoPC1_C1-C3) (Figure 8A and Figure 10A). MonoPC1_C1 corresponded to steady-state module Mono_C5. MonoPC1_C2 related to inflammatory module found in Mono_C2 and mP_C2. In addition, we identified a Neutrophil-like module MonoPC1_C3 (e.g. *S100a8, S100a9, Ltf, Lcn2*) highest expressed in neutrophils and up-regulated in monocyte lineage after T/HS, which is consistent with the monocyte-to-neutrophil shift observed in multi-Lin mP and can be mapped to mP_C1 (Figure 10B-10C). These three modules derived from bone marrow monocytes were also reflected in the circulating monocytes (Figure 10D).

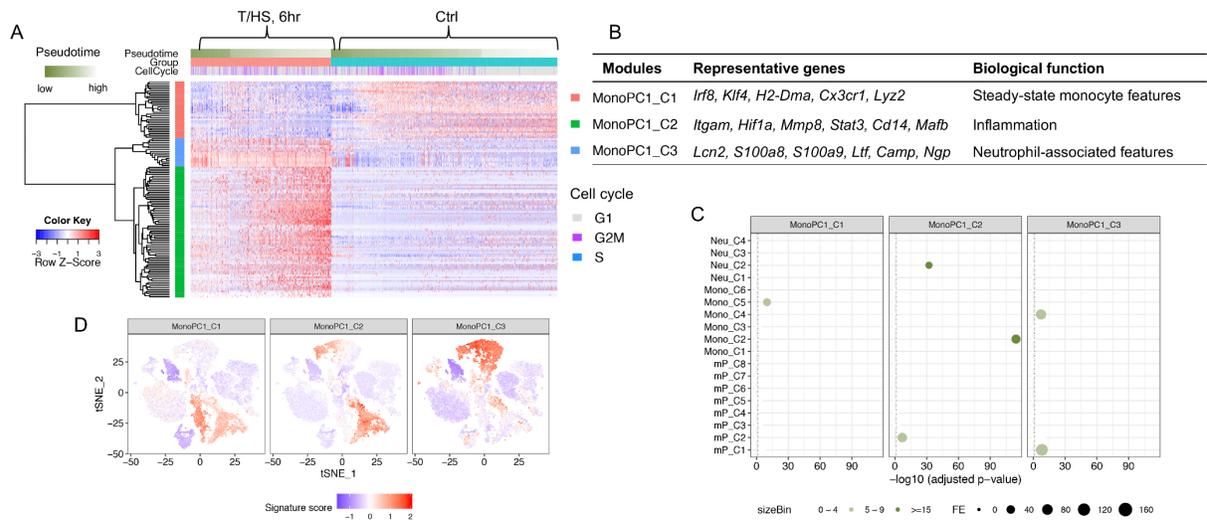


Figure 10 Transcriptomic changes in the BM monocyte lineages at 6 hrs after T/HS, Cont'd.

(A) Extraction of the major differences between groups from the PC1 as shown in Figure 8B. Ward clustering yielded three gene modules (MonoPC1_C1~C3) of PC1-associated genes (Pearson's correlation: adjusted p-value < 0.05 and $|r| \geq 0.3$). (B) Summary of the three gene modules identified in (A). (C) Enrichment analysis between the three gene modules and mP/monocyte/neutrophil gene modules identified at 6 hrs in mouse BM. Relationships are colored by binned number of overlapping genes. Only the relationship with ≥ 5 overlapping genes are visible. Mono_C1-C6 were characterized in Figure 8-9, Neu_C1-C4 in Figure 12 and mP_C1-C8 in Figure 7. (D) Visualization of the signature scores in the t-SNE plots.

We validated the unexpected changes in two critical markers (monocyte lineage marker CD115/*Csflr* and lineage TF *Irf8*) at the protein level in BM monocytes. Since CD115 gene expression was suppressed after T/HS, we used two gating strategies when selecting the cells in the monocyte lineage that included: (1) $\text{Lin}^- \text{CD115}^+ \text{Ly6G}^-$ and (2) $\text{Lin}^- \text{Ly6C}^{++} \text{Ly6G}^-$ (Figure 11A). In control mice, the cells gated by the two strategies appeared identical. Following T/HS, the proportion of $\text{Lin}^- \text{CD115}^+ \text{Ly6G}^-$ defined monocytes was significantly lower than that of $\text{Lin}^- \text{Ly6C}^{++} \text{Ly6G}^-$ defined monocytes, due to a significant decrease in CD115 protein expression. $\text{Lin}^- \text{Ly6C}^{++} \text{Ly6G}^-$ gated cells expressed less IRF8 and became more CD11b positive. To visualize the

colocalization of protein changes in Lin⁻Ly6C⁺⁺Ly6G⁻ gated cells, we chose Matlab/Cyt3 [25]. Similar to scRNA-seq, monocytes from the two experimental groups were largely separated in 2D dimension reduction space (first panel in Figure 11B). The higher CD11b expression and suppressed protein levels of IRF8 and CD115 were colocalized in most of the monocytes after T/HS (circled area in Figure 11B). Combined, the findings from the transcriptomics, regulatory networks and multi-dimensional protein analyses consistently characterized known and novel features in T/HS-induced monocytes compared to steady state monocytes. Therefore, circulating monocytes after T/HS are not a simply an immature version of steady state but instead a new myeloid trajectory evident at the mP level.

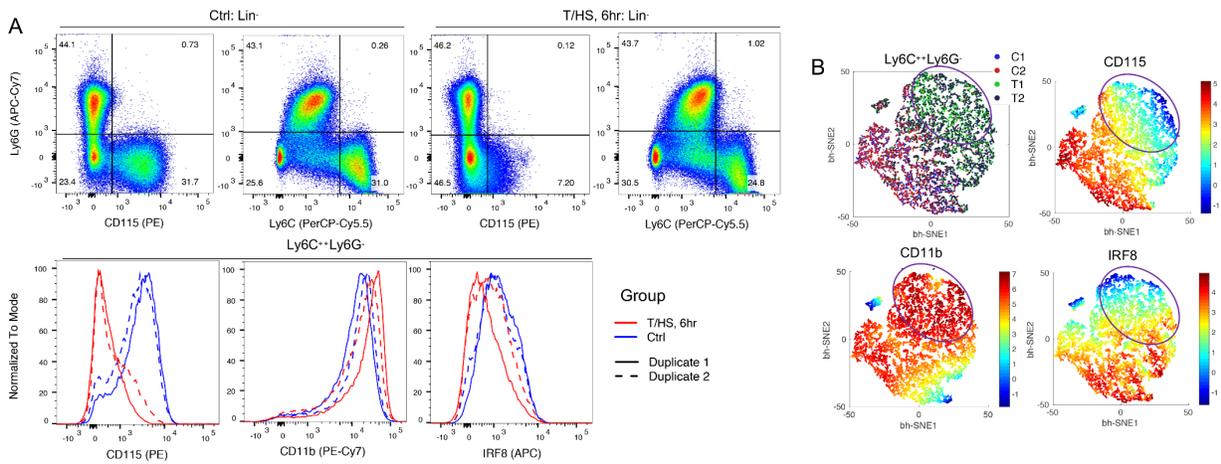


Figure 11 Validation of transcriptomic changes in the BM monocyte lineage at 6 hrs after T/HS by flow cytometry.

The results from one experiment (2 mice/group) are shown, which has been repeated for two extra times. (A) FlowJo analysis. (B) Matlab/Cyt3 analysis of Lin-Ly6C⁺⁺Ly6G⁻ monocytes (C: Ctrl; T: T/HS, 6hr).

2.3.5 Characterization of the transcriptomic changes in the BM neutrophil lineage after T/HS

To characterize the full developmental trajectory during emergency neutropoiesis, BM mP and BM neutrophils were analyzed together (Figure 12A-12B). Using the same scheme outlined in Figure 8A, we identified 4 gene modules (Neu_C1-C4) with differential expression along pseudotime (Figure 12C). Similar to emergency monopoiesis, Neu_C2, was enriched in *Cebpb* regulons and associated with inflammation. After T/HS, this module continued to increase during neutrophil differentiation (Figure 12D-12F). The well-established granulocyte-lineage TF *Cebpe* [39] and surface marker *Ly6g*, were associated with Neu_C1. Even though *Cebpe* mRNA and regulon expression trended downward after T/HS (Figure 12G-12H), the steady-state module represented by Neu_C1 was not obviously suppressed (Figure 12F). The surface expression of Ly6G in BM neutrophils was also comparable before and after T/HS (Figure 11A). This is in contrast to monocytes where the steady state features were obviously suppressed after T/HS.

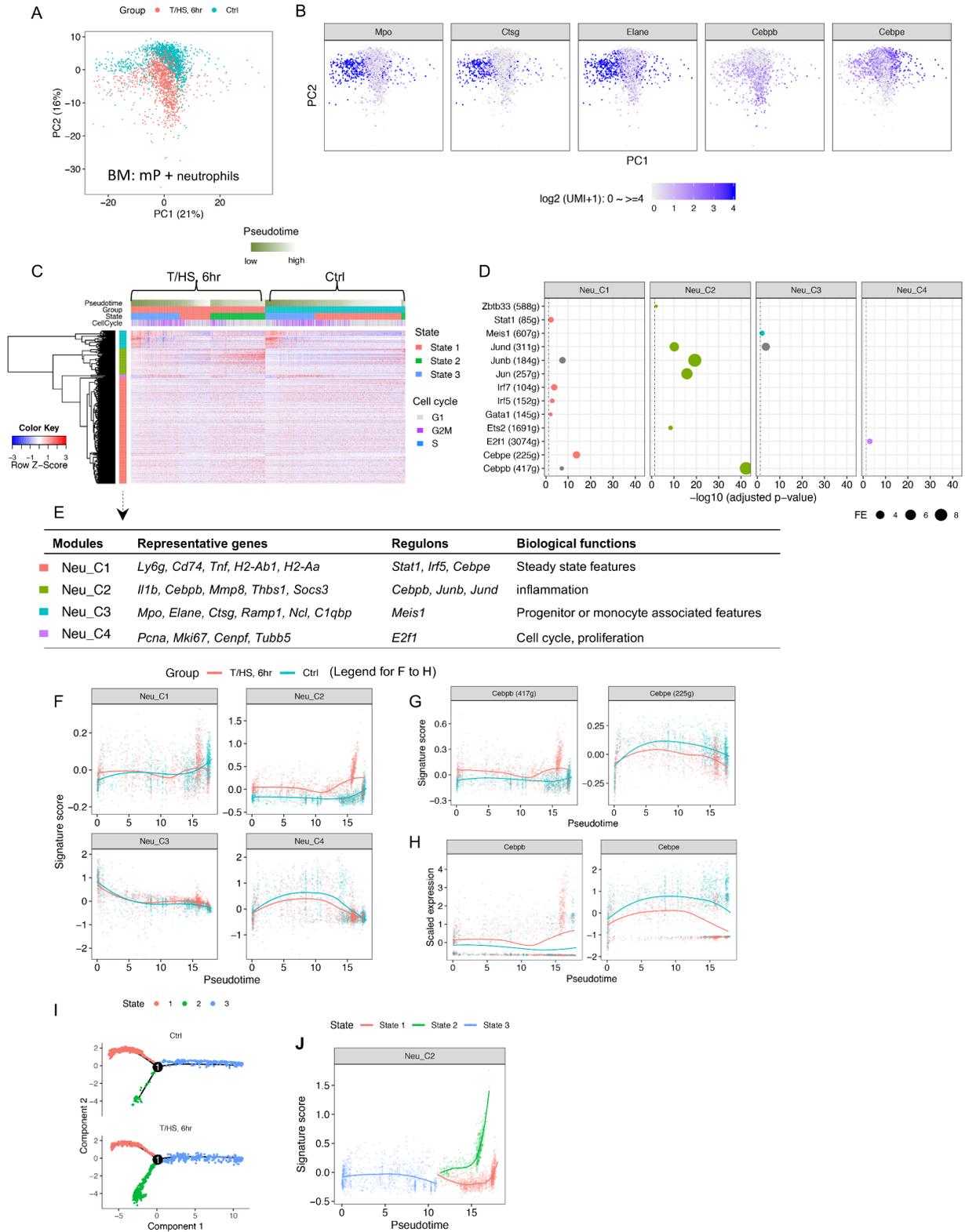


Figure 12 Characterization of transcriptomic changes in the BM neutrophil lineage at 6 hrs after T/HS.

(A) PCA plot of BM neutrophil lineage including all myeloid progenitors as shown in the 4th panel of Figure 4B, and expression of representative markers are shown in (B). (C) RNA profile of the BM neutrophil lineage is built upon pseudotime associated genes computed by Monocle 2. Cells (columns) are ordered first by groups, then by states identified using Monocle2 and lastly by pseudotime (from progenitors to committed cells). Genes (rows) are clustered by Ward clustering into four gene modules (Neu_C1~C4). (D) Enrichment analysis between the four gene modules and regulons. Hypergeometric p-value was computed. Only the relationships with Benjamini-Hochberg adjusted p-value < 0.05 with fold enrichment (FE) ≥ 2 and the number of overlapping genes ≥ 15 are shown. Relationships are color coded by top enriched gene modules (highest FE). (E) Summary of the four gene modules. (F) Expression of the four gene modules, (G) critical regulons and (H) corresponding TFs along pseudotime color coded by groups. Smoothing lines were fitted by Loess regression. (I) Trajectories constructed by Monocle2 are color coded by states and wrapped by groups. (J) Expression of Neu_C2 along pseudotime color coded by states.

Monocle [26] identified 3 states in the BM neutrophil lineage. State 3 corresponds to the progenitor state, and the other 2 states were neutrophil-committed states. Steady-state BM neutrophils were mostly in state 1. After T/HS, the proportion of the cells in state 2 increased dramatically, with only a small group of cells remaining in state 1 (Figure 12C and 12I). State 2 is characterized by high expression of the inflammatory module Neu_C2 (Figure 12J). Thus, we defined state 2 as the stimulated state and state 1 as the unstimulated state. T/HS induced a significant shift to the stimulated state during neutropoiesis.

Taken together, our analysis of monocytes and neutrophils after T/HS in mice indicates that these cells become more inflammatory represented as the up-regulation of the inflammatory modules Mono_C2 and Neu_C2. Compared with neutrophils, monocytes displayed an earlier BM branching point and more pronounced changes, characterized as partially losing steady-state features and also gaining neutrophil-associated genes. We also assessed for the presence of transcriptomic signatures described for MDSC [34] and found these to be simultaneously up-

regulated in both monocyte and neutrophil lineages after trauma (Figure 13). The gene modules and changes in TFs derived from the 6h time point in mouse BM were present at 3h and partially preserved to the 24h time point following injury (Figure 14).

A

GSE21927

Strains	MDSC	compared with	Control
BALB/c	CD11b+ cells from bone marrow cultured with GM-CSF+IL-6		CD11b+ cells from healthy spleen
BALB/c	CD11b+ cells from 4T1 tumor bearing mice		CD11b+ cells from healthy spleen
BALB/c	CD11b+ cells from C26-GM tumor bearing mice		CD11b+ cells from healthy spleen
C57BL/6	CD11b+ cells from EL-4 tumor bearing mice		CD11b+ cells from healthy spleen
C57BL/6	CD11b+ cells from MCA203 tumor bearing mice		CD11b+ cells from healthy spleen



- Genes up-regulated in MDSCs were defined as MDSC signatures.
- Total 5 signatures for different sources of MDSCs.

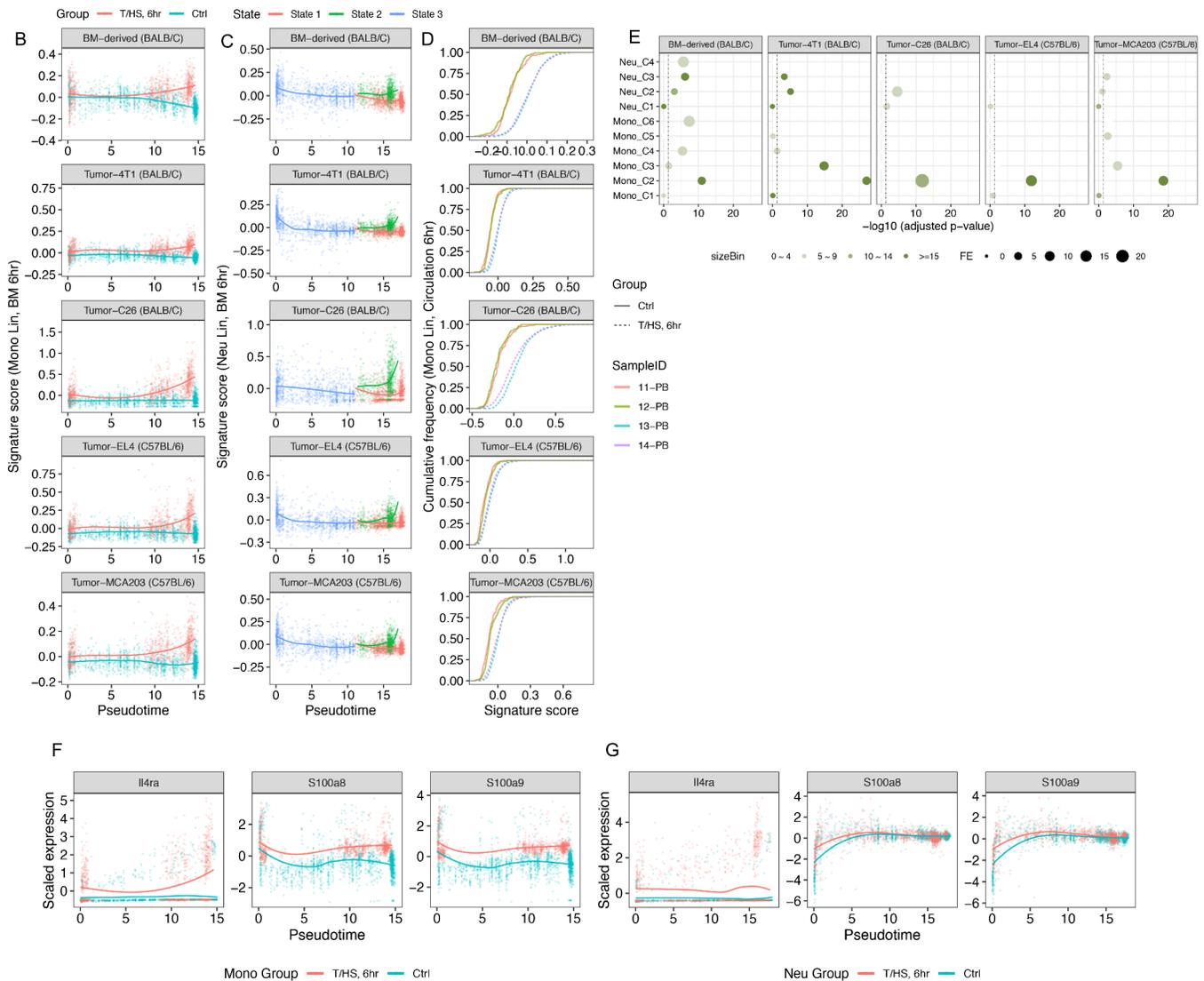
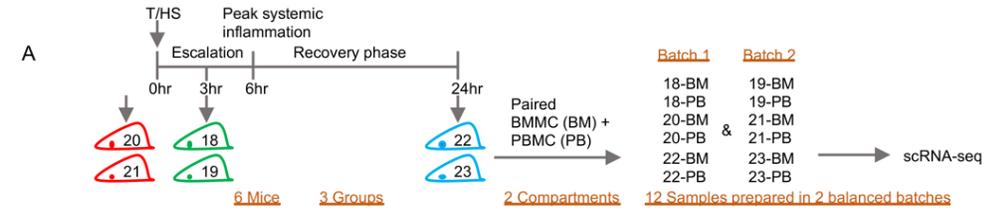


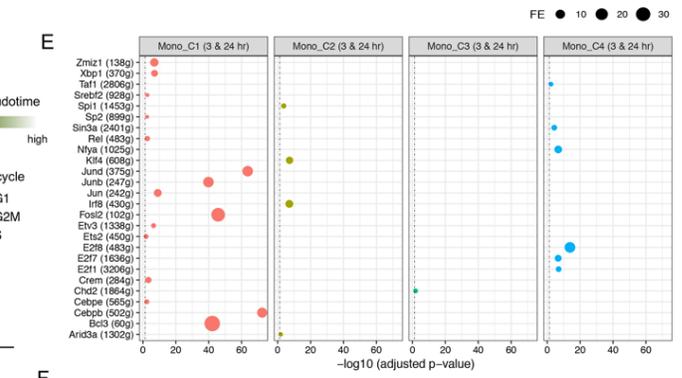
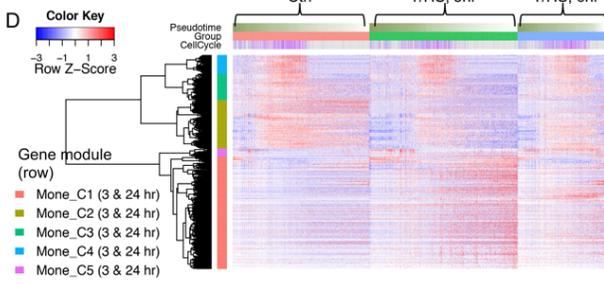
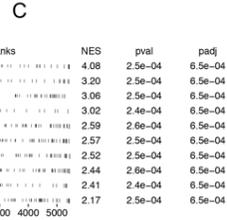
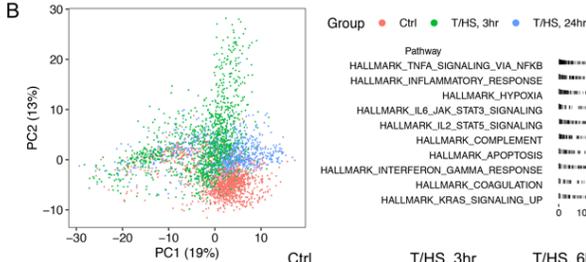
Figure 13 Simultaneous initiation of inflammatory and features of MDSC during emergency myelopoiesis.

(A) Development of MDSC gene signatures. (B-C) MDSC signatures are up-regulated in BM monocytes after T/HS (B) and in BM neutrophil stimulated state (state 2) (C) along pseudotime. A smoothing line was fitted by Loess

regression. (D) MDSC signatures are up-regulated in circulating monocytes after T/HS. Cumulative frequencies of signature scores (CDF curves) are shown. (E) Enrichment analysis between MDSC signatures and each gene module identified from BM monocyte and neutrophil lineages at 6hrs after T/HS. Hypergeometric p-value was computed. Relationships are color coded by binned number of overlapping genes. Only relationships with ≥ 5 overlapping genes are visible. The MDSC signatures were significantly enriched in the inflammatory modules (Mono_C2 and Neu_C2, *Cebpb* regulon associated), especially Mono_C2. (F-G) RNA expression of some functional markers associated with MDSC in BM monocytes (F) and BM neutrophils (G) after T/HS.

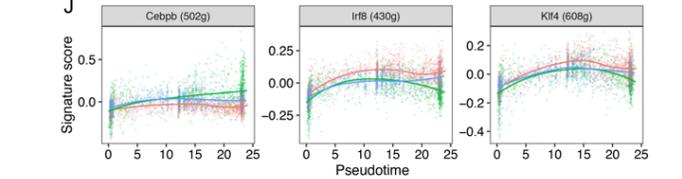
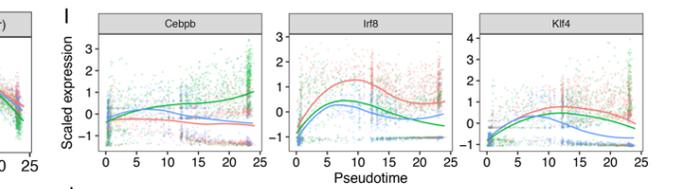
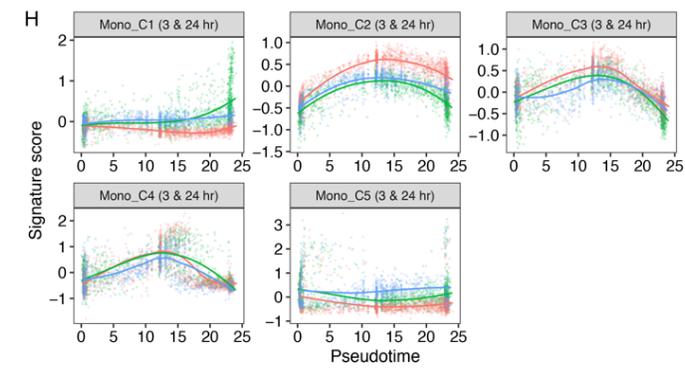
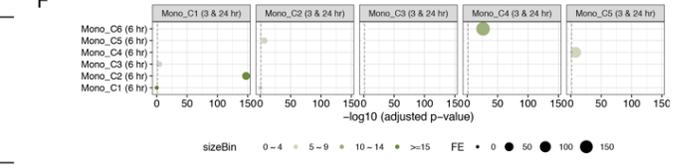


• BM monocyte lineage: 3hr & 24hr



G

Gene module (3 & 24 hr)	TF regulons	Gene modules (6 hr)
Mono_C1 (3 & 24 hr)	<i>Cebpb, Junb, Jund, Xbp1</i>	Mono_C2
Mono_C2 (3 & 24 hr)	<i>Irf8, Klf4</i>	Mono_C1 & C5
Mono_C3 (3 & 24 hr)	<i>Chd2</i>	N/A
Mono_C4 (3 & 24 hr)	<i>E2f8</i>	Mono_C6
Mono_C5 (3 & 24 hr)	N/A	Mono_C4



• BM neutrophil lineage: 3hr & 24hr

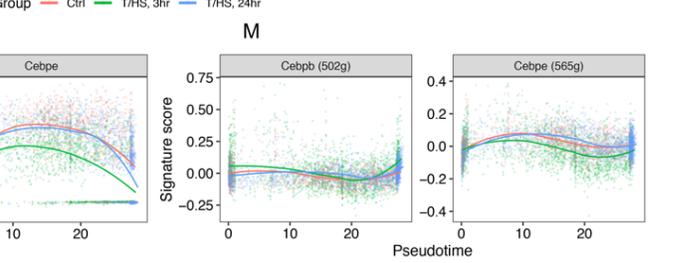
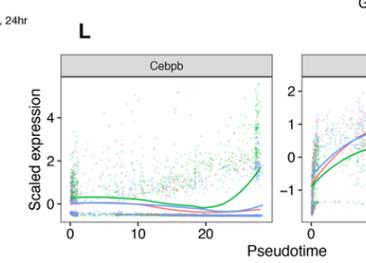
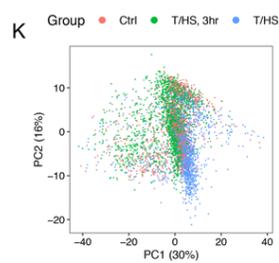


Figure 14 Characterization of transcriptomic changes in the BM monocyte and neutrophil lineages at 3 hrs and 24 hrs after T/HS.

(A) Experimental design of isolating paired BMDCs + PBMCs at 3 hrs and 24 hrs for scRNA-seq. (B-J) Analysis of BM monocyte lineage at 3hrs and 24hrs after T/HS. (B) PCA plot of BM monocyte lineage. The changes induced by T/HS are represented by PC2. (C) Top 10 hallmark pathways enriched in the positive side of PC2 by GSEA. (D) RNA profile was built upon PC2-associated genes (Pearson's correlation: $|r| \geq 0.2$ and adjusted p-value ≤ 0.05). Cells (columns) are ordered first by groups then by pseudotime. Cell cycle stages are also labeled. Genes (rows) are clustered into 5 gene modules, Mono_C1~C5 (3 & 24 hr). (E) Enrichment analysis between identified gene modules and regulons. Regulons were computed based on the myeloid cells including HSC/MPP and myeloid progenitors at 3hrs and 24hrs. Hypergeometric p-value was computed. Only the relationships with adjusted p-value < 0.05 and fold enrichment (FE) ≥ 2 and the number of overlapping genes ≥ 15 are shown. (F) Map the gene modules identified at 3 & 24 hrs to the ones at 6 hrs (enrichment analysis). Relationships are colored by binned number of overlapping genes. (G) The gene modules derived from the 6hr time point were largely recapitulated at 3hrs and 24hrs. (H-J) Expression of each gene module (H), critical TF (I) and corresponding regulon (J) along pseudotime. Smoothing lines were fitted by Loess regression. (K-M) Analysis of BM neutrophil lineage at 3hrs and 24hrs after T/HS. (K) PCA plot of BM neutrophil lineage. Although BM neutrophils did not show obvious global transcriptomic changes 3 & 24hrs, the critical TFs (L) and corresponding regulons (M) followed a similar trend to those seen at 6hrs shown in Figure 12.

2.3.6 Overview of the time-dependent transcriptomic changes in PBMCs from trauma patients

To extend our landscape analysis into the human response to injury, we isolated PBMCs from 10 trauma patients (Table 3) sampled at 3 time points (<4h, 24h, 72h) after systemic injury. Each patient was paired with an age- and sex-matched healthy control subject (Figure 15A). The 72h samples from two patients were not available leaving of 38 samples subjected to single-cell RNA sequencing. A total of 151,470 immune cells passed initial quality control, with a mean of 1757 genes/cell.

Table 3 Trauma Patient Clinical Characteristics.

Patient ID	ISS	Age	Sex	Injury	Admission chemistry				
					SBP	HR	GCS	BD	Lactate
MM3001	22	21	Male	MVC/motorcycle	71	127	15	12	4.8
MM3005	22	35	Male	MVC/PED	60	112	3	15	7
MM3008	19	26	Male	Penetrating	70	110	15	12	6.9
MM3009	13	71	Male	Fall	88	130	15	5	5.8
MM3012	17	56	Male	MVC	125	111	14	NA	3.7
MM3015	38	21	Male	Fall	45	74	3	17	3.4
MM3016	18	32	Male	MVC	127	148	15	19	2.4
MM3020	27	78	Male	Fall	83	51	8	11	NA
MM3038	22	44	Female	MVC	112	70	3	12	5.6
MM3040	25	58	Male	Fall/TBI/SDH	138	82	3	4	22

Abbreviations:

ISS: Injury severity score; HR: Heart rate; GCS: Glasgow Coma Scale/Score; BD: Base deficit; MVC: Motor vehicle collision; PED: Pedestrian; TBI: Traumatic brain injury; SDH: Subdural hematoma; NA: Not available.

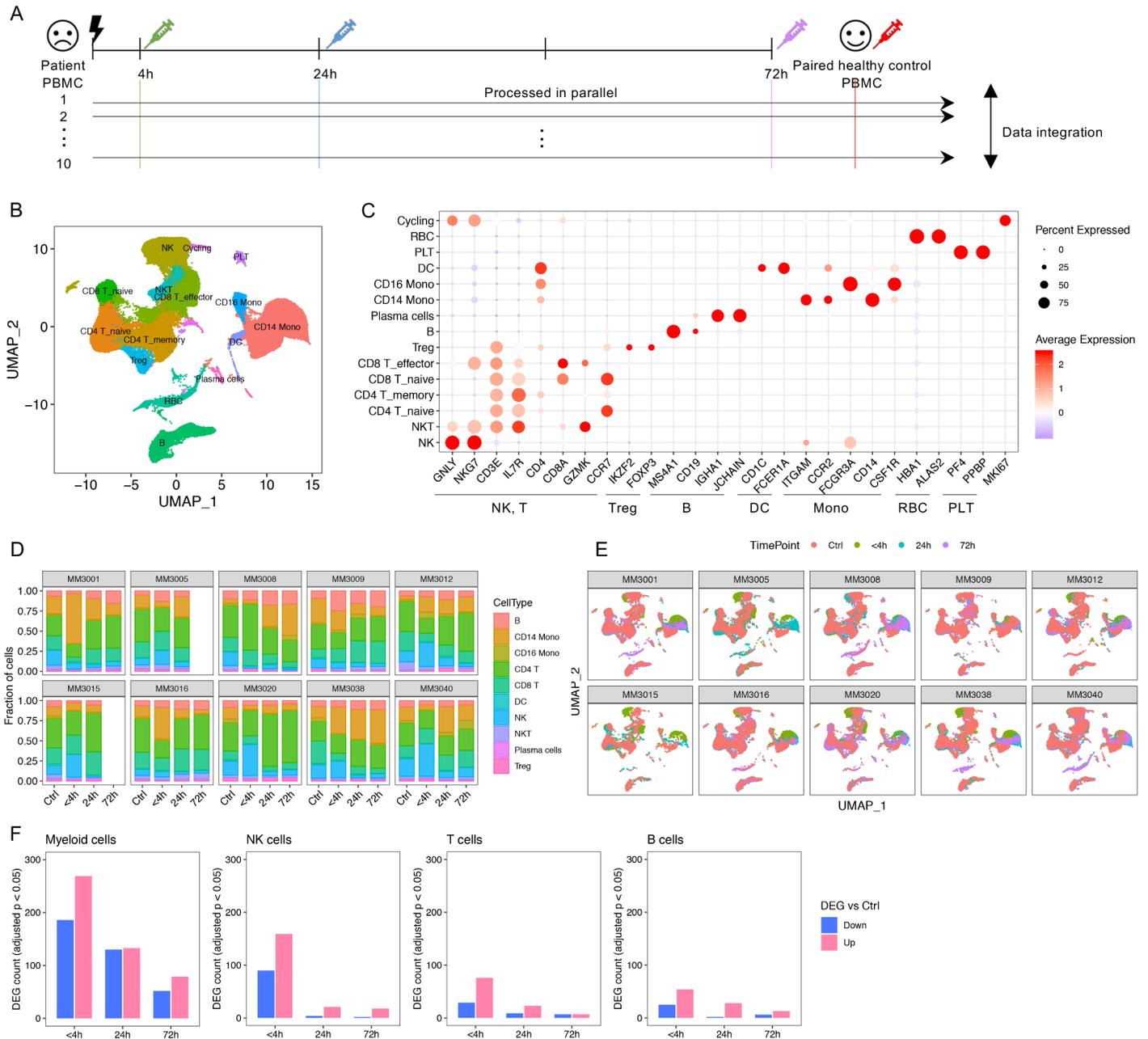


Figure 15 Overview of the transcriptomic changes in PBMCs from trauma patients over time.

(A) Experimental design for human scRNA-seq experiments. Blood samples for PBMCs isolation were obtained within 4hrs of injury and at 24hrs and 72hrs after injury from 10 patients. Blood drawn from a healthy age and sex matched uninjured volunteer was used to establish the baseline for each patient. The 72hr samples from 2 patients are not available, for a total of 38 samples (Ctrl: n=10; <4h: n=10; 24h: n=10; 72h: n=8). (B) UMAP plot of all human

PBMCs are color coded by major cell types. (C) Expression of major lineage markers in each cell type shown in (B). (D) Changes of cell type composition in each patient along with matched control subject. (E) UMAP plot as shown in (B) wrapped by patients and color coded by time points. (F) The number of significant DEGs (compared with healthy control, Bonferroni adjusted p-value < 0.05) at different time points in major cell types.

The common PBMCs cell types were easily distinguished in all 38 samples with clear differences in transcriptomic patterns (Figure 15B-15C). The frequencies of major cell types in healthy controls were consistent with a previous report [46]. After trauma, changes in cell frequencies were variable between patients (Figure 15D). However, similar to mouse the data, the myeloid compartment (mostly monocytes) exhibited a significant and time-dependent shift shown by UMAP (Figure 15E) and contained the largest number of DEGs among the major cell types (Figure 15F). NK cells displayed the second largest number of DEGs followed by lymphocytes. The greatest changes were observed at the <4h time point.

2.3.7 Characterization of the transcriptomic changes in human circulating monocytes

Because there were no clear boundaries between dendritic cells (DC), CD14⁺ monocytes, and CD16⁺ monocytes in the UMAP plot (Figure 15B), we analyzed these cell populations separately to deconvolute the circulating monocytes at a higher resolution (Scheme shown in Figure 16A). This yielded 14 clusters of myeloid cells that included, 1 DC (cluster 8), 1 CD16⁺ (cluster 4), 1 cycling myeloid cells (cluster 13), 1 monocyte-platelet aggregates (cluster 9), 3 CD14^{int} monocytes (clusters 6, 10 and 12) and 7 CD14⁺ monocytes (clusters 0, 1, 2, 3, 5, 7 and 11). The CD14⁺ monocyte population displayed the most striking differences across time (Figure 16B-16D). To order the clusters in a more biologically meaningful way, we designated each cluster

based on assigned cell cluster number (CC#), major cell subset and enrichment time point (Figure 16E). If a cluster tended to be distributed evenly along time series (highest odd ratio ≤ 2.5), it was labeled as “pan”. The clusters were ordered first by monocyte developmental orders (CD14⁺→CD14^{int}→CD16⁺) [47] and then by time series (4h→24h→72h→control) (Figure 16F-16G).

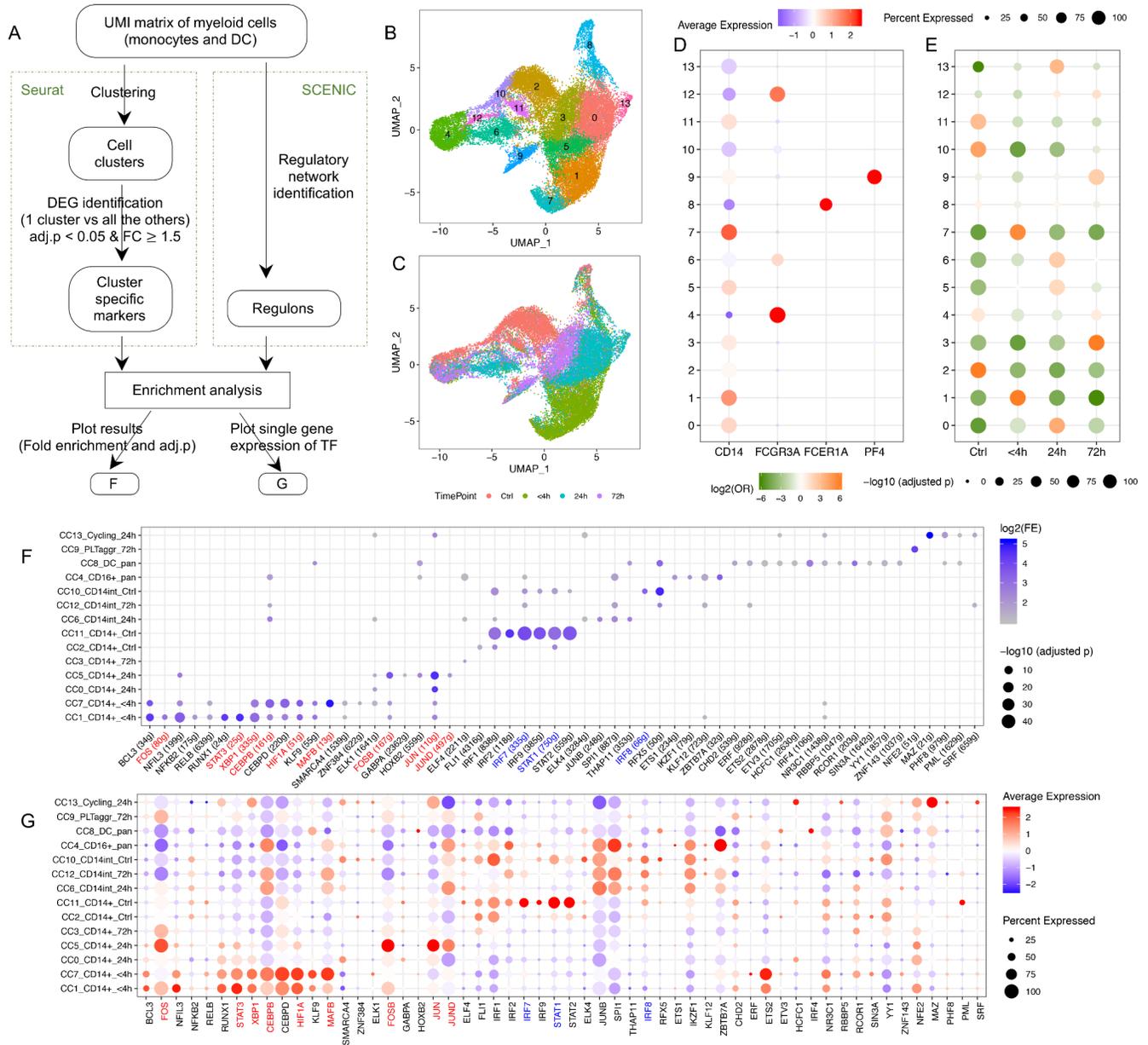


Figure 16 Characterization of the transcriptomic changes in human circulating monocytes after trauma.

(A) Schema describing the workflow for this figure. (B-C) UMAP plot of all human peripheral blood mononuclear myeloid cells color coded by identified clusters (B) or by time points (C). (D) Expression of monocyte, dendritic cell and platelet representative markers. (E) Overlap between myeloid clusters and time points were evaluated by χ^2 test. Two-sided p values were computed and adjusted by the Benjamini-Hochberg method. OR: odds ratio. (F) Enrichment analysis between cluster specific markers and regulons. Hypergeometric p-value was computed. Only the relationships

with Benjamini-Hochberg adjusted p-value < 0.05 with fold enrichment (FE) ≥ 2 and the number of overlapping genes ≥ 5 are shown. (G) Gene expression of enriched TFs corresponding to the regulons shown in (F). Color coded TFs or regulons in (F-G) are the ones also identified in mouse monocytes. Red: up-regulated; Blue: down-regulated after trauma.

To prioritize critical transcription factors (TFs), we first computed SCENIC [27] regulons using all myeloid cells. Then, we performed enrichment analysis between regulons and identified top markers for each cluster (Figure 16A). Each cluster displayed distinct enriched regulons. There are roughly two blocks of TFs/regulons; one associated with CD14⁺ monocytes at early time points after injury, and another associated with CD14⁺ cells in the controls (Figure 16F). The expression of corresponding TFs was largely consistent with the regulon enrichment results (Figure 16G). Since regulon computation is independent of the Seurat integration workflow, this analysis provides independent confirmation for the identified clusters.

2.3.8 Generation of six CD14⁺ monocyte signatures

To further generalize the changes in CD14⁺ monocytes, we identified the pairwise DEGs between each pair of CD14⁺ monocyte clusters. This identified 129 genes above the threshold (adjusted p-value < 0.05 & fold change ≥ 2). These genes could be clustered into 6 signatures in which C1-, C2- and C3-associated genes were induced after trauma and C4, C5 and C6 represented the features dominant under steady state; these were suppressed after trauma (Figure 17A and Table 4). To biologically interpret these signatures, we performed enrichment analysis (Figure 17B and Table 5). C1-C3 were associated with the innate immune response, the response to oxygen containing compounds, the inflammatory response and regulation of hematopoiesis. Enriched

regulons included *JUN*, *STAT3*, *XBPI*, *MAFB*, *CEBPD*, and *CEBPB*, among others. C4 associates with an interferon (IFN)-specific program, highly enriched in the regulons of *STAT1*, *STAT2* and IFN regulatory factors (*IRF1*, *IRF7* and *IRF9*) [48]. It is noteworthy that a small but dominant population of CD14⁺ monocytes (cluster 11, Figure 16F) was highly associated with IFN signaling and significantly enriched in healthy controls. Many C5 genes are MHC II molecules and enriched in the regulon of MHC II activating TF *RFX5* [49]. There were only 8 genes in C6, which were not enough to identify significantly enriched GO terms or regulons. Thus, we generalized the changes in monocytes after systemic injury into six signatures with central TFs and biological interpretations.

Table 4 Six signatures derived from human CD14+ monocytes.

Signature Names	Genes
C1	<i>CLU, NKG7, S100A8, S100A9, LGALS1, S100A12, CTSD, RETN, RNASE2, PLAC8, PLBD1, FOLR3, JUN, ALOX5AP, HP, EGR1</i>
C2	<i>NAMPT, ACSL1, IL1R2, SOCS3, CD63, PIMI, CXCL8, VCAN, SLC2A3, AGFG1, CD55, SLC11A1, SAMSNI, MCEMP1, GCA, FKBP5, BCL2A1, SERPINB1, SLC25A37, CYP1B1, MCTP2, CCND3, ADM, G0S2</i>
C3	<i>FTH1, THBS1, EREG, MARCKS, SRGN, CD300E, TIMP1, CTSL, HMOX1, LITAF, CD163, GK, GLUL, HLA-DQA1, ASPH, AREG, IL1B, GK5, AQP9, PHC2, PLSCR1, GPR183, ETS2, CEBPB, CXCL2, SAP30, MAFB, FCGR1A, DDIT4, TPM4, MAP3K8, HIF1A, HLA-DRB5, ID2</i>
C4	<i>OAS2, EPSTI1, IFIT1, RNF213, PARP14, IFIT3, IFI44, XAF1, LY6E, IFI44L, MX1, ISG15, IFI6, STAT1, OAS3, IFIT2, MX2, HERC5, EIF2AK2, SAMD9L, RSAD2, APOBEC3A, OAS1, TNFSF10, GBP1, IFITM3, LAP3, TMEM123, MT2A, SP110, STAT2</i>
C5	<i>HLA-DRB1, HLA-DRA, HLA-DPA1, HLA-DPB1, FGL2, CD74, TXNIP, LGALS2, HLA-DMB, HLA-DMA, HLA-DQB1, DUSP6, CPVL, ZFP36L2, AP1S2, TGFBI</i>
C6	<i>NCF1, AHNAK, TAGLN2, CRIP1, JAML, RAB11FIP1, NUP214, LTA4H</i>

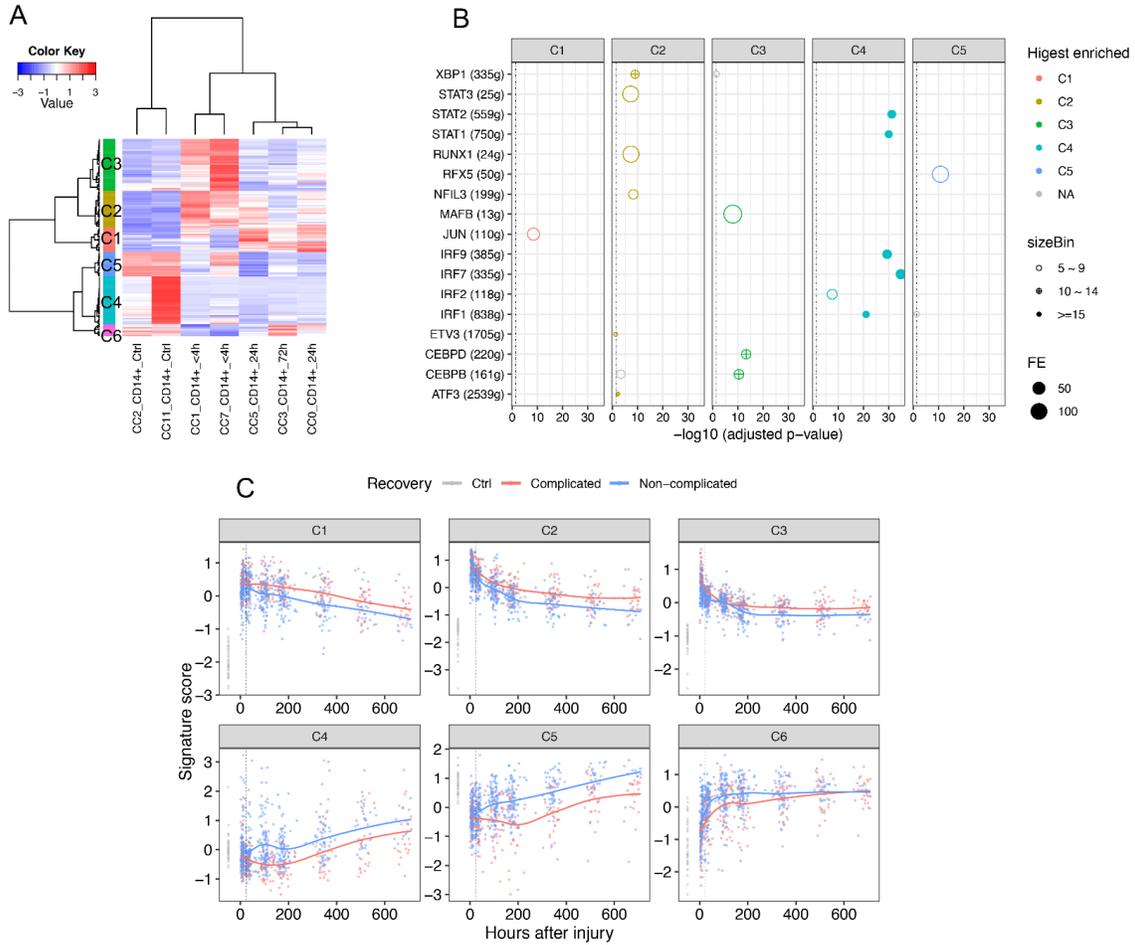


Figure 17 Generation and validation of six CD14+ monocyte signatures.

(A) RNA profile of pairwise DEGs (Bonferroni adjusted p-value < 0.05 and fold change ≥ 2) between seven CD14+ monocyte clusters. Columns represent the average gene expression for each cluster. Genes (rows) are clustered into six signatures (C1~C6). (B) Enriched regulons for the signatures shown in (A). Hypergeometric p-value was computed. Only the relationships with Benjamini-Hochberg adjusted p-value < 0.05 with fold enrichment (FE) ≥ 2 and the number of overlapping genes ≥ 5 are shown. Relationships are color coded by top enriched gene modules (with highest FE). (C) Validation of the six signatures in published trauma dataset (37 healthy controls vs. longitudinal data from 167 patients). Expression of each signature along timeline (up to 28 days after injury) are shown. Smoothing lines were fitted by Loess regression. The vertical dotted line labels the 24 hrs timepoint after injury.

Table 5 Selected enriched GO terms for human CD14+ signatures.

Signature name	Selected GO terms
C1	Defense response to bacterium
	Leukocyte activation
	Response to biotic stimulus
	Innate immune response
C2	Response to lipid
	Response to organic cyclic compound
	Response to oxygen containing compound
C3	Acute phase response
	Acute inflammatory response
	Inflammatory response
	Regulation of hemopoiesis
C4	Response to interferon alpha
	Response to interferon gamma
	Defense response to virus
C5	MHC class ii protein complex
	Lymphocyte co-stimulation
	Interferon gamma mediated signaling pathway

The human and mouse data are generally consistent in the monocyte compartment before and after trauma (Figure 18 and color-coded regulons/TFs in Figure 16F and 16G). The six human monocyte signatures can be mapped to mouse bone marrow gene modules. Human monocyte signatures C2 and C3 overlap with mouse inflammatory modules Neu_C2 and Mono_C2. *S100A8* and *S100A9*, both neutrophil-associated genes, are contained in human C1. Human C4 (IFN signaling) and C5 (MHCII signaling) correspond to mouse steady-state modules Mono_C5 (*Irf8*, *Irf7*) and Mono_C1 (lymphocyte activation, including *Cd74* and *H2-Aa*), respectively. Globally, shared DEGs derived from human and mouse monocytes after trauma display an intermediate level of correlation (Spearman correlation: $\rho > 0.6$).

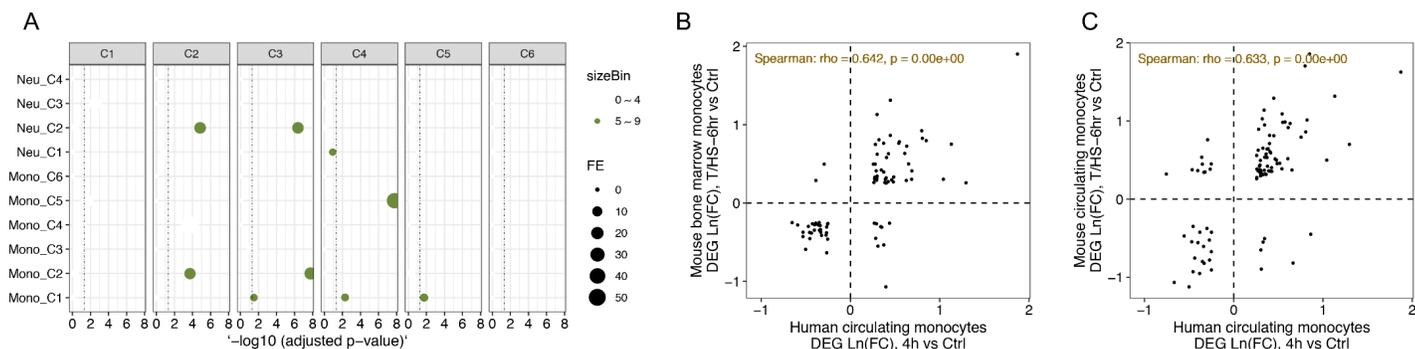


Figure 18 The human and mouse data are generally consistent in the monocyte compartment before and after trauma.

(A) Enrichment analysis between six human signatures (C1-C6) and the monocyte/neutrophil modules identified from mouse BM (Mono_C1-C6 & Neu_C1-C4) at 6hrs after T/HS. Hypergeometric p-value was computed and adjusted by Benjamini-Hochberg method. Relationships are colored by binned number of overlapping genes. Only relationships with ≥ 5 overlapping genes are visible. Correlation of monocyte DEGs identified from mouse BM vs. human circulating monocytes (B) or from mouse circulating vs. human circulating monocytes (C). FC: fold change.

Our findings in human PBMCs further extend the conclusions derived from the mouse T/HS model. Specifically, changes in monocytes do not simply involve a switch between two fixed trajectories representing steady state versus post trauma. Instead, monocytes shift away from steady state in a graded manner (Figure 16C): C1-C3 vs. C4-C6 correspond to the two distinct directions of the trajectory shifting away or toward steady state.

2.3.9 Validation of the six signatures in bulk RNA datasets

To provide confirmation for the gene signature patterns derived from our single-cell analysis, we queried a published dataset [6] of the transcriptomic changes of whole-blood

leukocytes from 167 severely injured humans in a 28-day time course, along with a baseline obtained from healthy controls. The changes in the six signatures could be largely recapitulated in this database (Figure 17C). Compared to healthy controls, C1-C3 were significantly induced and C4-C6 were suppressed to different degrees after injury. The response to systemic injury was most pronounced at the first sampling time point (~12 hours after injury) and then gradually returned to baseline. The changes were also associated with different clinical trajectories. Complicated cases (TTR >14 days) showed both higher magnitude and persistence of the changes compared with the non-complicated cases (TTR ≤ 14 days). It is noteworthy that the changes in six monocyte signatures in whole-blood leukocytes may also include the contribution of neutrophils, considering the similar changes between monocytes and neutrophils we observed in mouse bone marrow.

2.3.10 Two subtypes of trauma patients defined by the six signatures with differential prognostic value

To explore potential patient heterogeneity in the six-signature response, we clustered the patients based on a signature score matrix of the six signatures (Figure 19A). Because early identification of patients at risk for adverse outcomes could be clinically useful, we extracted the first sampled time point from all 167 trauma patients (mean ± standard deviation: 8.0 ± 3.4h). Trauma patients exhibited obvious heterogeneity at the early time point. The six signatures roughly clustered the patients into two subtypes (SG1 vs. SG2, Groups clustered based on Signature scores, Figure 19B). Compared with SG2, SG1 patients expressed higher C1-C3 and lower C4-C6, and experienced worse clinical outcomes including (Table 6), longer hospital length of stay, more severe multi-organ dysfunction, and higher incidence of infectious and non-infectious complications. Kaplan-Meier analysis demonstrated that SG1 patients underwent significantly

slower 28-day recovery than SG2 patients (Figure 19C). Well-established prognostic factors for trauma include injury severity [50], brain injury [51] and serum lactate levels [52]. To determine the influence of these factors, we assessed patient baseline characteristics (Table 6). Whereas many parameters were distributed evenly, injury severity score (ISS) and maximal serum lactate within 6h after admission were not. Considering the counts of myeloid cells from this dataset may not come from the same sample sent for microarray, we also deconvoluted myeloid composition using RNA data matrix by CIBERSORT [53]. Even though SG1 patients had higher ISS and lactate levels within 6h, multivariate analysis using a Cox model indicated that SG1 remained an independent risk factor for slower recovery (Figure 19D) after adjusting for these potential co-variants including myeloid composition ~12 hr after injury. These results suggest that patients have intrinsically different responses to systemic injury. Our findings provide additional information for differential prognosis that cannot be explained by injury severity or other known prognostic factors.

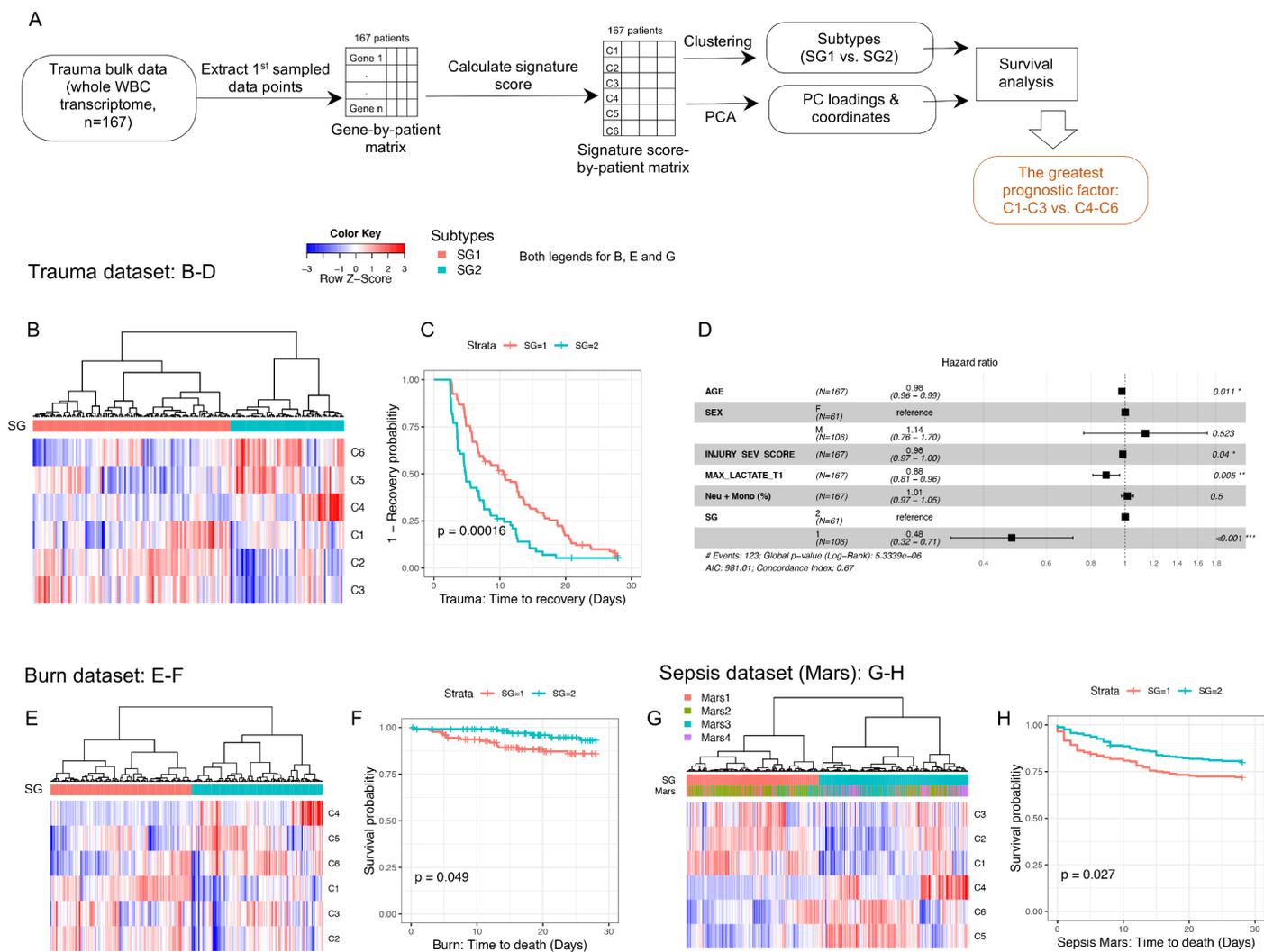


Figure 19 Six signatures define two patient subtypes associated with different prognosis.

(A) Schema describing the workflow for Figure 19-20. (B-D) Trauma dataset (n=167). (B) Trauma patients were clustered into two subtypes (SG1 vs SG2) using the signature score matrix. (C-D) Time-to-event analyses (event = recovery). (C) Kaplan-Meier curve was plotted by the two subtypes to visualize 28-day recovery. Log-rank p value is shown. (D) Hazard ratio of the subtypes after adjusting potential covariates using cox proportional hazards model. Compared with SG2 (shown as the reference), SG1 is significantly associated with slower recovery after adjusting for the potential co-variants. (E-H) Burn/sepsis patients were clustered into two subtypes, and Kaplan-Meier curve was plotted to visualize 28-day survival. Log-rank p value is shown. (E-F) Burn dataset (n=241). (G-H) Sepsis dataset (n=479).

Table 6 Basic characteristics and outcome parameters between two subtypes of trauma patients.

	SG1 (n=106)	SG2 (n=61)	p value
<i>Basic characteristics</i>			
Age	33.43 (10.8)	34.89 (11.62)	0.4540
Sex: Male	66 (62.3%)	40 (65.6%)	0.7400
Weight	85.84 (22.43)	89.2 (20.6)	0.1600
Height	174.2 (11.24)	174.44 (9.7)	0.8580
No Bad Head	93 (87.7%)	56 (91.8%)	0.6050
Injury Severe Score (ISS)	33.77 (12.29)	27.13 (14.05)	0.0020**
Apache II Score	27.74 (6.44)	26.66 (4.85)	0.0930
Transfused blood 0-12 hours since injury (ml)	2553.39 (2068.78)	2203.25 (1981.93)	0.0915
Total volume crystalloids received 0-12 hours since injury (ml)	13307.42 (7833.21)	12167.21 (5915.39)	0.7150
<i>Chemistry</i>			
Highest lactate 0-6 hrs	5.03 (2.25)	4.81 (3.95)	0.0499*
Highest lactate 6-12 hrs	4.07 (2.28)	3.27 (1.65)	0.0865
Highest lactate 12-18 hrs	3.18 (1.79)	2.45 (1.36)	0.0678
Highest lactate 18-24 hrs	2.92 (1.92)	2.57 (1.46)	0.6460
Max glucose 0-24 hrs	195.43 (63.41)	187.87 (45.87)	0.8820
Max insulin requirement 0-24 hrs	1.38 (2.07)	1.61 (2.26)	0.3860
Worst base deficit 0-6 hrs	-10.2 (4.64)	-9.14 (4.51)	0.1830
Worst base deficit 6-12 hrs	-4.85 (4.07)	-3.96 (4.35)	0.1640
Worst base deficit 12-18 hrs	-2.29 (3.76)	-1.18 (3.84)	0.1520
Worst base deficit 18-24 hrs	-1.25 (4.09)	-0.4 (3.39)	0.5040
<i>Cell fraction in white blood cells (%) (~12 hrs)</i>			
Myeloid cells (Neu + Mono)	0.86 (0.1)	0.87 (0.08)	0.7390
Lymphocytes	0.11 (0.08)	0.11 (0.07)	0.5730
<i>Outcome parameters</i>			
Survival	101 (95.3%)	59 (96.7%)	1.0000
Max Denver 2 Score	2.58 (2.11)	1.43 (1.7)	0.0001**
Hospital Length Of Stay	26.85 (18.57)	21.05 (16.65)	0.0096**
Max MOF ⁺ (Marshall Score)	6.07 (2.88)	4.5 (2.26)	0.0002***
Day of max Marshall Score	3.67 (4.53)	2.48 (2.74)	0.2470
Max MOF Neuro	3.6 (0.97)	3.72 (0.71)	0.7150
Max MOF Cardio	2.74 (0.97)	2.36 (1.05)	0.0269*
Max MOF Resp	2.01 (1.14)	1.27 (1.07)	<0.0001***

Table 6 continued

	SG1 (n=106)	SG2 (n=61)	p value
Max MOF Renal	1.06 (0.75)	0.82 (0.34)	0.0305*
Max MOF Hepatic	0.77 (1.1)	0.45 (0.81)	0.0669
Max MOF Hema	0.68 (0.65)	0.52 (0.67)	0.0601
Nosocomial Infections	70 (66.0%)	22 (36.1%)	0.0002**
Non-Infectious Complications	66 (62.3%)	20 (32.8%)	0.0004**
Surgical Site Infection	29 (27.4%)	8 (13.1%)	0.0349*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Continuous variables were shown as mean (standard deviation) and evaluated by Wilcoxon test. Categorical variables were shown as count (percentage) and evaluated by Fisher's exact test. Two-sided p values were computed.

[†]Max MOF Neuro: Worst central nervous system score over 28 days. Max MOF Cardio: Worst cardio score over 28 days. Max MOF Resp: Worst respiratory score over 28 days. Max MOF Renal: Worst renal score over 28 days. Max MOF Hepatic: Worst hepatic score over 28 days. Max MOF Hema: Worst hematologic score over 28 days. For all MOF scores, high is bad. Max MOF: The sum of the above 5 organ components without neurologic component, since this study precluded the cases with severe head injuries.

In addition to blunt trauma, burns and sepsis are common clinical problems that lead to acute critical illness. To determine if burns and sepsis result in the emergence of similar leukocyte gene expression patterns, we examined burn and sepsis datasets. We analyzed the first sampled time point after hospitalization (burn) or ICU (sepsis) admission. Similar to the trauma dataset, the burn and sepsis patients also fell into two subtypes. SG1 patients showed higher C1-C3 and lower C4-C6 signature scores and worse 28-day survival vs. those in SG2 (Figure 19E-19H). Considering there are 2^6 combinations of the six signatures (up vs. down), we performed PCA on the 6-signature score matrix to comprehensively evaluate the prognostic values of the six signatures (Figure 20). The sum of the first three PCs can explain >80% of the variation. Across these diseases, the PCs corresponding to the degree of separation between C1-C3 vs. C4-C6 among patients (1st PC in trauma and sepsis, and 2nd PC in burns), demonstrated the highest and the most significant association with prognosis.

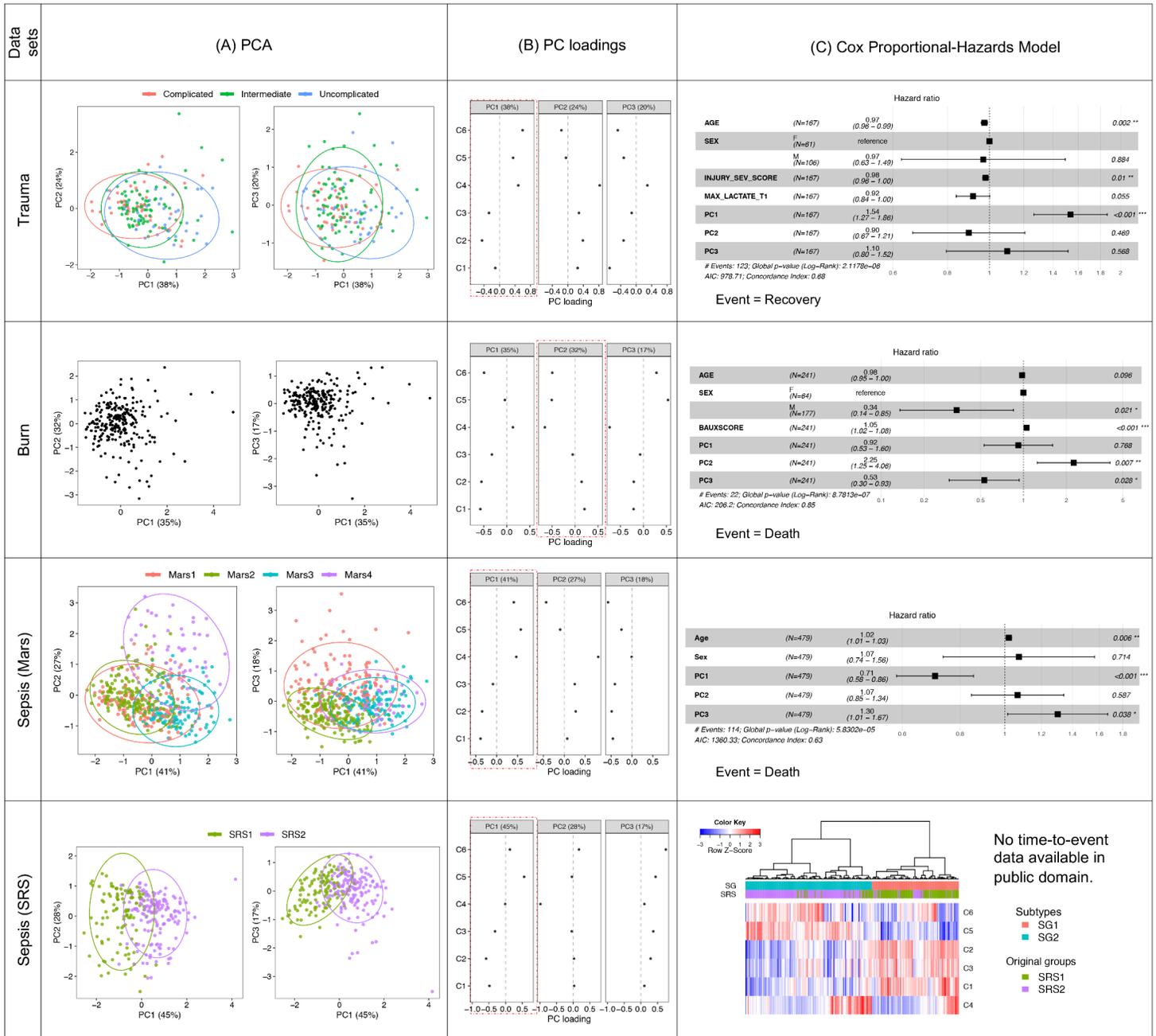


Figure 20 Systemic characterization of the prognostic value for the six signatures.

(A) PCA was performed on the signature score matrix. Patients were color coded by the groups demonstrated in the original paper. Ellipses indicate 95% confidence interval. (B) The signature loadings on first three PCs are shown. The PCs representing the degree of separation between C1-C3 vs. C4-C6 among patients are highlighted in red blocks. (C) The prognostic values of the first three PCs were evaluated by a Cox model after adjustment of age, sex and

severity (if available). The names of datasets are labeled in the first column. No groups documented for burn dataset, so all the patients are in black. Sepsis SRS dataset do not provide time-to-event data. Thus, clustering results are shown instead.

At the single-cell level, trauma induced an increase in C1-C3 and simultaneously decrease in C4-C6 in CD14⁺ monocytes (Figure 17A). However, we have only discussed the relative expression of each signature among patients (see methods for the description of the calculation of signature scores). To fill this gap between single-cell observations and patient subtypes, we next determined the relative changes in the genes comprising signatures C1-C6 in individual patients. To quantify this, an “intrinsic signature score” was calculated for each signature. We then generalized the six intrinsic signature scores from C1-C6 into a single score, the “intrinsic deviation score” (IDS), to roughly reflect the degree that C1-C3 exceeds C4-C6 in each patient (Figure 21A). SG1 patients displayed a much higher IDS than SG2 patients and this difference could not be explained by age, sex, injury severity or early lactate levels (Figure 21B-21F). The IDS was highly and linearly correlated with the PCs representing the separation based on the expression levels of C1-C3 vs. C4-C6 in the population (Figure 22). Thus, the changes in CD14⁺ monocytes we characterized at single-cell level may reflect an underlying biological process that results in patient heterogeneity. Since C1-C3 align with pro-inflammatory gene programs and suppressed C4-C6 relate to impaired immune responses, it is possible that the excessive and sustained over “deviation” in these myeloid gene sets contributes to worse prognosis.

2.3.11 Generation and validation of a classifier for subtype designation

We have demonstrated that the SG subtypes were consistently associated with different outcomes. Thus, assignment of patients to SG1 or SG2 early after admission could assist with clinical decision-making. Considering SG subtypes were defined based patient populations, we sought a strategy to translate the observation in individual patients. Thus, we sought a reliable internal, patient-specific normalization strategy to predict patient SG classification in the future. Considering the distinct distribution of IDS between SG1 and SG2, we built a random forest classifier using the first time point from the 167 published blunt trauma patients [6], taking the six intrinsic signature scores for each patient to predict subtype assignment (Figure 21A). After 5-fold cross validation we obtained a test error of 0.114 ± 0.046 and $AUC = 0.954$, indicating that we found an optimal internal normalization to provide a practical way to predict patient SG classification.

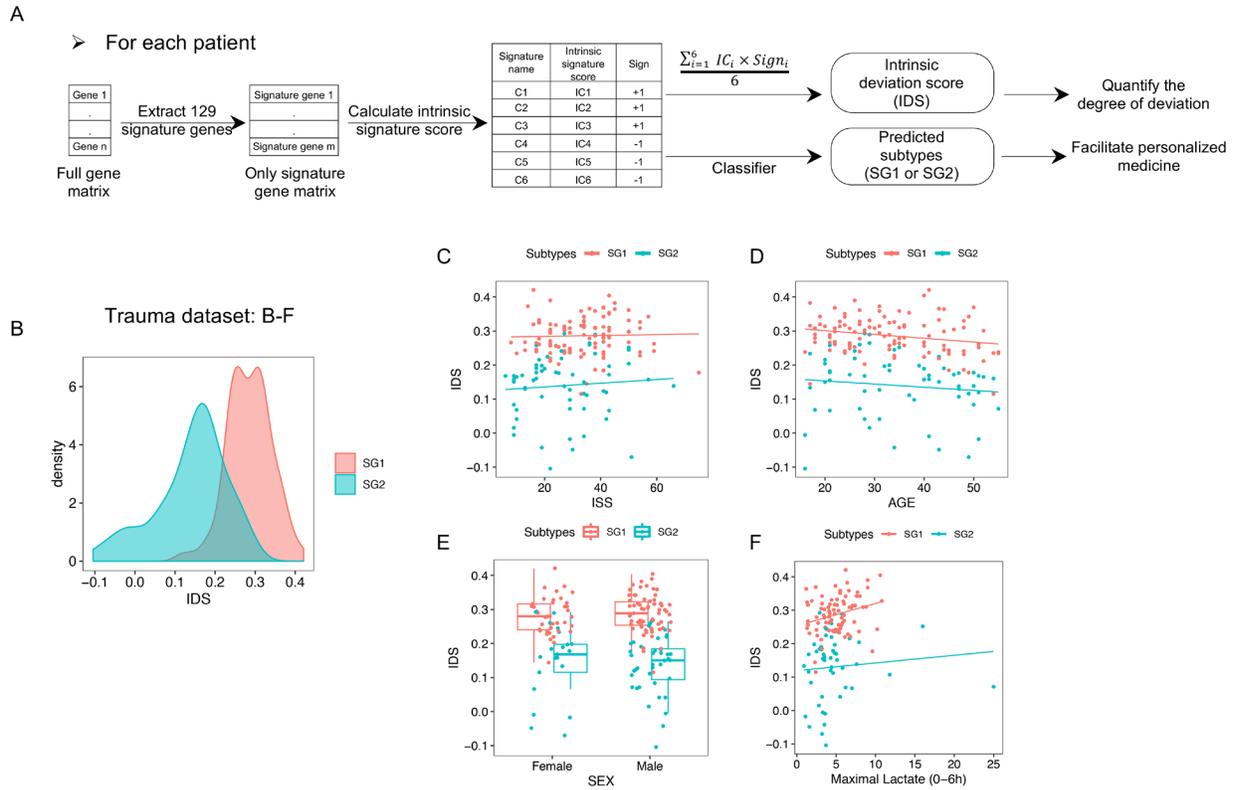


Figure 21 Generation of intrinsic signature score (IDS) to give a potential biological explanation for the patient heterogeneity with critical illness.

(A) Schema describing the workflow for Figure 21 and Figure 23. (B-F) Two subtypes of trauma patients (n=167) have different IDS distribution (B), which cannot be explained by different (C) injury severity (ISS), (D) age, (E) sex and (F) maximal lactate levels within 6 hrs after admission. Data points are color coded by two subtypes.

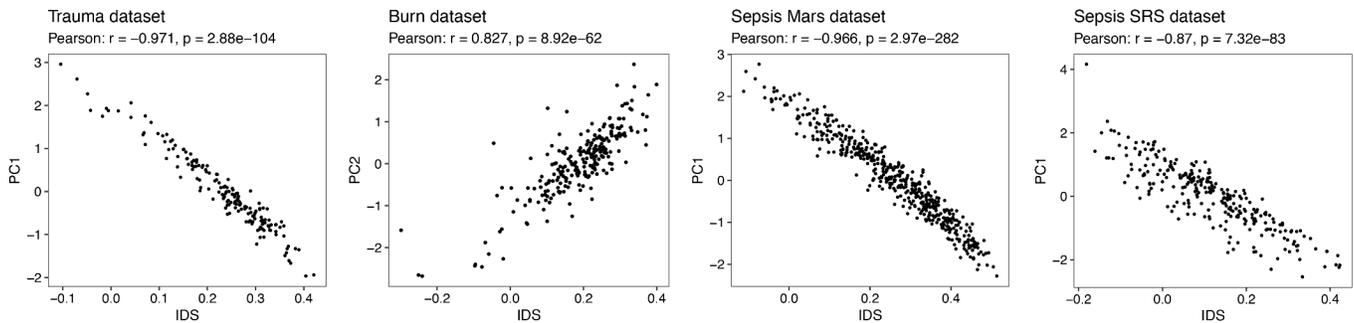


Figure 22 Correlation between IDS and the PC highlighted in red blocks shown in Figure 20.

We applied this classifier to human burn, sepsis [15] and experimental endotoxemia datasets [54]. The predicted SG1 burn and sepsis patients were consistently associated with worse survival (Figure 23A-23D). Experimental endotoxemia in human volunteers has been used to induce a transient systemic inflammatory response (recovery within 24h). Experimental subjects treated with either placebo or endotoxin displayed lower IDS than trauma patients and were all assigned to SG2 (Figure 23E-23F). Experimental endotoxemia induced C1-C3 gene signatures but only minimally suppressed C5 and C6 in circulating leukocytes over 24h. Instead of suppression, endotoxin promoted C4 expression (anti-viral program) (Figure 23G), which was consistent with previous publications [55, 56]. Our analyses highlight the shared and distinct features between a systemic inflammatory response that rapidly resolves (endotoxemia) and one that does not (sustained critical illness) at the single-cell level.

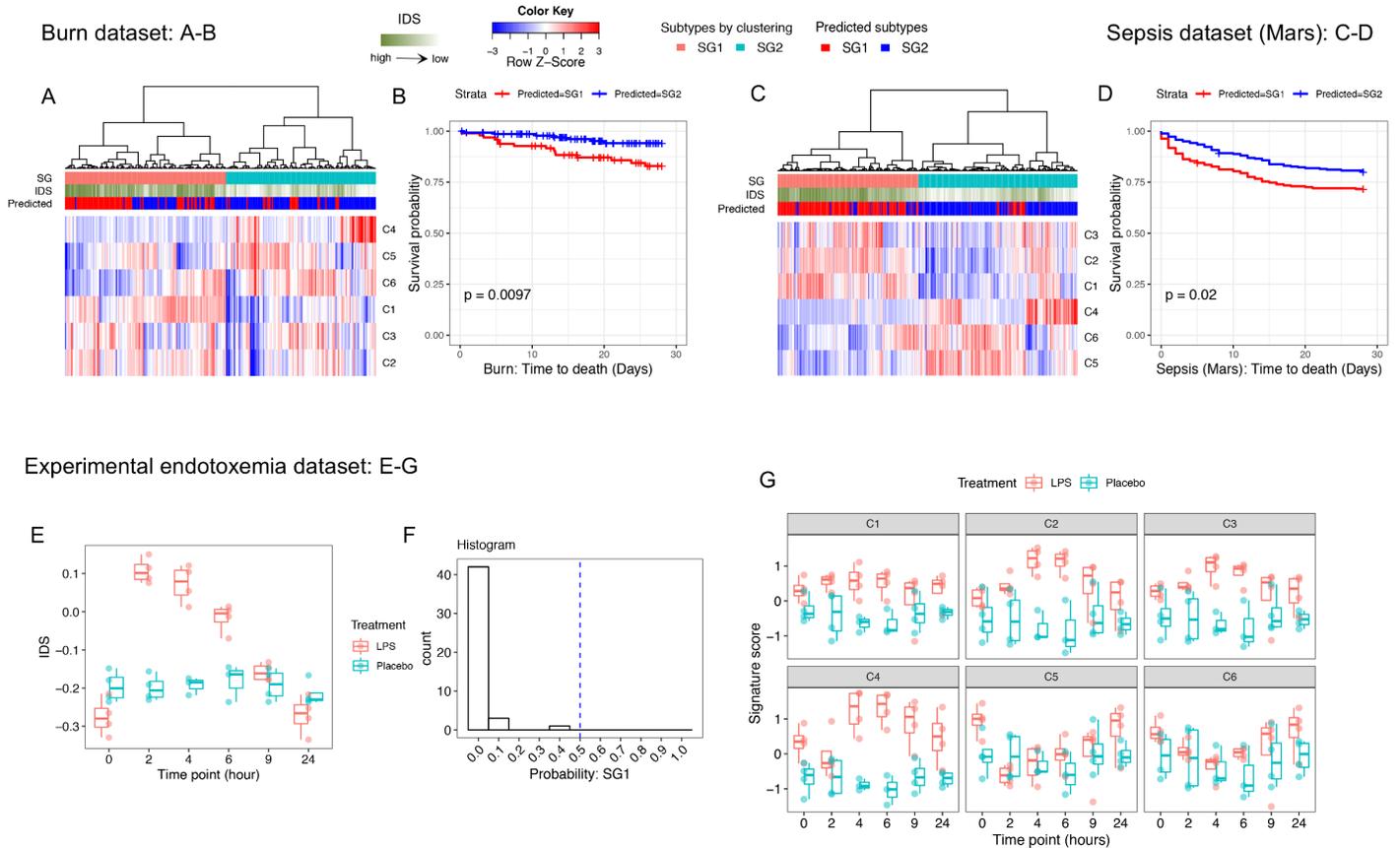


Figure 23 Generation and validation of the classifier for SG subtype designation.

(A-D) For burn/sepsis patients, predicted subtypes and calculated IDS are added to Figure 19E/19G. Kaplan-Meier curve was plotted to visualize 28-day survival between predicted subtypes. Log-rank p value is shown. (A-B) Burn dataset (n=241). (C-D) Sepsis dataset (n=479). (E-G) Endotoxemia dataset (LPS: n=4, placebo: n=4). (E) Visualization of IDS between two groups within 24 hours after LPS or placebo administration. (F) Histogram of the predicted probabilities of SG1 in all the data points shown in (E). (G) Expression of the six signatures in healthy volunteers within 24 hours after administration of LPS or placebo.

2.4 Discussion

The primary goal of this work was to describe the landscape of transcriptomic changes in circulating immune cells induced by severe injury. Complimentary findings in injured mice and humans found dramatic changes in circulating Ly6C⁺/CD14⁺ monocytes that result from the changes in BM. In mice, the trauma-induced changes in monocytes are traceable to progenitor cells in the BM and characterized by three simultaneous features, including (1) induction of features of inflammation, (2) suppressed steady-state features and (3) up-regulation of some neutrophil-associated genes. Neutrophils show similar changes, but smaller in magnitude. Our observations in injured humans showed that the monocyte changes can be generalized into six signatures with distinct and biologically relevant regulatory networks/TFs. These six signatures diverge after injury and further define two patient subtypes associated with different prognosis after severe injury.

The degree of overlap in the inflammatory response between humans and mice has been a topic of debate [13, 17]. Some of the lack of consistency between the species is likely due to genetic differences, however the different composition of cell populations may also be a factor [11]. In this study, we independently analyzed and compared the monocyte responses between injured humans and mice in the early phase of the response. The DEGs revealed an intermediate degree of correlation, suggesting consistencies are generally overlapping in the response of the monocyte compartment between the species early after injury. More importantly, many critical TFs and signaling pathways were shared, indicating that the major biological processes are preserved within this cell population.

Two pathways (G vs. M) of monocyte development derived from progenitors have been documented by Yanez *et al* [57]. The new trajectory derived from BM we characterized in mice

can be generally mapped to G pathway (Figure 24). In humans, we further demonstrated that the gene expression pattern of monocytes deviates from steady state in a continuous manner after injury, rather than a simple binary pattern. We generalized the degree of deviation into a score (IDS). Experimental endotoxemia induces a lower deviation compared with trauma. More importantly, severely injured patients also exhibit distinct magnitudes of deviation, with thresholds represented in two subtypes that cannot be explained by injury severity alone. Thus, by characterizing the full range of transcriptomic patterns observable in monocytes after major systemic perturbations in vivo, our studies go beyond the current model for monocyte development.

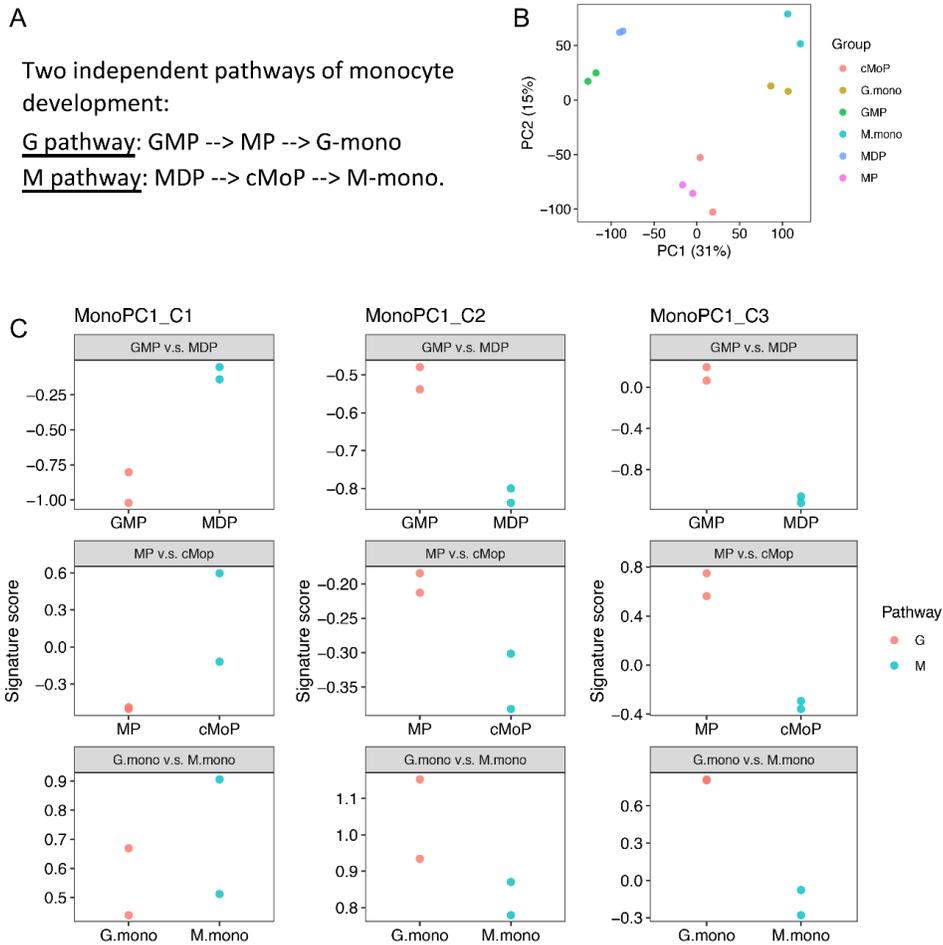


Figure 24 Similarity between the new trajectory of monocytes after T/HS and the G pathway characterized by Yanez *et al.*

(A) Introduction of the major points in the paper from Yanez *et al* [57]. (B) Gene expression matrix was downloaded from GSE88982 GEO dataset. PCA was performed on the six subsets involved in monocyte development. This is largely consistent with Figure 3A in the original paper. (C) Signature scores of MonoPC1_C1-C3 (characterized in Figure 10) were calculated for each sample shown in (B). Monocyte subsets are color coded by the corresponding pathways they demonstrated. Similar to the changes in BM monocytes after T/HS, the G pathway shows increase in MonoPC1_C2-C3 and decrease in MonoPC1_C1 compared with the M pathway. GMP: granulocyte-monocyte progenitors; MP: monocyte-committed progenitors; G-mono: G pathway-derived monocytes; MDP: monocyte-dendritic cell progenitors; cMoP: common monocyte progenitors; M-mono: M pathway-derived monocytes.

We provide evidence that the monocyte gene expression patterns that appear after trauma are also observable in other common acute immunological insults leading to critical illness, including burns and sepsis. A very recent report also identified major changes in CD14⁺ monocytes in patients suffering bacterial urinary tract infections, consistent with the finding that the activation of inflammatory and suppression of MHC II programs in this cell population is a generalizable feature of the early response to trauma and infection [58]. Similar features have also recently been reported in COVID-19 patients [59]. Thus, the transcriptomic features of immune response we identified within myeloid cells may be a pattern common in critical illness due to many etiologies.

Several transcriptional or clinical sub-classifications have been documented for sepsis, including two transcriptional subtype analyses (SRS1-2 [14] and Mars1-4 [15]) and one clinical classification (phenotypes α , β , γ and δ [60]). Transcriptional subtypes SRS1 and Mars1 have the worst prognosis in the original reports; however, >60-70% SRS1 patients were mapped to Mars2 rather than Mars1 (Figure S7D of Mars paper [15]). Our subtype SG1 (high deviation and worse prognosis) largely fits with SRS1, Mars2 and clinical phenotype δ . Mars3, Mars4 and SRS2 map to SG2 (low deviation and better prognosis), with Mars4 similar to endotoxemia-like response (Figure 25A and Figure 23G). The majority of Mars1 has an intermediate deviation (Figure 25B-25C). We notice that the Mars1-specific PC (PC3, Figure 20) is also a prognostic PC achieving statistical significance, suggesting that the biological explanation for the worse outcome in the Mars1 patients is distinct from SRS1 patients (This unsolved question will be answered in Chapter 4). In the future, all of these separate phenotyping efforts may be usefully combined to achieve a more accurate stratification for precision medicine. We advance that goal with our single-cell analyses by linking outcomes in critical illness to specific changes in gene expression in a subset of myeloid cells.

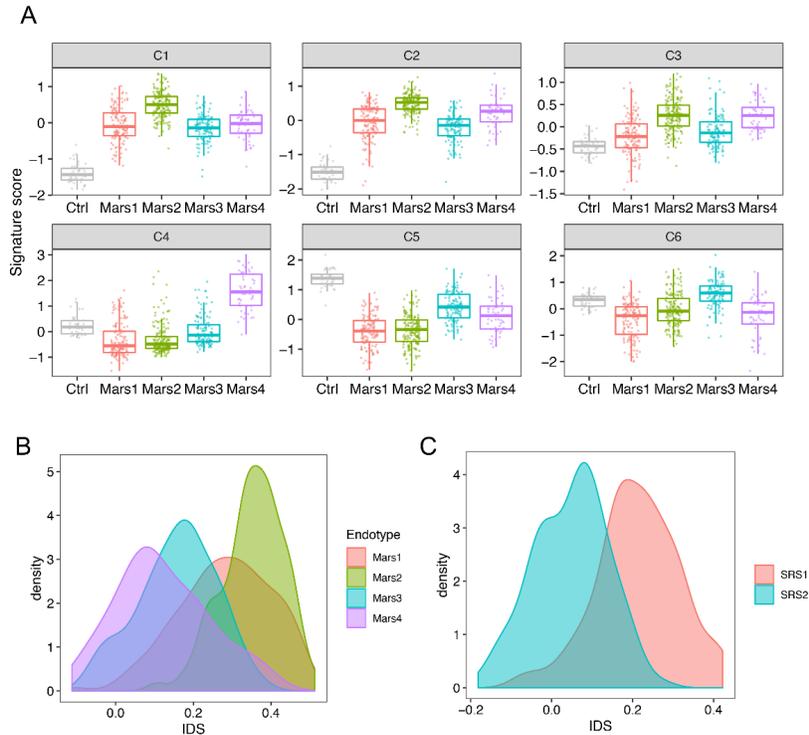


Figure 25 Interpretation of sepsis subtypes using our findings.

(A) Expression of the six signatures in sepsis Mars subtypes compared with healthy control. Based on Figure 23G, Mars4 displays an endotoxemia-like response. (B) IDS distribution among four sepsis Mars subtypes. (C) IDS distribution between two sepsis SRS subtypes.

A limitation of our study is the focus of our single-cell studies on the first three days. However, we confirmed that gene expression patterns we identified early persisted in patients and further defined two subtypes of trauma patients (identifiable as early as ~12 hours after injury) with differential prognosis, which were also recapitulated in burn and sepsis patients. We also limited our evaluation of neutrophils to the BM compartment in mice. Further studies will be required to confirm these gene expression patterns persist in circulating neutrophils.

In summary (Figure 26), our landscape findings provide a new paradigm for the immune response to trauma. In the near term, the two subtypes of trauma patients could be translated

quickly for early identification of the patients at high risk (SG1). In the long term, our findings point to studies on the regulatory mechanisms in myeloid progenitors and CD14+ monocytes as a fruitful area for further research on the mechanisms leading to immune dysfunction after severe injury. Our landscape analysis will act as a new starting point for further study of the regulatory mechanisms and identify the potential target for precision medicine in trauma, which may also be beneficial for other causes of critical illness.

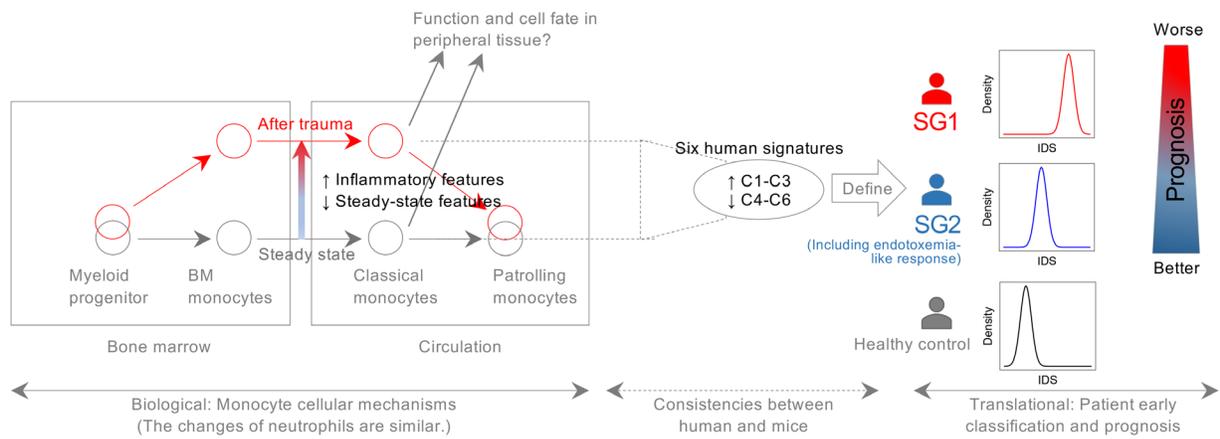


Figure 26 Graphic summary of single-cell RNA-seq analyses.

3.0 Longitudinal Analysis of Transcriptomic Subtypes in Trauma Patients

3.1 Introduction

Trauma is very common among humans and contributes to 10% of deaths and 16% of disabilities worldwide [1]. Furthermore, the diversity in clinical trajectories (e.g. rapid resolution vs. multi-organ dysfunction) points to heterogeneity in the immune response among critically ill trauma patients [5]. We have shown previously that the transcriptomic patterns in human blood CD14⁺ monocytes could be generalized into six gene signatures that identified two patient subtypes (SG1 vs. SG2) using whole-blood leukocyte transcriptome data from 167 patients at 12h after injury [18]. Patients classified as SG1 had significantly worse outcomes (delayed recovery, more severe organ dysfunction and a higher incidence of infection and non-infectious complications) than patients designated as SG2. The main transcriptomic features of SG1 patients includes an upregulation of pro-inflammatory genes and a simultaneous suppression of genes representing MHC II (Major histocompatibility complex II) expression and interferon signaling compared with SG2 patients. To better understand the prognostic value of SG subtype designation over time, we sought to characterize the longitudinal changes in SG subtypes in association with a range of clinical outcomes.

3.2 Methods

Longitudinal, whole-blood leukocyte RNA microarray data were downloaded from Gene Expression Omnibus (GSE36809) [6]. Samples annotated as either “low RNA quality” or “incomplete time points” in metadata were excluded from the following analyses. SG subtypes were computed for each data point from trauma patients (each patient at a specific time point) as we reported previously [18]. The sampling time points were binned into the nearest time point, for a total of 7 time bins: 12h, 1d, 4d, 7d, 14 d, 21 d or 28d. If a patient had >1 time point falling in the same time bin, only the first time point in the bin was analyzed. Kaplan-Meier analysis was performed using the *survival* R package (v3.1.8) to evaluate 28-day recovery between (1) SG subtypes at a specific time point or (2) longitudinal SG subgroups. The time-to-event data at a specific time point was defined as the Time-to-Recovery (TTR) from the original data minus the analyzed sampling time point. For longitudinal SG subgroups, the later sampling time point (in the time bin of 1d) was used as time zero. For non-survivors, “hospital length of stay” minus the analyzed sampling time point was used as the time-to-event data, and recovery status was annotated as “no”. Patients that recovered prior to the analyzed sampling time point were excluded. To compare other parameters between groups, continuous variables were depicted as median \pm median absolute deviation (MAD) and tested by Wilcoxon rank-sum test. Categorical variables were depicted as counts (percentages) and tested by Fisher’s Exact test. A two-sided p value was computed. $p < 0.05$ was considered as significant.

3.3 Results

The gene array transcriptomic dataset generated from whole-blood leukocytes, obtained from severely injured humans by The Inflammation and the Host Response to Injury Large-Scale Collaborative [6], was used to assess clinical outcomes as a function of SG subtype designation over time. Patients were assigned to either SG1 or SG2 subtypes based on leukocyte transcriptomic patterns at 12h, 1d, 4d, 7d, 14 d, 21 d and 28d after injury. SG subtype designation displayed dynamic changes after trauma within 12h-7d post injury, with a general trend to transition from SG1 to SG2 over time (Figure 27A). The 12h time point had the highest percentage of SG1 patients. Approximately 50% of patients identified as SG1 at 12h converted to SG2 by 1d, a trend that continued at later time points. However, there was also a small number of patients that switched from SG2 to SG1 between 12h and 1d and over time (Figure 27A). By 4-7d post injury, subgroup designation remained relatively stable in surviving patients.

To investigate the prognostic value of SG status at later time points, we performed Kaplan-Meier (K-M) analyses to establish the differences in TTR between SG subtypes at each time point separately. Due to limited numbers of SG1 patients after 7d post injury, we only analyzed the data points falling in the time bins ≤ 7 days. SG1 designation consistently associated with a subsequent slower recovery (Figure 27B-27D).

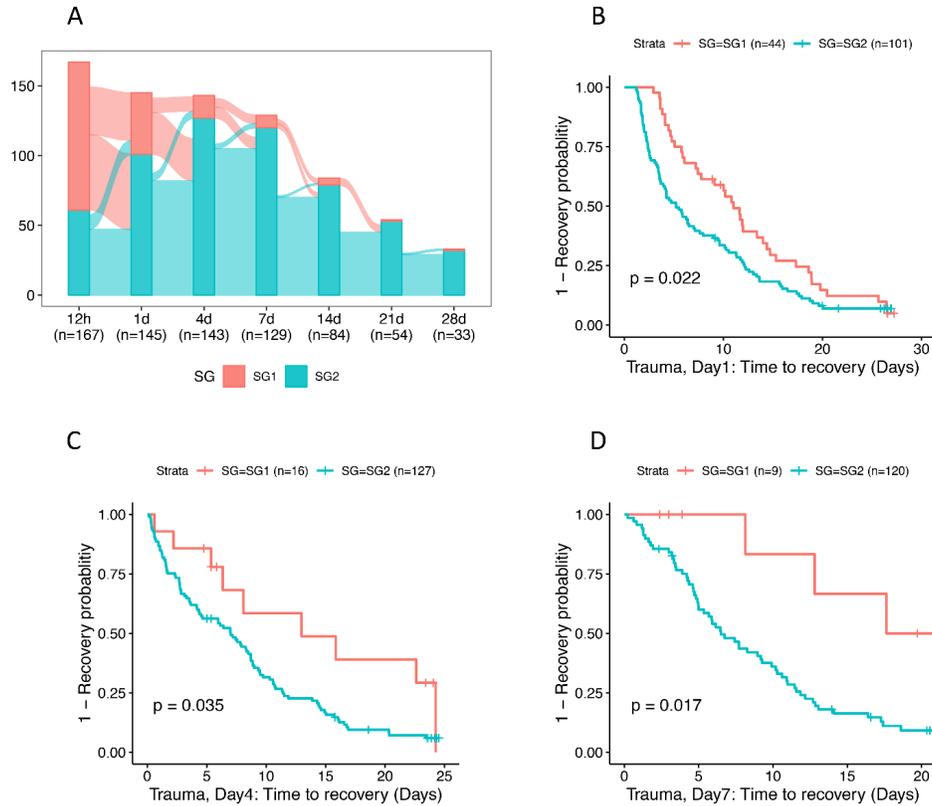


Figure 27 Prognostic value of SG subtypes at 1d, 4d and 7d post injury.

(A) Overview of SG subtypes over time after injury. Sampling time points were binned in the nearest time bin. For the patients with >1 time point in a time bin, the 1st time point was kept. (B-D), K-M analysis at each later time point. Log-ranked p value was provided. (B) Patients were grouped by the SG status at 1d. (C) Patients were grouped by the SG status at 4d. (D) Patients were grouped by the SG status at 7d. Time zero for K-M analysis was set as the analyzed time point.

Our analysis indicates that categorizing patients as SG1 as early as 12h or less could be useful to identify patients at risk for a delayed recovery [18]. The finding that some patients converted from SG2 to SG1 at 1d suggested that assessing SG status at both 12h and 1 day could improve prognostic accuracy. Therefore, we next performed K-M analysis in subgroups of patients based on the SG status at both time points (Figure 28). For the patients who were SG1 at one of the two time points (at either 12h or 1d), the SG status at another time point only showed a trend

without achieving statistical significance (Figure 28A-28B). However, for the patients who were SG2 at either time point, the SG status at the other time point markedly associated with recovery rates. As shown in Figure 28C, for patients that were classified as SG2 at 1d and SG1 at 12h the recovery was dramatically slower than those classified as SG2 at both time points ($p = 0.0024$). A similar pattern was seen in the patients who were SG2 at 12h but then became SG1 at 1 d ($p = 0.073$, Figure 28D).

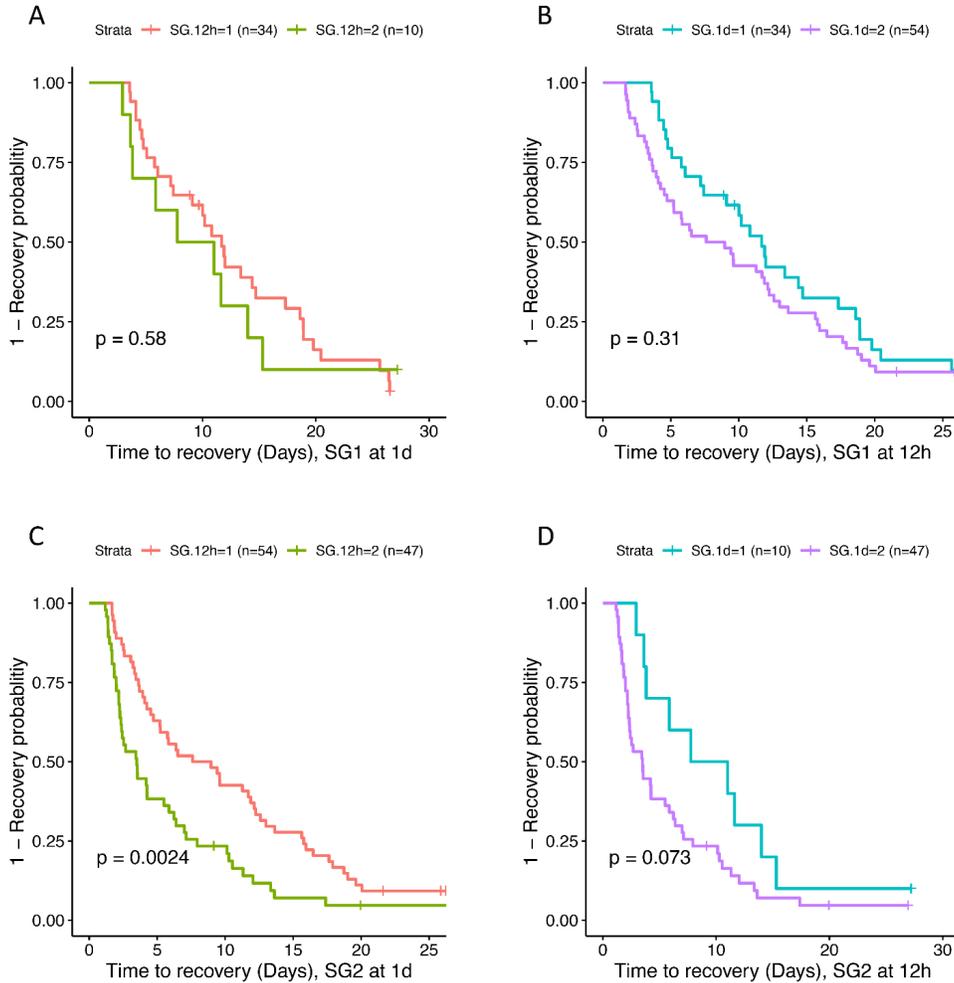


Figure 28 K-M analysis in subgroup analysis.

Patients were divided into subgroups into their SG status at either 12h or 1d. Log-ranked p value was provided. (A) Of all SG1 patients at 1d, the prognostic value of SG status at 12h was evaluated. Time zero for K-M analysis was set as the time point falling in the bin of 1d. (B) Of all SG1 patients at 12h, the prognostic value of SG status at 1d was evaluated. (C) Of all SG2 patients at 1d, the prognostic value of SG status at 12h was evaluated. (D) Of all SG2 patients at 12h, the prognostic value of SG status at 1d was evaluated.

In addition to TTR, a measure of global recovery from organ dysfunction, we also examined other outcome parameters (Appendix Spreadsheet 1 and 2) and potential associated factors (Appendix Spreadsheet 3 and 4) in our subgroup analysis. Generally, SG1 was also associated with worse outcome parameters, such as higher Marshall (MOF) or Denver 2 score, in

each of the four subgroup analyses. Among the patients that were SG2 at either time point (12h or 1d), the patients that were SG1 at the other time point exhibited significantly longer ICU and hospital stays as well as more days on the ventilator. These findings show that an SG1 designation at either time point has additional prognostic value beyond each other and associates with a subsequent delay in recovery.

Interestingly, in the subgroup analysis among SG2 patients at 12h, patients that converted to SG1 at 1d showed early evidence of cardiovascular dysfunction (e.g. higher heart rate, higher respiratory rate, lower PaO₂, and lower systolic blood pressure after arrival, Appendix Spreadsheet 3). This association was not found in other subgroups. These patients also showed higher cardiac MOF scores and delayed cardiac recovery (Appendix Spreadsheet 1). These results suggest that early cardiovascular dysfunction may be a factor for the conversion from SG2 to SG1 between 12 and 24h post-injury. Perhaps more importantly, these findings suggest that a favorable immune profile on admission can deteriorate within the first 24h and that this deterioration may be driven by hemodynamic instability.

In our previous study, we showed that SG1 designation at 12h only correlated significantly with limited number of known prognostic factors, including higher Injury Severity Score and higher lactate levels in the initial 6h, compared with SG2 patients. Furthermore, after adjusting for potential co-variants including the two above using Cox regression model, we demonstrated that SG status at 12h has independent prognostic value beyond elevated circulating lactate or injury severity [18]. In contrast to our findings at 12h, we report here that SG1 status at 1d was correlated significantly with more known prognostic factors, including higher APACHE II score, higher crystalloids, and blood transfusion during the first 12h, and higher lactate and lower base deficit at later time points within 24h post injury (Appendix Spreadsheet 5). Taken together, these results

show that SG subtype designation associates with unique outcome parameters based on the time point, and that assessing subtype at both 12h and 1d provides even more information on patient outcomes than assessment at a single time point.

To visualize the prognostic value of SG status determination at both 12h and 1d, we divided all the patients with data available at both 12h and 1d (n = 145) into four groups, labeled as SG status at 12h --> SG status at 1d, including two groups of SG-non-converters (SG1-->SG1 or SG2-->SG2) and two groups of converters (SG1-->SG2 or SG2-->SG1). For categorical outcomes (nosocomial infections and non-infectious complications), SG1 non-converters had the highest incidence and SG2 non-converters had the lowest incidence. The two groups of converters showed an intermediate incidence of these outcomes (Figure 29A-29B). Notably, all SG1 non-converters exhibited a TTR of 5 days or greater (Figure 29C). Similar patterns were observed in global recovery. As shown in Figure 29D, SG1 non-converters underwent the slowest recovery, followed by the two groups of converters, and then SG2 non-converters with the fastest recovery.

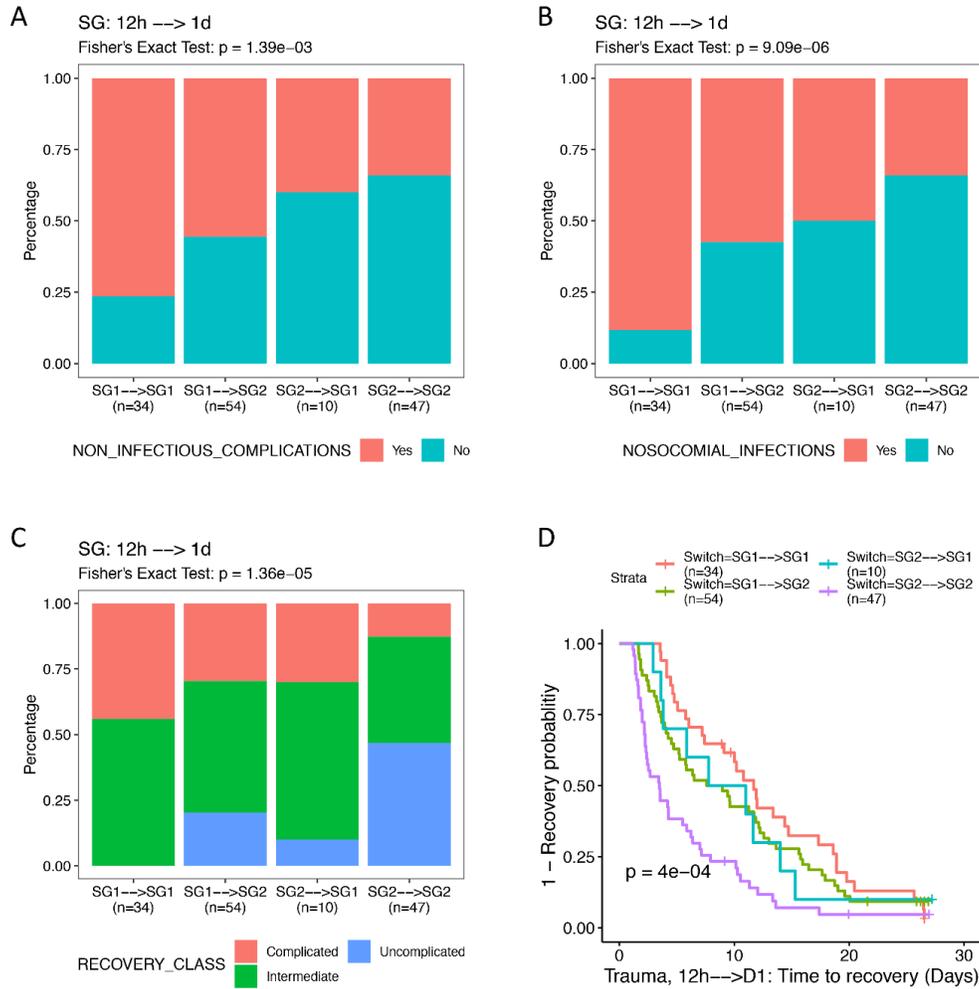


Figure 29 Differential outcomes across four the SG1-SG2 combinations for 12h and 1d.

(A) Incidence of non-infectious complications among the four groups. (B) Incidence of nosocomial infections among the four groups. (C) Incidence of complicated (TTR > 14 days or dead), intermediate (5 days ≤ TTR ≤ 14 days) and uncomplicated (TTR < 5 days) among four groups. (D) K-M analysis. Time zero was set as the sampling time point falling in the bin of 1d.

3.4 Discussion

This study was undertaken to extend our previous findings using a patient classifier developed from transcriptomic patterns identified in circulating CD14+ monocytes from injured humans [18]. Using transcriptomic data from bulk leukocytes, we were able to use the classifier to designate patients as SG1 (worse prognosis) or SG2 (better prognosis) at multiple time points after injury. We confirmed that patient designation as SG1 as early as 12h after injury correlated strongly with slow recovery. We went on to show that a subset of patients with early hemodynamic instability convert from SG2 to SG1 by 1d, showing that unstable patients can evolve to a pathogenic gene signature after admission. Finally and most importantly, monitoring SG subtype at two time points over the first day can provide additional prognostic value beyond each other and leads to a very high capture of patients likely to experience slow recovery and organ dysfunction.

Interestingly, the patients that converted from SG2 to SG1 were associated with worse cardiovascular parameters (higher HR and lower BP) early post injury. Thus, the switch from SG2 to SG1 within day 1 may result from cardiovascular dysfunction, traumatic shock, or ongoing bleeding. The idea that the transcriptomic profile can shift away from a favorable state to an unfavorable state early after arrival points to the importance of monitoring transcriptomic subtypes over time should this strategy be adopted for clinical decision support. In addition, our findings confirm that the prevention or reversal of hemodynamic instability after admission is of paramount importance to minimize immune dysfunction [61, 62].

We have demonstrated previously that SG subtypes also associate with a dysfunctional immune response across several etiologies of critical illness, including sepsis and burns [18]. Thus,

longitudinally monitoring transcriptomic subtypes as described in other etiologies of critical illnesses [14, 15] may also be helpful and warrants further exploration.

4.0 Single-cell ATAC-seq Reveals Global Epigenetic Alterations in Immune Cells during Severe Trauma

4.1 Introduction

Trauma induces a system-wide physiologic crisis including strong activation of the innate immune system. When excessive, this gives way to immune dysfunction leading to impaired resistance to infection [1]. However, there is considerable patient-to-patient heterogeneity in these responses and the associated clinical trajectories [5]. This suggest that patient intrinsic factors contribute to the variability of the immune response following injury. In chapter 2, we extracted gene signatures from scRNA-seq data derived from circulating CD14+ monocytes from severely injured patients [18]. Using these gene signatures within a whole-blood leukocyte gene array trauma dataset [6], we characterized the two transcriptomic subtypes (Signature Groups: SG1 vs. SG2). Compared with SG2, SG1 trauma patients exhibited higher up-regulation of inflammation and greater suppression of MHC II and IFN signaling, and had worse clinical outcomes (prolonged organ dysfunction and higher infection rates). Similar outcome-associated SG patterns were observed in burn and sepsis patients [3].

scATAC-seq (single-cell assay for transposase-accessible chromatin using sequencing) [63] can provide representative information on open chromatin accessibility within cell subsets derived heterogenous populations of cells. Currently, scATAC profiles have not been widely studied in specific diseases, including the immune dysfunction associated with acute critical illness. We postulated that scATAC-seq would allow us to explore open chromatin profiles and the regulatory mechanisms across cell types in PBMCs isolated from trauma patients and explore

regulatory mechanisms for the transcriptomic subtypes we defined previously, and may also provide additional information beyond that cannot be obtained from scRNA-seq alone.

The ChromHMM 15-state model [64] provides detailed information on various human epigenomes for primary cells and tissues including the major circulating leukocyte cell types. This allowed us to map the differentially accessible (DA) open chromatin regions identified from scATAC-seq with reference chromatin states. The DA peaks associated with the genes used to define SG subtypes in CD14⁺ monocytes (referred as SG signatures in the following context) were largely mapped to active chromatin states of CD14⁺ monocyte reference epigenome. Unexpectedly, scATAC-seq also revealed that trauma induced global epigenetic alterations across major immune cell types that pointed to increased accessibility in chromatin structure and de-repression of polycomb targets, which should normally be suppressed in hematopoietic lineages. We developed a strategy to extract the gene signatures associated with these novel global epigenetic alterations identified by scATAC-seq. Using whole-blood leukocyte transcriptomes, we were able to validate the pathological epigenomic gene signatures identified in cells from representative trauma patients in independent and large datasets of patients with critical illness spanning different etiologies including trauma, burns and sepsis. Importantly, these features were associated with worse prognosis across all three causes of acute critical illness. Furthermore, the gene signatures associated with the global epigenetic alterations were distinct from the well-established transcriptomic patterns associated with acute systemic inflammation, characterized previously by us and others in the context of trauma.

4.2 Methods

4.2.1 Patient and human volunteer enrollment

Trauma patients and healthy volunteers were enrolled in an observational study approved by the University of Pittsburgh Institutional Review Board. Informed consent was obtained from all the subjects or next of kin. Patients suffering blunt or penetrating trauma that were admitted to the intensive care unit and suffering hypotension (systolic blood pressure <90 mmHg) or tachycardia (heart rate > 108) on admission were eligible for enrollment. Four patients representing two different clinical trajectories (quicker recovery: MM3001 and MM3016, versus slower recovery: MM3008 and MM3038) were selected for scATAC-seq. Blood samples were obtained within 4hrs of injury and at 24hrs and 72hrs after injury. Each sample was paired with an age and sex matched healthy control, for a total of 16 samples.

4.2.2 PBMC isolation, cryopreservation, and thawing

Human PBMC were isolated by standard Ficoll centrifugation (Ficoll-Paque PLUS, Cat#17-1440-03, GE Healthcare) as described previously[18]. SepMate tubes (Cat#85415, StemCell) were used to regulate the procedure. The isolated cells were cryopreserved and thawed for analysis according to the 10x Genomics protocol. Briefly, PBMC were resuspended in 0.5ml resuspension medium (40% FBS in DMEM) and 0.5ml freezing medium (40% FBS + 30% DMSO in DMEM) in a 1:1 ratio. Cells were chilled in CoolCell (Cat#CLS432002, Corning) in -80 overnight and transferred to liquid nitrogen. Cryopreserved vials were thawed in the water bath at 37°C and then transferred to a 50mL conical tube after thawing was complete. A milliliter of

thawing medium (PBMC: 10% FBS in DMEM; BMMC: 20% FBS in PBS) was added dropwise (5 sec/drop), followed by 2mL, 4mL, 8mL, 16mL thawing medium at ~ 1-min intervals. After this, the cells were washed and resuspended in calcium and magnesium free buffer (PBMC: PBS with 0.04% BSA) for cell counting and nucleus isolation.

4.2.3 Nucleus isolation

Isolated single-cell suspensions were processed for nuclei isolation immediately prior to single-cell ATAC library preparation, strictly following the 10x Genomics protocol (CG000169_RevD). Briefly, single-cell suspensions were centrifuged 300 rcf for 5 min. After removing the supernatant, cell pellets were resuspended with DNase I (Cat#07900, StemCell) and incubated at room temperature for 15 min. After centrifugation at 300 rcf for 5 min, cells were resuspended in lysis buffer and placed on ice for 3 min. Then, wash buffer was added, and samples were centrifuged at 500 rcf for 5min and then resuspend in diluted nuclei buffer.

4.2.4 Single-cell ATAC library preparation and sequencing

Chromium Next GEM Single-Cell ATAC Library & Gel Bead Kit v1.1 (Cat#1000175, 10x Genomics), Chromium Next GEM Chip H Single-Cell Kit (Cat#1000162, 10x Genomics) and Single Index Kit N Set A (Cat#1000212, 10x Genomics) were purchased for scATAC library preparation. The experimental steps strictly followed 10x Genomics Next GEM single-cell ATAC kit v1.1 protocol (CG000209_RevD). We targeted 5000 nuclei for each sample. Single-cell nuclei suspensions were first digested with the 10x ATAC enzyme, Transposase, and then prepared for Gel Bead-In Emulsions (GEMs) and barcoded. GEMs were purified by Dynabeads and SPRIselect

reagent cleanup steps. Then, samples were indexed, PCR amplified and processed for double sided size selection using the SPRIselect reagents to create the final libraries. Post-library QC was performed with an Agilent TapeStation 4150 with 2 μ L from each library. A periodical pattern representing preserved nucleosomal structure indicated good quality for the prepared library. Libraries were pair-end and dual-indexing sequenced on an Illumina NovaSeq 6000 in UPMC Genome Center.

To clarify, the 4 time points (including the paired healthy control) for the same patient were processed in parallel. The 16 samples were processed in 4 parallel batches. That single one cluster of DC and single one cluster of CD16+ monocytes were identified across different time points in contrast to multiple clusters of CD14+ monocytes after trauma further confirmed that the batch effects were minimal compared with the biological changes (Figure 32A-32D).

4.2.5 Single-cell ATAC data processing

The BCL files generated by Illumina sequencer were processed by Cell Ranger Atac pipeline (v1.2.0) for demultiplexing using *cellranger-atac mkfastq* and generated count matrix for peaks for each sample using *cellranger-atac count* (mapped to GRCh38 human reference genome). *cellranger-atac aggr* were used to aggregate the outputs from multiple samples. The aggregated peak-barcode count matrix were processed using Signac R package (v1.0.0.9000) [65] and Seurat R package (v3.2.0) [24]. The peaks accessible in ≥ 10 cells and the cells with ≥ 200 peaks detected were retained. The cells with total number of fragments in peaks ≤ 3000 or ≥ 30000 were excluded to remove the cells with low sequencing depth or potential multiplets. The cells with fraction of fragments in peaks $\leq 15\%$, poor nucleosome banding pattern (nucleosome signal ≥ 4) or low TSS enrichment score (TSS ≤ 2) were excluded to remove potential low-quality cells. The cells with

ratio of reads in genomic blacklist regions ≥ 0.05 were excluded to remove artifactual signals. Then, the count matrix was normalized by TF-IDF normalization followed by singular value decomposition (SVD), known as latent semantic indexing (LSI) [66]. The 2nd ~ 30th dimensions were used for graph-based clustering and non-linear dimension reduction. The resolution for graph-based clustering was tuned back and forth until the identified clusters were biologically meaningful. Transcription factor binding profiles were obtained from JASPAR2020 R package.

4.2.6 Identification of DA peaks

We borrowed the H value (Shannon Entropy) and Q value [67] that were previously used to quantify tissue specificity of RNA expression and to assign the genes to a specific tissue for the identification of cluster-specific peaks among multiple clusters. To identify the DA peaks between two clusters, we adopted the method provided by Signac R package (v1.0.0.9000), using logistic regression and adding the total number of fragments and experimental batches as latent variables. The peaks of interest were annotated as the nearest genes using ChIPseeker R package (v1.20.0).

4.2.7 Chromatin state enrichment analysis

For a set of peaks (single peak or a group of peaks) and a specific chromatin state provided by 15-state ChromHMM model (E124), fold enrichment (FE) and hypergeometric p value were computed as mentioned in the Epigenome Roadmap paper [64], quantified by the overlap of bases in genome. After that, for each peak set, FE for a given state i was further scaled within 0 to 1 as below:

$$Relative\ enrichment_i = \frac{FE_i - \min(FE)}{\max(FE) - \min(FE)}$$

where $FE = (FE_1, \dots, FE_{15})$.

4.2.8 Generation of state-barcode matrix and identification of state clusters

To evaluate the global epigenetic changes across 15 chromatin states, we began from the fragment file (`fragments.tsv.gz`) generated by *cellranger-atac aggr* and created the cut site count matrix for the 15 chromatin states of the reference epigenome E124. To clarify, different from the peak-barcode matrix, we counted the number of all the cut sites falling into each chromatin state in the genome for each single CD14+ monocyte, regardless whether they were in peaks or not. Then, the new count matrix was processed by Signac [65], largely similar to the methods used in preparing the peak-barcode matrix. There were two major differences. First, since all the cells for this analysis already passed the filtering step when we analyzed the peak-barcode count matrix. Thus, no extra filtering step was needed for this analysis. Second, the state-barcode matrix was no longer a sparse matrix and much more resistant to sequencing depth, and the 1st dimension of LSI was no longer highly correlated with sequencing depth. Thus, the 1st dimension of LSI was also included for the downstream clustering and nonlinear dimension reduction.

4.2.9 Enrichment analysis between gene sets

For a list of genes with rank, gene set enrichment analysis (GSEA) [31] was performed using the `fgsea` R package (v1.6.0). For a list of genes without rank, fold enrichment and hypergeometric p value were computed. Raw p values were corrected by Benjamini-Hochberg

method for multiple testing. $FDR < 0.05$ was considered as significant. MSigDB gene sets v5.2 were used.

4.2.10 Hierarchical clustering

1 minus Pearson's correlation was used as the distance matrix. `hclust()` function was used for hierarchical clustering with the method as "ward.D2" [68].

4.2.11 Calculation of signature scores

To calculate the signature score for a set of genes, the expression values for each gene were first z-score transformed across all the analyzed samples. After that, the expression values of the signature genes were averaged within each sample, and the calculated mean was the signature score for each corresponding sample.

4.2.12 Time-to-event analysis

For the longitudinal gene array datasets (trauma and burn), only the 1st sampled gene array data points were included for the time-to-event analysis. Due to the very few cases of in-hospital deaths (mortality rate: 4.2%), event was set as recovery status in trauma patients. The definition of recovery was based on the annotation from the original dataset [6]. For non-survivors, "hospital length of stay" was used as the time and recovery status was annotated as "No". For burn and sepsis patients, the event was set as in-hospital death. A Kaplan-Meier curve was plotted by survival R package (v3.1.8) for visualization 28-day prognosis and the Log-rank p value was

computed. Cox proportional hazards model was performed by `coxph()` function in R for adjusting covariates.

4.3 Results

4.3.1 Overview of trauma-induced open chromatin changes in PBMCs

To characterize the accessible chromatin regions in circulating cells after severe injury, we harvested PBMCs from four trauma patients across three time points (<4hr, 24hr, 72hr) post-injury (Figure 30A). These cells were analyzed in parallel with age and sex matched healthy controls yielding a total of 16 samples that were subjected to scATAC-seq. The four seriously injured patients (Injury Severity Score [ISS] ≥ 18) were selected to represent two distinct clinical trajectories; (i) rapid recovery and a short duration of organ dysfunction (patients MM3001 and MM3016) and (ii) slow recovery and persistent organ dysfunction (patients MM3008 and MM3038), based on multiple organ dysfunction (MOD) scores [69] over time (Figure 30B). UMAP plots were constructed based on ATAC-seq peaks with $\sim 57,000$ cells passing the quality control (see Methods for the filtering criteria), and yielded clusters of the major mononuclear immune cell types (Figure 31). Generally, cells from MM3001 and MM3016 (rapid recovery) were more similar to each other, while MM3008 and MM3038 (slow recovery) tended to show more distinct patient-specific patterns (Figure 31D).

A

	Ctrl	<4hr	24hr	72hr	Clinical description
MM3001	✓	✓	✓	✓	Quicker recovery
MM3008	✓	✓	✓	✓	Slowest recovery
MM3016	✓	✓	✓	✓	Quickest recovery
MM3038	✓	✓	✓	✓	Slower recovery

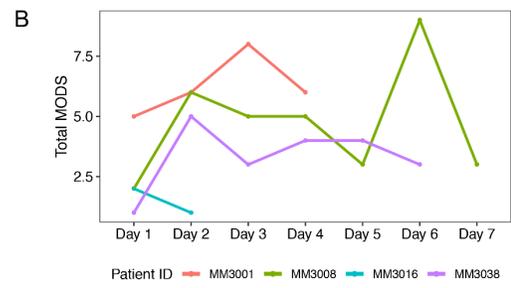


Figure 30 Overview of patient clinical information for scATAC-seq

(A) Experimental design and patient clinical information. (B) Changes in Multiple Organ Dysfunction Scores (MODS) in the four patients.

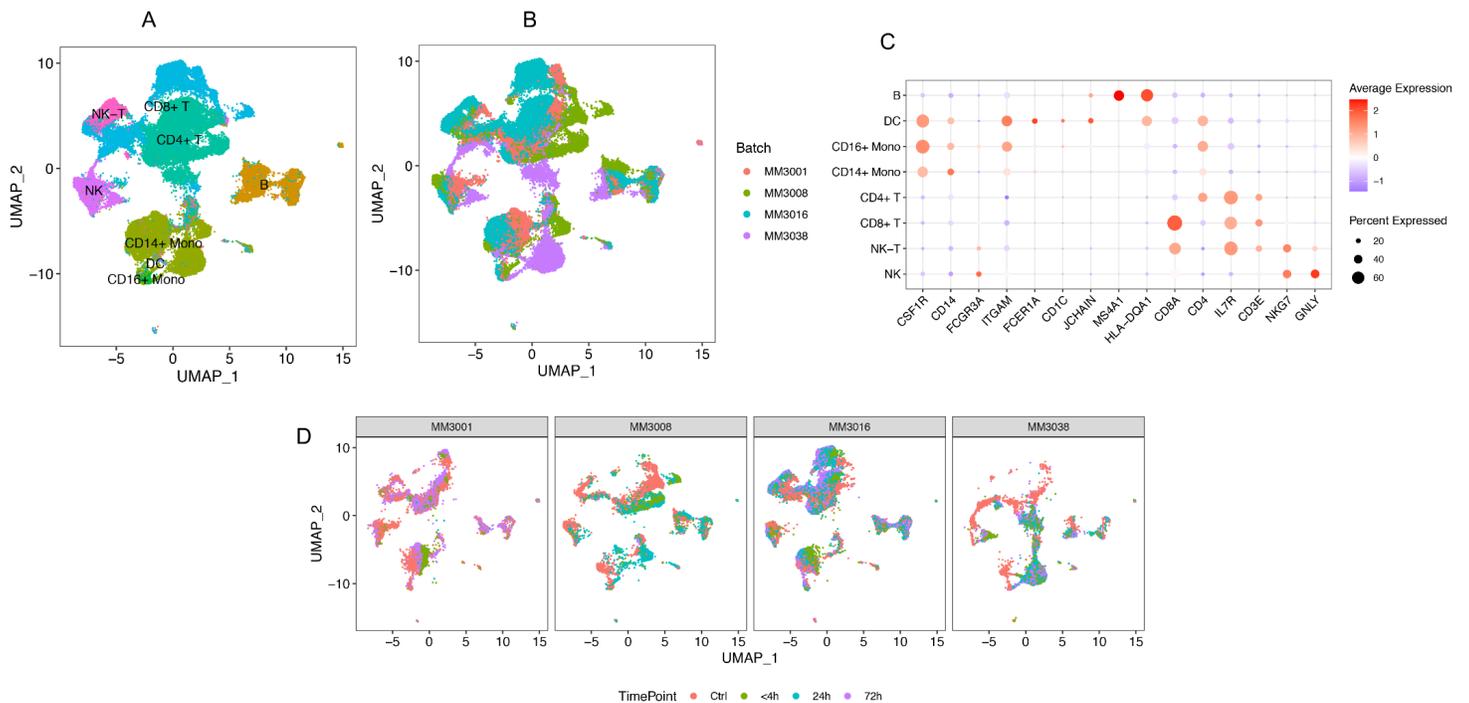


Figure 31 Overview of open chromatin profile in PBMCs isolated from trauma patients in a time-series manner and paired with healthy controls.

(A) UMAP plot of PBMCs color coded by major cell types. (B) UMAP plot of PBMCs color coded by different experimental batches. All the samples for the same patients were processed in the same batch. (C) Gene activities (computed by Signac R package) of major lineage markers in different cell types. (D) UMAP plot of PBMCs color coded by different time points and wrapped by individual patient. UMAP was performed based on the globally accessibility of peaks.

4.3.2 Overview of trauma-induced open chromatin changes in circulating monocytes

Given our previous scRNA-seq analyses of myeloid cells in severely injured patients [18], we first analyzed for corresponding chromatin alterations within this immune compartment. The most substantial changes took place in CD14+ monocytes (Figure 32A-32D), consistent with our

previous scRNA-seq analyses of trauma [18]. The single clusters of dendritic cells (DC, C10) and CD16⁺ monocytes (C6) provided an internal control for batch effect and human heterogeneity, indicating that the differences between CD14⁺ monocyte clusters can be interpreted as meaningful epigenomic changes. These clusters are referred to as “Peak_CC#” (Cell Cluster) in Figure 35D, since they are the cell clusters identified based on the ATAC-seq peak profiles.

The ten CD14⁺ clusters were generally clustered into three blocks (Figure 32E). Block I clusters were enriched in either control samples (C4 and C11) or the later time points from cases with rapid recovery (C1 and C8). Block II clusters enriched in either the early time points from the cases with quicker recovery (C2 and C3) or later timepoints from the cases with slower recovery (C0 and C5). Cluster 9 in block III was highly associated with the 24 and 72h timepoints from MM3008, the case with the slowest recovery (Figure 32F). Thus, the clusters in block I shared features associated with a more favorable clinical state similar to healthy controls, while the clusters in block II and block III were distinct from controls and therefore likely to include features associated with an unfavorable clinical state.

We mapped the DA peaks between CD14⁺ clusters from block I, II, or III versus the corresponding control cluster (C4 or C11) to the ChromHMM 15-state model for CD14⁺ monocytes (E124) [64] (Figure 32G). Within block I, the DA peaks were mainly associated with active states. In contrast, there was an increase in the relative enrichment of normally repressed states, such as bivalent regions seen in block II and this extended to include polycomb or quiescent chromatin regions in block III.

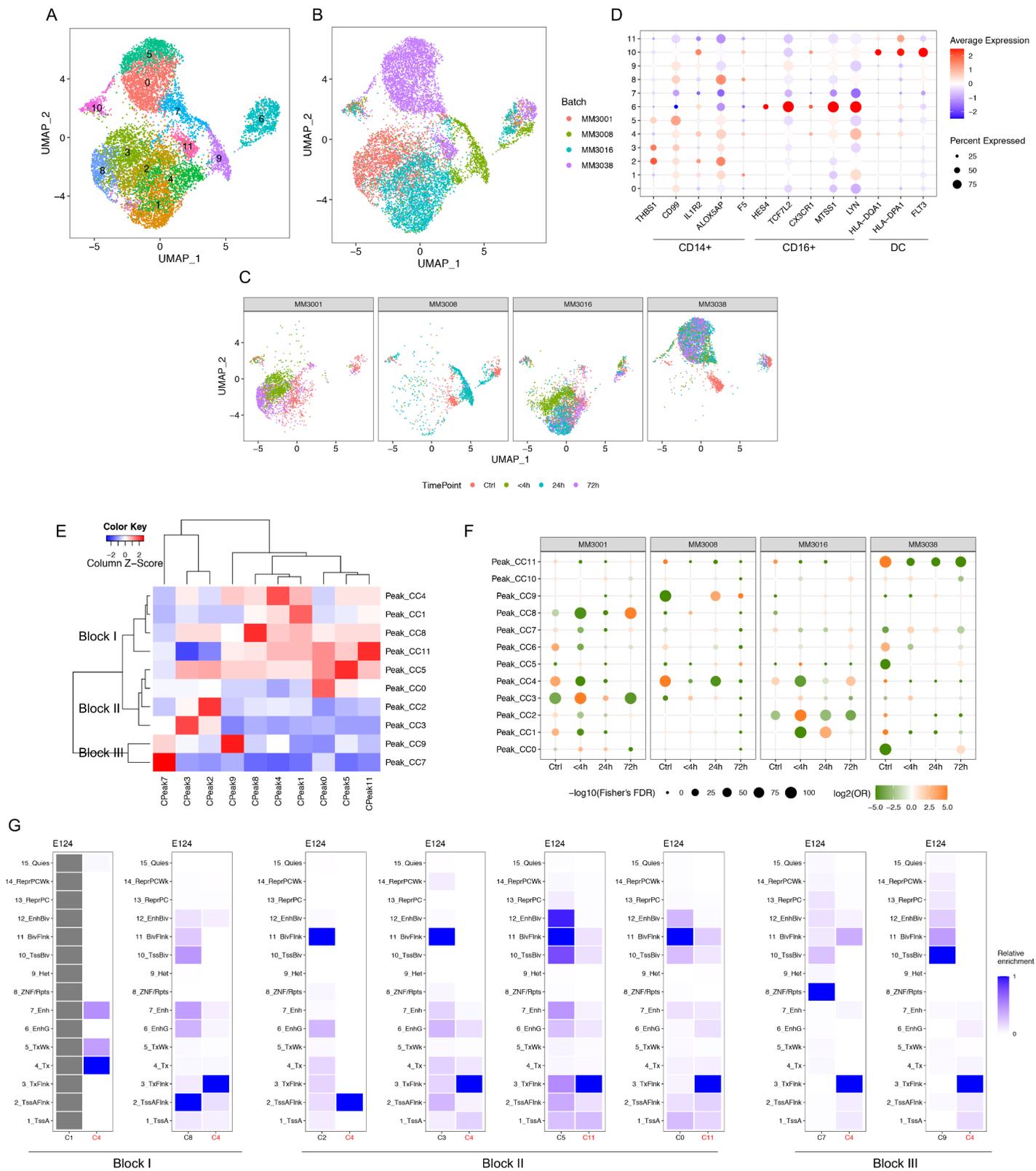


Figure 32 Characterization of the trauma-induced open chromatin profiles in myeloid cells.

(A-C) UMAP plots of peripheral mononuclear myeloid cells color coded by identified cell clusters (A), color coded by experimental batches (B) or color coded by time points and wrapped by individual patients (C). UMAP was performed based on the globally accessibility of peaks. These clusters are also referred as “Peak_CC#” to designate clusters identified based on peak profiles. (D) Gene activities (computed by Signac R package) of myeloid cell markers [70] in different myeloid clusters shown in (A). (E) Hierarchical clustering of myeloid clusters shown in (A). CPeak#: cluster-specific peaks for the corresponding cluster. The matrix represents the average accessibility of CPeaks in each myeloid cluster. Color represents the relative average accessibility of each group of CPeaks across myeloid clusters (column-wise scale). (F) Enrichment between myeloid cell clusters and different time points for each patient. (G) Different enrichment patterns in the ChromHMM 15-state model (E124) across three blocks of CD14+ clusters. The clusters significantly enriched in control samples (C4 and C11) are labelled in red. Gray blocks show unavailable results, since no DA peaks were significantly up-regulated in C1 when compared to C4.

4.3.3 Different regulatory mechanisms for the DA peaks at different CD14+ chromatin states

To interpret the unexpected chromatin state changes observed in block II and III, we further examined the chromatin state for each DA peak. We first performed the 15-state enrichment analysis for each DA peak identified between the clusters from block II and cluster 9 versus the control-associated clusters (C4 and C11). Next, we performed hierarchical clustering. The DA peaks could be generally clustered into five Epigenetic Clusters that we refer to as EC1-5 (Figure 33A). EC3-EC5 represented the peaks associated with active chromatin states in the reference genome. Peaks in EC3 were more specific to enhancer regions (Enh), while EC4 contained peaks in proximity to active transcription start sites (TssA), and EC5 peaks flanked the active TSSs (TssAFlnk). The peaks in EC2 were enriched in bivalent domains. The peaks enriched in other chromatin states were clustered into EC1. The DA peaks were then further divided into trauma or

control groups based on whether the DA peak associated cluster enriched in trauma or control samples, respectively. Because cluster 7 did not have an obvious association with either trauma or control, we excluded C7 from this analysis. This yielded 10 groups of DA peaks based on experimental groups and epigenetic clusters (Figure 33B).

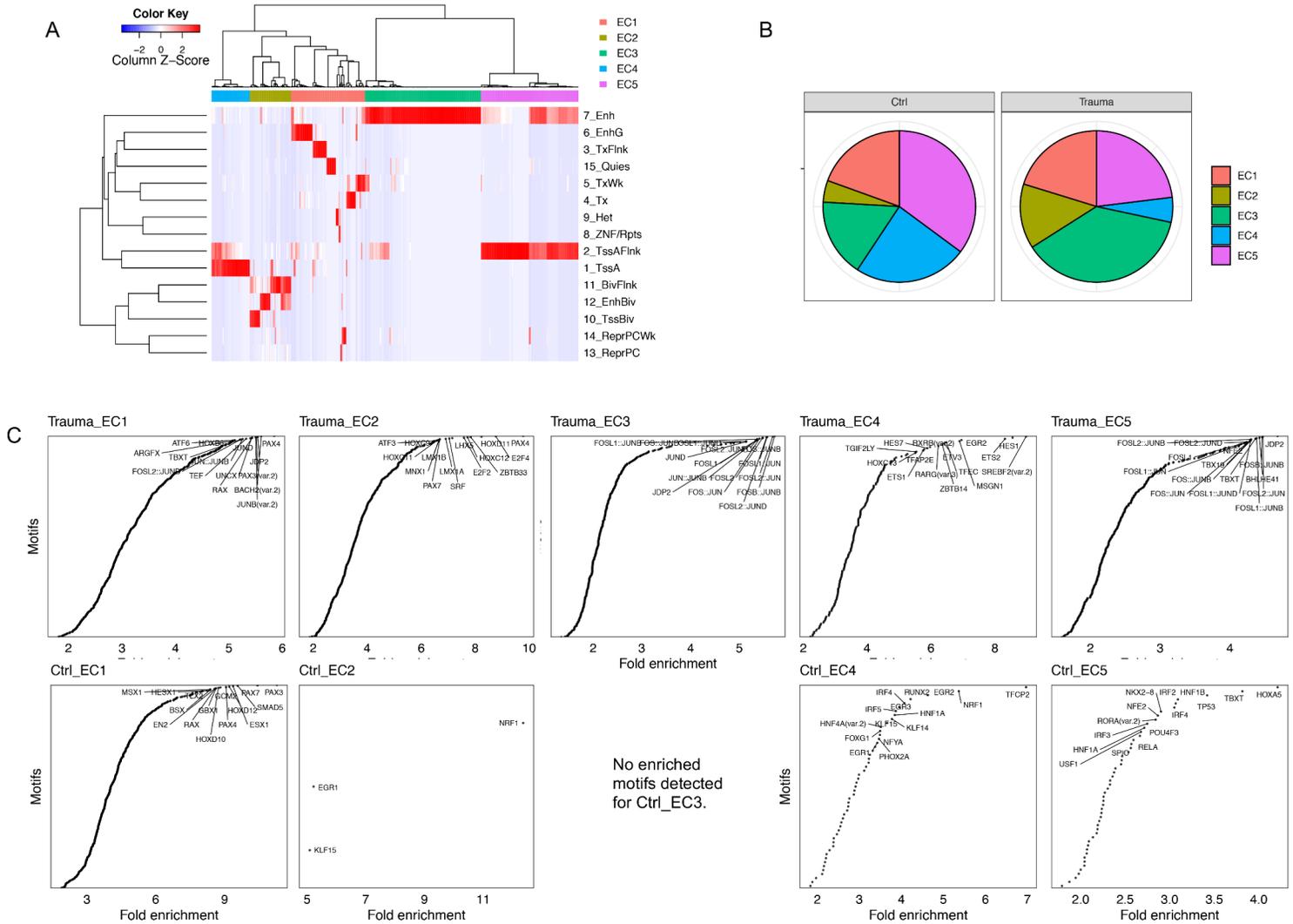


Figure 33 Different regulatory mechanisms for the DA peaks at different chromatin states.

(A) Hierarchical clustering of CD14+ DA peaks based on a matrix of relative fold enrichment in the ChromHMM 15-state model. The peaks were resolved into 5 epigenetic clusters EC1-EC5. (B) The ratio of DA peaks under each EC from control and trauma groups. (C) Motif enrichment results for DA peaks for each EC associated with either trauma or control. All significantly enriched motifs (the number of observed motifs ≥ 5 and $FDR < 0.05$) were plotted and sorted by fold enrichment. The Top 15 enriched motifs are labeled.

We next analyzed for enrichment of transcription factor (TF) binding site motifs in the various DA peaks. Of the peaks more accessible in control CD14+ monocytes (Figure 33C), EC3

had no significantly enriched motifs. The EC4 and EC5 peaks were largely associated with the motifs dominant in steady or differentiated states (i.e., IRF/SPIC/KLF families [45, 71, 72]). In contrast, the active state peaks with more accessibility after trauma (Figure 33C), especially those in EC3 and EC5, were associated with inflammation-related motifs (i.e. FOS/JUN/JUNB/JUND [73]). Previously, we extracted gene signatures from scRNA-seq data derived from circulating CD14⁺ monocytes from severely injured patients [18]. Based on a whole-blood leukocyte gene array dataset from severely injured humans [6], these gene signatures were used to identify two transcriptomic subtypes (Signature Groups: SG1 vs. SG2) [18]. Compared with SG2, SG1 trauma patients exhibited a higher up-regulation of inflammation and greater suppression of MHC II and IFN signaling, and had worse clinical outcomes (prolonged organ dysfunction and higher infection rates). The DA peaks that mapped to the genes present in SG transcriptomic signatures [18] were largely associated with the active state (Figure 34). These results indicated that the up-regulation of inflammation and down-regulation of MHC II or IFN signaling were largely associated with the chromatin regions already accessible at steady-state. After trauma, these regions may undergo additional histone modifications (H3K4me1, H3K27Ac or H3K27me3) leading to enhanced or reduced accessibility.

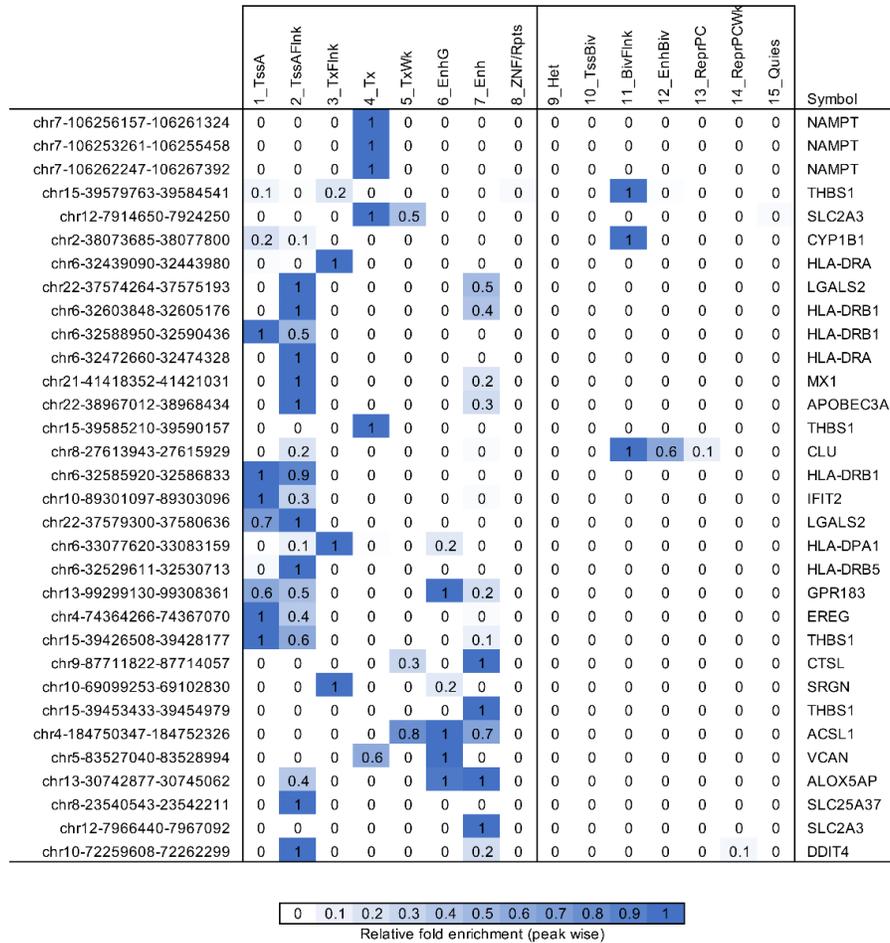


Figure 34 ChromHMM 15-state model enrichment.

Enrichment in the ChromHMM 15-state model (E124) for the DA peaks identified in (Figure 32G) whose nearest genes were also included in SG signatures we previously extracted [18]. The enrichment analysis was performed between a specific chromatin state and a specific peak. SG signatures associated DA peaks were largely enriched in active chromatin states (states 1-8) from the reference epigenome.

Interestingly, the number of peaks at repressed states (EC2, bivalent domains) were markedly increased after trauma (Figure 33B). These peaks were associated with a number of HOX/PAX family TF motifs (Figure 33C), which are known to be critical in the regulation of embryogenesis including morphogenesis [74]. The activities of many developmental genes including those encoding the HOX/PAX family TFs are known to be cross-regulated and in turn

controlled by bivalent chromatin domains and polycomb repressive complexes [75]. The enrichment of HOX and PAX motifs in bivalent chromatin domains conformed to expectations. However, it was not clear what led to this increased accessibility after trauma. Since there was an enrichment of bivalent and polycomb targeted regions in the DA peaks more accessible in block II and III clusters (Figure 32G), we wondered whether de-repression of regions encoding developmental regulators was associated with a wider alteration in chromatin structure in the cells from block II and III.

4.3.4 Characterization of CD14+ cell clustering based on global epigenetic alterations.

To test this hypothesis, we developed a strategy to evaluate global epigenetic alterations (Figure 35A). A state-barcode matrix was generated by counting total cut sites for each of the 15 states in individual cells for all CD14+ monocytes. Seven clusters were identified based on the new count matrix (Figure 35B). These clusters are referred to as “State15_CC#” in order to distinguish these clusters from the Peak_CC# described in Figure 32. By hierarchical clustering, the seven State15 clusters could be generally classified into two categories (Figure 35C). State15_CC4, 5, 2 and 1 aligned with the global accessibility pattern similar to the reference epigenome. In contrast, State15_CC0, 3 and 6 exhibited greater accessibility in regions that were less accessible in the reference epigenome. These accessible epigenomic regions were in turn linked to polycomb targeted or heterochromatin domains. Furthermore, the Peak_CC# clusters in block I were largely enriched in the State15_CC4, 5, 2 or 1, while the Peak_CC# clusters in block II and III were mostly enriched into State15_CC0, 3 or 6 (Figure 35D), supporting our hypothesis that block II and III clusters associated with wider alterations of chromatin structure.

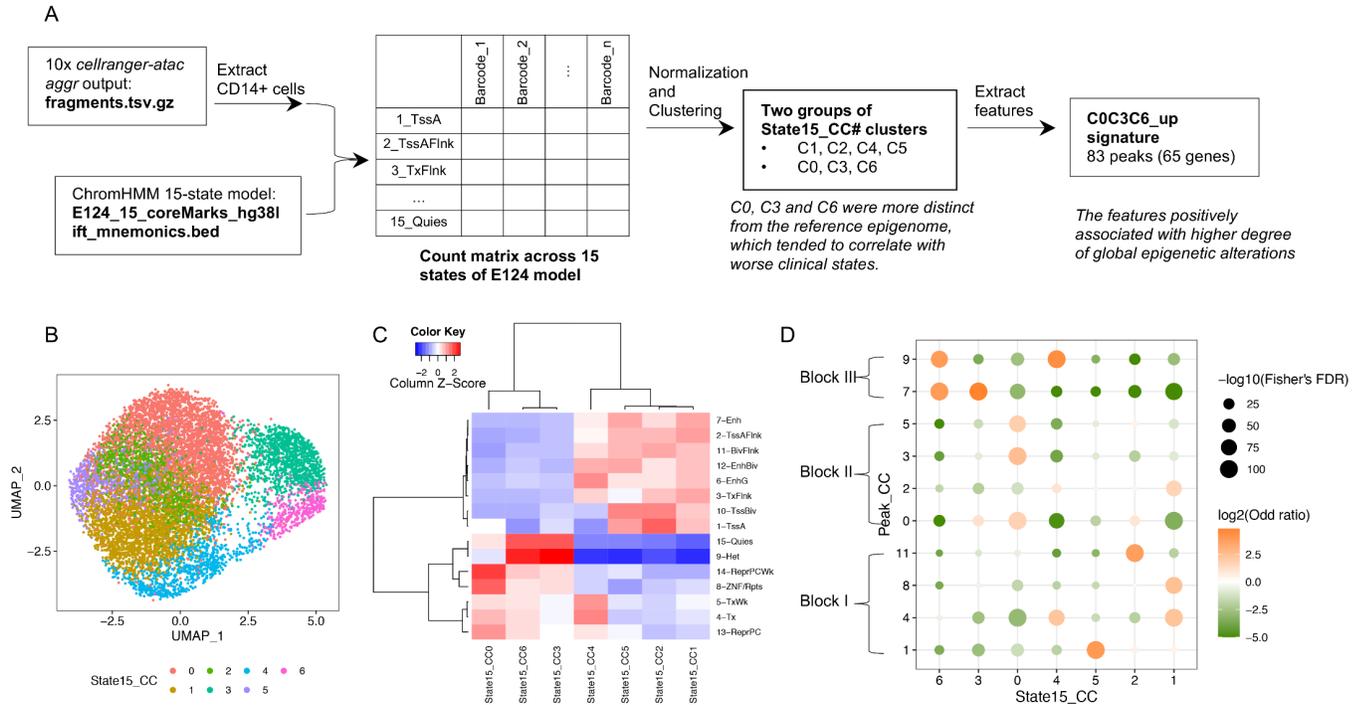


Figure 35 Characterization of global epigenetic changes across ChromHMM 15 states in CD14+ monocytes.

(A) Schematic of workflow of how we generated a state-barcode count matrix and extracted the COC3C6_up signature. (B) UMAP was performed and color coded by the cell clusters identified based on the state-barcode matrix. These clusters are referred to as “State15_CC#”. (C) Hierarchical clustering of the State15_CC# clusters shown in (B). (D) Co-occurrence between the Peak_CC# clusters (shown in Figure 32) and State15_CC# clusters, tested by Fisher’s Exact test. p values were corrected by Benjamini-Hochberg method for multiple testing.

4.3.5 Generation and validation of gene signatures representing the global epigenetic alterations.

Next, we sought to directly extract the features associated with the global epigenetic alterations by identifying the DA peaks between State15_CC0, 3, and 6 vs. State15_CC1, 2, 4 and 5 (Figure 35A, Table 7 and Appendix Spreadsheet 6). Interestingly, 83 peaks were significantly

more accessible in State15_CC0, 3, and 6, while no peaks were significantly more accessible in State15_CC1, 2, 4, and 5. These 83 peaks corresponded to 65 unique genes, which we refer to as the “C0C3C6_up” signature (peaks or genes). Among MSigDB [42] (v5.2) C2 curated gene sets (Appendix Spreadsheet 6), the top enriched gene sets ($p < 0.05$ & the number of overlapping genes ≥ 5 , sorted by decreasing fold enrichment) associated with C0C3C6_up signature largely involved bivalent domains and polycomb targets (SUZ12 or domains with H3K27me3). Consistently, the top enriched GO terms were mostly associated with neuron development or morphogenesis, which are known to involve developmental genes marked by bivalent domains and regulated by polycomb-group proteins (PcG) [75].

Table 7 C0C3C6_up signature

Peak	Nearest gene
chr3-93470153-93471006	<i>PROS1</i>
chrY-56825450-56851853	<i>SPRY3</i>
chr21-10693195-10738780	<i>TPTE</i>
chr21-10413349-10432882	<i>BAGE2</i>
chr9-40922055-40931475	<i>MIR1299</i>
chr21-7930519-7958237	<i>KCNE1B</i>
chrY-11290781-11297100	<i>GYG2P1</i>
chr5-49599254-49603103	<i>EMB</i>
chr21-10329531-10333013	<i>BAGE2</i>
chrY-11304439-11323747	<i>GYG2P1</i>
chr16-46385961-46392033	<i>ANKRD26P1</i>
chr15-20338949-20359301	<i>CHEK2P2</i>
chr6-168048129-168057957	<i>FRMD1</i>
chr16-34581040-34589122	<i>UBE2MP1</i>
chr9-40906828-40916361	<i>MIR1299</i>
chr5-49656325-49662011	<i>EMB</i>
chr20-31057589-31069509	<i>DEFB115</i>
chr21-10457775-10466597	<i>BAGE2</i>
chrY-11331673-11334570	<i>GYG2P1</i>
chr5-131984427-132004795	<i>ACSL6</i>
chr14-99173913-99220651	<i>BCL11B</i>
chr10-86653791-86671651	<i>LDB3</i>
chr14-99223431-99275903	<i>BCL11B</i>
chr10-46519808-46531356	<i>GPRIN2</i>
chr22-11033735-11039390	<i>FRG1FP</i>
chr9-41229221-41230848	<i>MIR4477A</i>
chr21-7915852-7929905	<i>KCNE1B</i>
chr3-13625519-13637992	<i>FBLN2</i>
chr13-18211584-18213670	<i>FAM230C</i>
chr20-28804507-28817951	<i>FRG1CP</i>
chr1-226976773-226988501	<i>COQ8A</i>
chr1-228357438-228379773	<i>OBSCN</i>
chr10-41902942-41915241	<i>LOC441666</i>
chr2-240711188-240726164	<i>KIF1A</i>
chr10-38526648-38530001	<i>LINC00999</i>
chr2-240889169-240900861	<i>MAB21L4</i>
chr10-49735174-49746098	<i>OGDHL</i>
chr19-27352841-27362955	<i>LINC00662</i>

Table 7 continued

Peak	Nearest gene
chr22-10775030-10783980	<i>FRG1FP</i>
chr8-142520691-142550077	<i>ADGRB1</i>
chr1-227919342-227928735	<i>MIR5008</i>
chr14-16096743-16105285	<i>OR11H12</i>
chr7-149810832-149829633	<i>SSPO</i>
chr2-232476744-232489024	<i>ECEL1</i>
chr1-228332736-228347496	<i>OBSCN</i>
chr4-49116744-49121450	<i>CWH43</i>
chr2-89826282-89837073	<i>LOC101927050</i>
chr5-180618615-180634135	<i>FLT4</i>
chr17-16922328-16936127	<i>TBC1D27P</i>
chr16-46392851-46396684	<i>ANKRD26P1</i>
chr20-29873849-29879518	<i>FRG1EP</i>
chr7-30905464-30917521	<i>AQP1</i>
chr1-53076915-53089065	<i>PODN</i>
chr1-228285806-228295363	<i>OBSCN</i>
chr15-29101631-29119156	<i>APBA2</i>
chr7-152402010-152407690	<i>KMT2C</i>
chr3-75667652-75670740	<i>LINC00960</i>
chr3-127026273-127040010	<i>PLXNA1</i>
chr1-227883027-227897635	<i>MIR5008</i>
chr12-47974824-47991735	<i>COL2A1</i>
chr11-128462571-128480938	<i>ETSI</i>
chr3-133831782-133842339	<i>SRPRB</i>
chr21-8987971-8996653	<i>MIR3648-2</i>
chr2-134457484-134464875	<i>TMEM163</i>
chr16-34590530-34596332	<i>UBE2MP1</i>
chr8-12583550-12589880	<i>LOC729732</i>
chr20-30811892-30816844	<i>MLLT10P1</i>
chr2-130822654-130829713	<i>ARHGEF4</i>
chr10-48444884-48452694	<i>ARHGAP22</i>
chr16-46398079-46401836	<i>ANKRD26P1</i>
chr10-86678858-86687490	<i>LDB3</i>
chr20-31070330-31076126	<i>DEFB115</i>
chr2-240756647-240774217	<i>KIF1A</i>
chr10-43213785-43222742	<i>RASGEF1A</i>
chr10-43036725-43047037	<i>RET</i>
chr1-43595745-43610176	<i>PTPRF</i>

Table 7 continued

Peak	Nearest gene
chr1-2651463-2656432	<i>MMEL1</i>
chr14-104098469-104109681	<i>ASPG</i>
chr3-126957013-126966118	<i>CHCHD6</i>
chr1-10650776-10672122	<i>CASZI</i>
chr18-79451022-79463960	<i>NFATC1</i>
chr7-100949044-100961824	<i>MUC3A</i>
chr4-2064429-2078691	<i>POLN</i>

We sought to validate the new gene signature inferred from profiling of chromatin states using a comprehensive trauma dataset. We queried a large-scale whole-blood leukocyte transcriptomic dataset from trauma patients (n=167) [18] (Figure 36A). Compared with the healthy controls, the C0C3C6_up signature was up-regulated after trauma. Notably, this upregulation tended to persist in patients with a slow recovery based on persistent organ dysfunction (time-to-recovery [TTR] ≥ 14 days) and was maintained at an even higher level in the patients that failed to recover within 28 days after injury (Figure 36B). We extracted the genes from this bulk dataset that were highly correlated with the C0C3C6_up signature (Spearman correlation: $|\rho| > 0.3$), and performed GSEA [31] using the correlation coefficient as the rank (Figure 36A). The gene sets which included bivalent domains and polycomb targets were highly and significantly enriched in genes positively correlated with the C0C3C6_up signature (Figure 36F-36H, Appendix Spreadsheet 7). Thus, the transcriptional signature that we extracted from single-cell global epigenetic alterations induced by trauma (4 representative patients) was reflected in a bulk whole-blood leukocyte transcriptomic dataset (167 patients). In addition, the gene sets enriched in the genes negatively correlated with global epigenetic signature were associated with DNA repair and RNA processing (Appendix Spreadsheet 7), indicating that the global epigenetic alterations we identified involve in multiple biological processes. Importantly, no genes from the previous SG

signatures [18] showed a Spearman correlation coefficient $|\rho| > 0.3$ (Figure 36C-36E), indicating that the chromatin alterations revealed by epigenetic profiling in this study, while confirming the earlier work, yielded a new set of actionable trauma induced epigenomic changes.

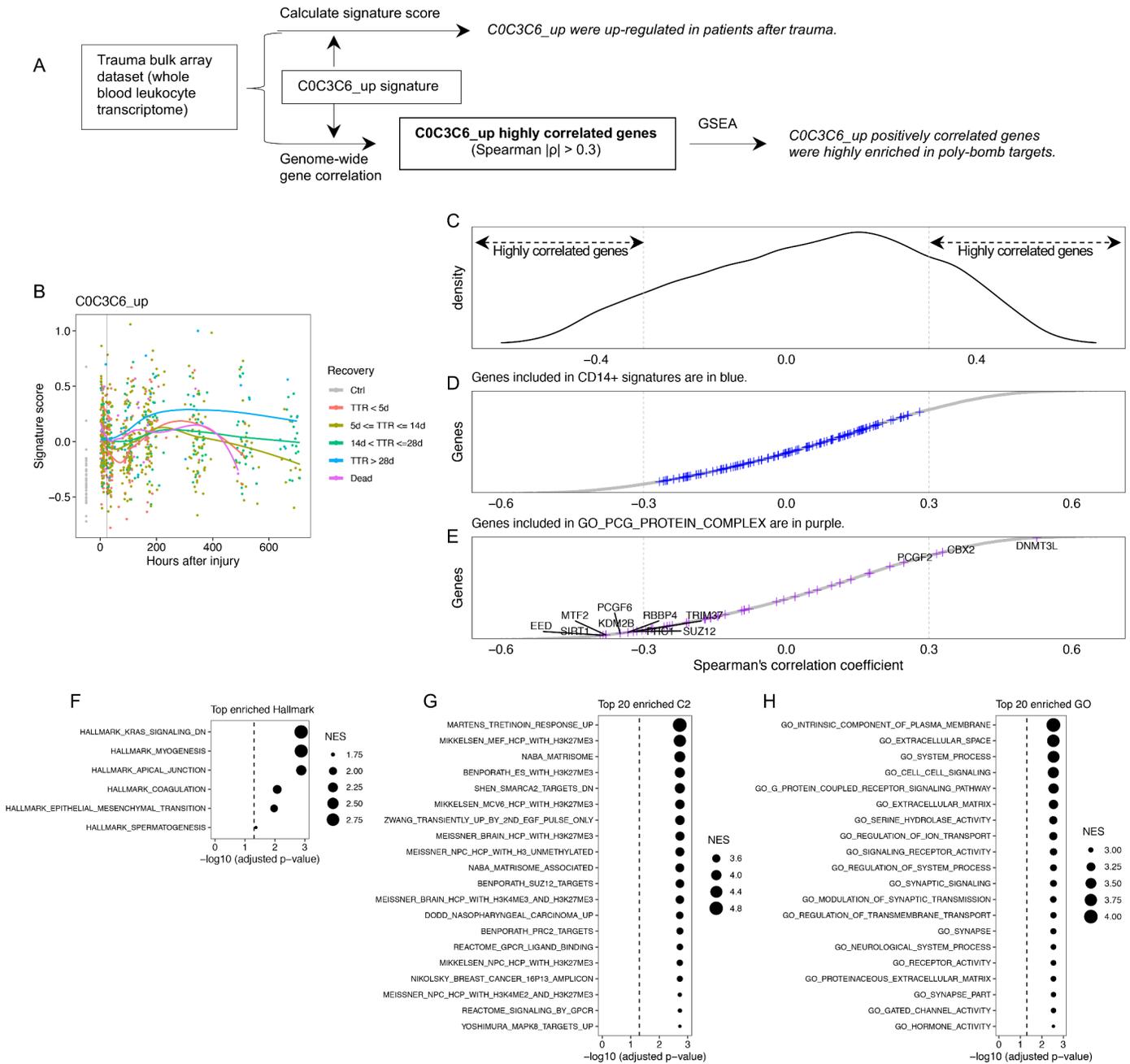


Figure 36 Validation of the C0C3C6_{up} signature in trauma bulk microarray data.

(A) Schematic of the workflow of the analyses shown in this figure. (B) The changes in C0C3C6_{up} signature scores after trauma, color coded by different clinical trajectories (TTR: Time-To-Recovery). (C-E) Spearman's correlation coefficients were computed for a genome-wide gene correlation with the C0C3C6_{up} signature score. Density plot of correlation coefficients ρ were plotted in (C). $|\rho| = 0.3$ was labeled as the vertical dashed lines. The genes from the

CD14⁺ signatures we previously used to define SG subtypes [18] were color coded in blue as shown in (D), and were largely associated with inflammation, MHC II expression and IFN signaling. The genes associated with PcG protein complex were color coded in purple as shown in (E). The genes with $|\rho| > 0.3$ in either category above were further annotated with the gene symbols. (F-H) GSEA results of hallmark gene sets (F), curated gene sets (G) and GO terms (H) (MSigDB gene sets v5.2) using the highly correlated genes identified above ($|\rho| > 0.3$). The significantly enriched pathways associated with genes positively correlated with C0C3C6_up are shown (adjusted p-value < 0.05 and NES > 0) and sorted by normalized enrichment scores (NES). Significance (adjusted p-value = 0.05) was annotated by the vertical dashed line. If the number of significantly enriched pathways was more than 20, only the top 20 pathways are shown in the figure.

4.3.6 Global epigenetic alterations were the common changes across major immune cell types.

Because the genes associated with morphogenesis or neuron development would not be expected to be expressed in myeloid, or even in hematopoietic lineages, we wondered whether other types of immune cells also underwent similar changes. To assess this, we reanalyzed the scATAC-seq data and calculated the average accessibility of the C0C3C6_up signature (83 peaks) in other major immune cell types. Notably, B cells, NK and T cells, DC and CD16⁺ monocytes also showed an increase in the accessibility of the C0C3C6_up signature after trauma that was observed in CD14⁺ monocytes (Figure 37).

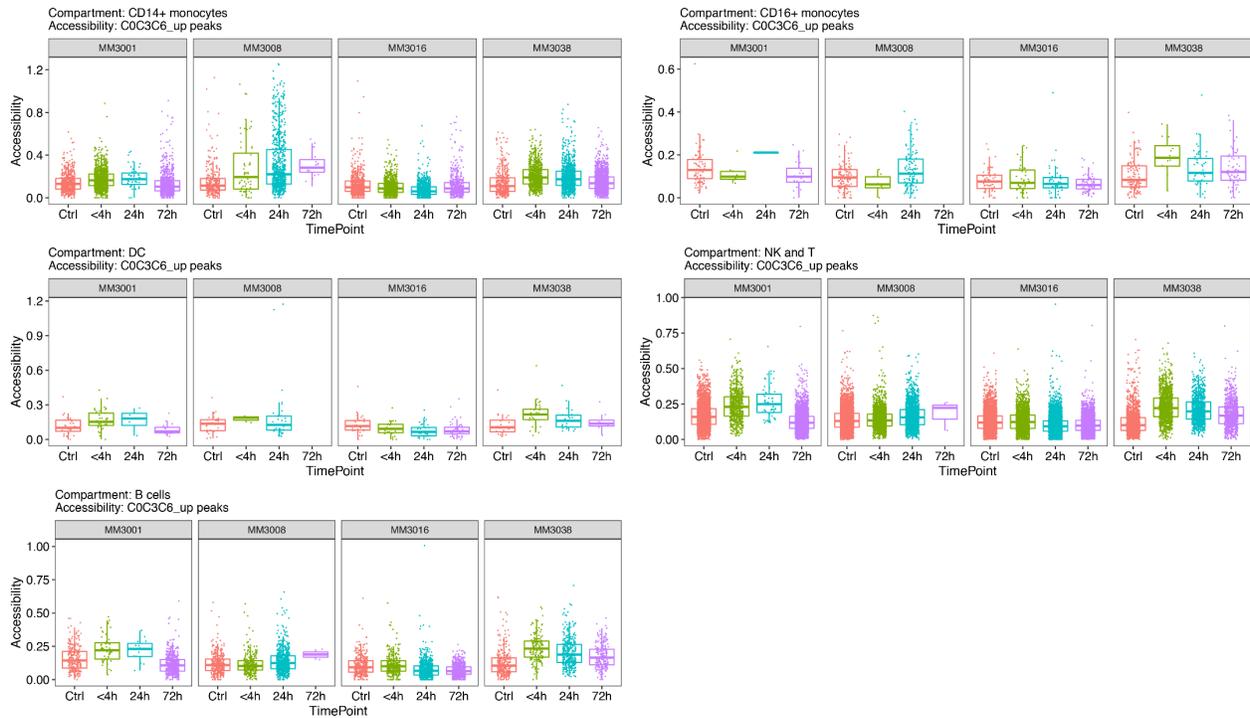


Figure 37 The changes in the accessibility of C0C3C6_up peaks in other immune cell types besides CD14+ monocytes.

The average accessibility of C0C3C6_up peaks in CD14+ monocytes, CD16+ monocytes, DC, NK and T, and B cells is shown for four trauma patients across time and for healthy controls.

4.3.7 The independent prognostic value of epigenetic subtypes in trauma patients.

Next, we sought to determine if we could identify patient heterogeneity based on gene signatures for the global epigenetic response to systemic injury. Using the trauma bulk gene array from 167 severely injured patients, we extracted the C0C3C6_up signature genes to cluster the patients. This was done based on 37 genes from the C0C3C6_up signature present in the trauma bulk gene array data. Based on the first timepoint (~12 h after injury), patients were clustered into three groups (Epigenetic Group, EG1-EG3) (Figure 38A, Appendix Spreadsheet 8). EG3 was the

most distinct and had the highest expression of C0C3C6_up signature genes and polycomb targets (Figure 38B and Figure 39A). Of the genes associated with the PcG complex, EG3 trauma patients tended to express lower PRC2 core components (SUZ12, EED, EZH2, RBBP4/7), lower PRC2 cofactors (AEBP2, MTF2), and lower SIRT1 (maintaining gene silencing [76] and genome stability [77]), but higher H3K27me3 demethylases (KDM6A and KDM6B) compared with EG1 and EG2 trauma patients (Figure 39B). These changes suggest global epigenetic alterations associated with increased accessibility in chromatin structure and de-repression of polycomb targets in a subset of severely injured patients (EG3).

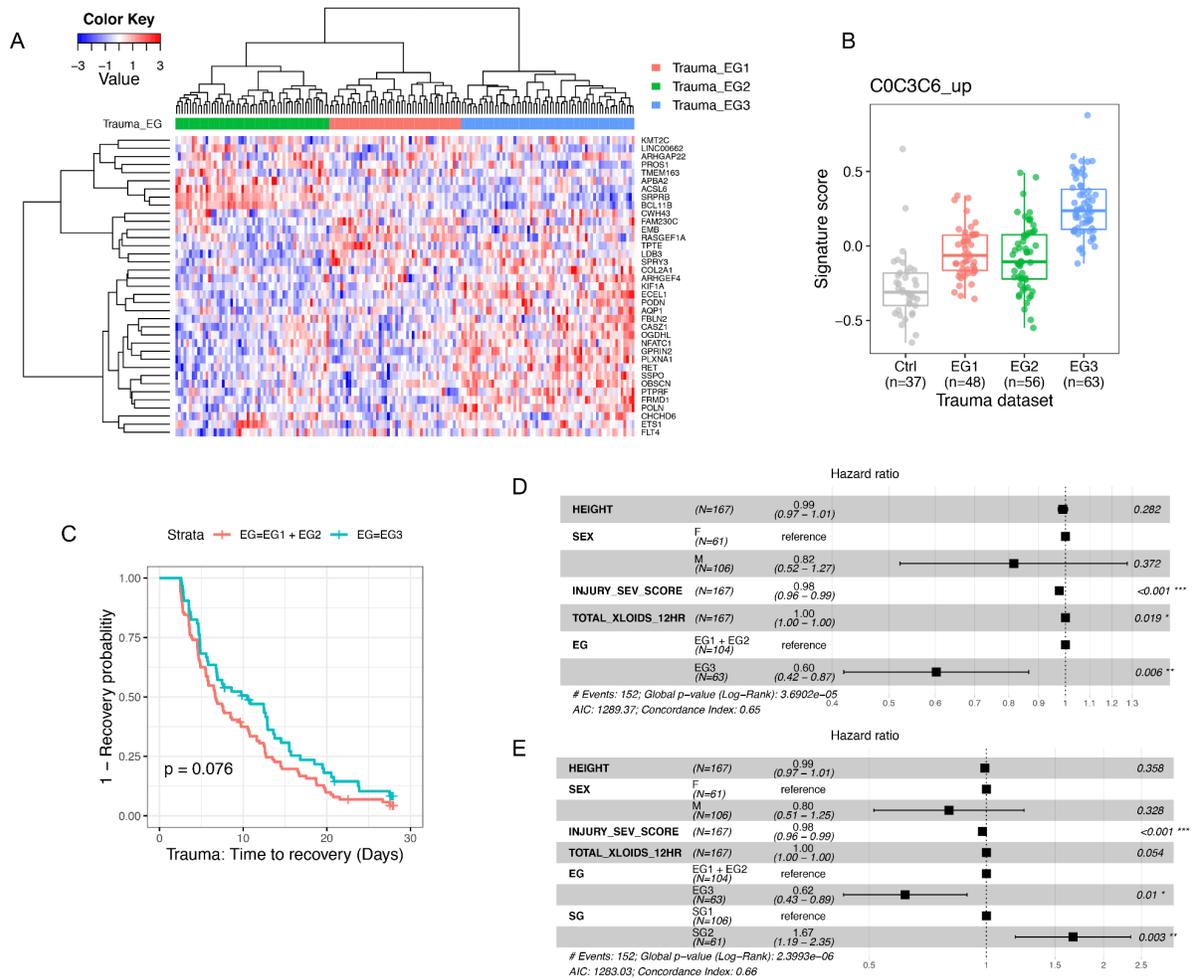


Figure 38 Epigenetic subtypes and prognostic value in trauma patients.

(A) Hierarchical clustering of trauma patients by C0C3C6_up signature genes with an available expression value in the trauma dataset. The 1st sampled time point (all within 12 hrs of injury) for each patient (n = 167) was extracted for the clustering analysis. Patients were clustered into different subtypes referred to as epigenetic groups (EGs). (B) C0C3C6_up signature scores among trauma EGs derived from (A) and the healthy controls. (C-E) Time-to-event analysis between Trauma_EG3 (highest level of C0C3C6_up signature) versus Trauma_EG1+EG2. Event was set as recovery (absence of organ dysfunction). (C) Univariate analysis by Kaplan-Meier estimate. (D) Multivariate analysis using Cox model to adjust potential co-variants of EGs. (E) Multivariate analysis using Cox model to further adjust for SG subtype designations.



Figure 39 Expression of epigenetic signatures or genes among trauma Epigenetic Groups (EGs) along with healthy controls.

(A) Signatures of polycomb targets. (B) Critical epigenetic regulators, including GO: PCG_PCG_PROTEIN_COMPLEX along with H3K27me3 demethylase <KDM6A, KDM6B> and CTCF.

To further characterize the global transcriptomic profiles between trauma epigenetic subtype EG3 vs. EG1+EG2, we identified the DEGs between these two groups of patients (Figure 40A). The genes up-regulated in EG3 trauma patients were largely associated with bivalent domains and polycomb targets (PRC2, EED, SUZ12 or domains with H3K27me3) (Figure 40B-40D, Appendix Spreadsheet 10). This was highly consistent with the GSEA analysis using C0C3C6_up highly correlated genes (Figure 36F-36H, Appendix Spreadsheet 7) and the gene set over-representation test directly using C0C3C6_up signature genes (Appendix Spreadsheet 6). This indicated that the 37 gene subset from the C0C3C6_up signature genes could be viewed as representative of the global epigenetic alterations and sufficient to define epigenetic subtypes in trauma patients.

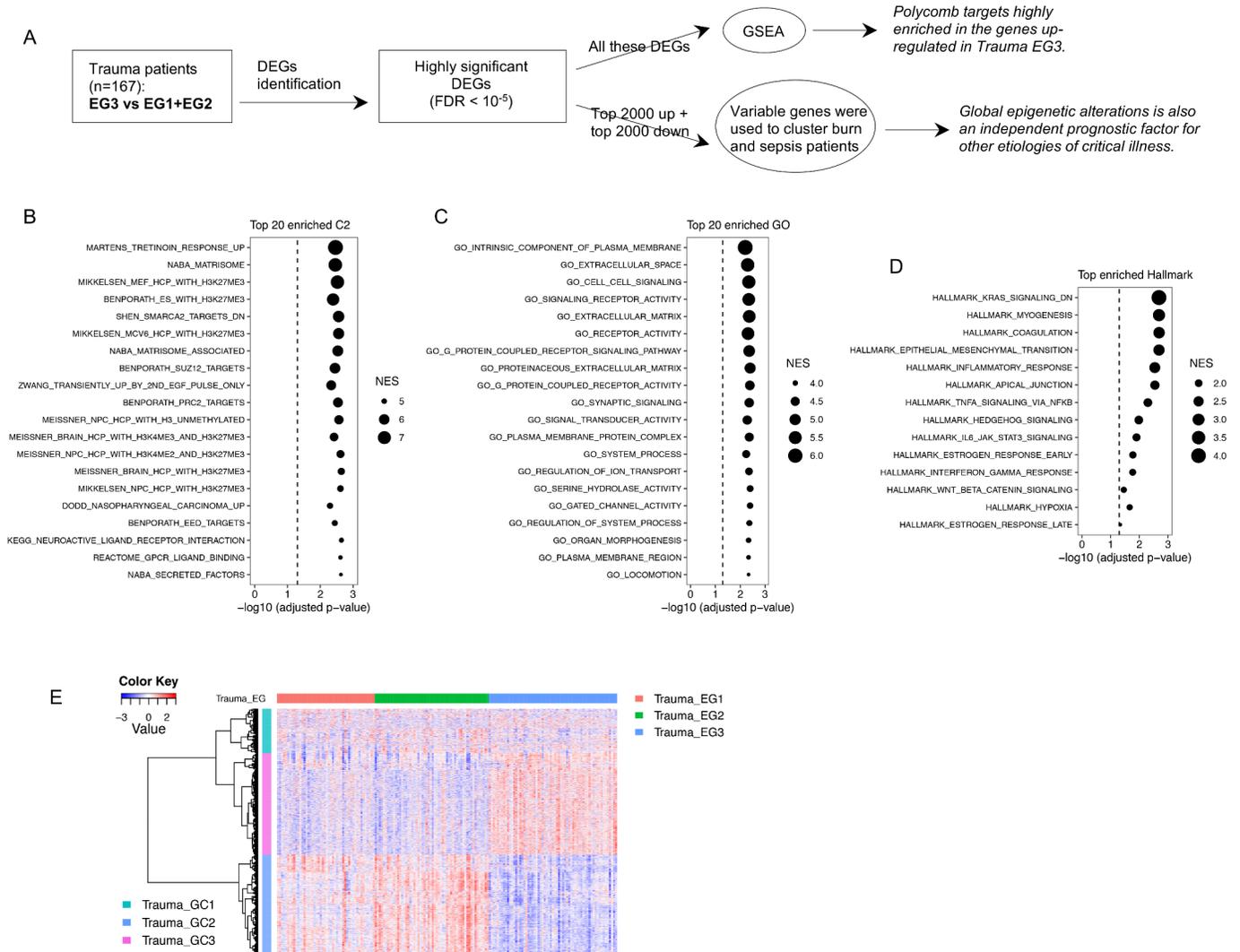


Figure 40 Gene expression profile among trauma Epigenetic Groups (EGs) (related to Figure 38).

(A) Schematic workflow for generating trauma epigenetic DEGs that were further used to cluster burn and sepsis patients. (B-D) GSEA results of curated gene sets (B), GO terms (C) and hallmark gene sets (D) (MSigDB gene sets v5.2), using highly significant DEGs ($FDR < 10^{-5}$ identified by Limma R package). The significantly enriched pathways associated with Trauma_EG3 are shown (adjusted p-value < 0.05 and $NES > 0$) and sorted by the normalized enrichment score (NES). Significance (adjusted p-value = 0.05) was annotated as the vertical dashed line. If the number of significantly enriched pathways was more than 20, only the top 20 pathways are shown in the figure. (E) The profile of top 4000 DEGs (top 2000 up and top 2000 down DEGs in Trauma_EG3) among trauma patients. The 4000 DEGs can be clustered into 3 gene clusters Trauma_GC1-GC3. These genes were used to cluster burn and sepsis

patients, and additional filter was applied in order to only keep the variable genes with standard deviation of scaled expression ≥ 0.5 among the analyzed samples.

We next compared clinical outcomes between the EG3 versus the EG1+EG2 trauma patients. EG3 trauma patients tended to show a slower recovery (Kaplan-Meier analysis, log-rank p value = 0.076, Figure 38C), compared with EG1+EG2. The co-variants of sex, height, ISS (Injury Severe Score) and total crystalloids received within 12 hours after injury (TOTAL_XLOIDS_12HR, a marker of worse disease state) were significantly and differentially distributed between EG3 patients vs. EG1+EG2 (Appendix Spreadsheet 9). After adjusting for these co-variants using Cox regression, the p value for EG classification became highly significant (p = 0.006, Figure 38D). Noticeably, after adding SG designation [18], the transcriptomic subtypes we previously defined, into the Cox model both SG1 and EG3 were significantly associated with slower recovery (Figure 38E). This analysis suggests there are two different mechanisms that contribute to the differential prognosis following systemic injury that are reflected by the SG and EG signatures.

The gene sets of polycomb targets (PRC2/SUZ12/EED/H3K27me3) were derived from human embryonic stem cells (ES) [78]. Thus, these polycomb targets should include both hematopoietic and non-hematopoietic developmental genes. To explore whether the de-repressed targets are relevant to hematopoiesis and immune cell states, we plotted the gene expression profiles of all polycomb targets (union of the 4 gene sets) across trauma patients (Figure 41A). These genes fell into 3 gene clusters (PcGtargets_TraumaGC1-GC3). Interestingly, the PcG targets relevant to hematopoietic lineages (e.g. leukocyte/lymphocyte differentiation/regulation) were much more enriched in PcGtargets_TraumaGC2, which was suppressed in EG3 patients. The de-repressed polycomb targets (PcGtargets_TraumaGC1+GC3) were largely associated with

morphogenesis or non-hematopoietic tissue development (Figure 41B). This evidence suggests that the de-repression of polycomb targets, encoding non-hematopoietic genes, in immune cells of EG3 trauma patients, may lead to their dysfunctional states.

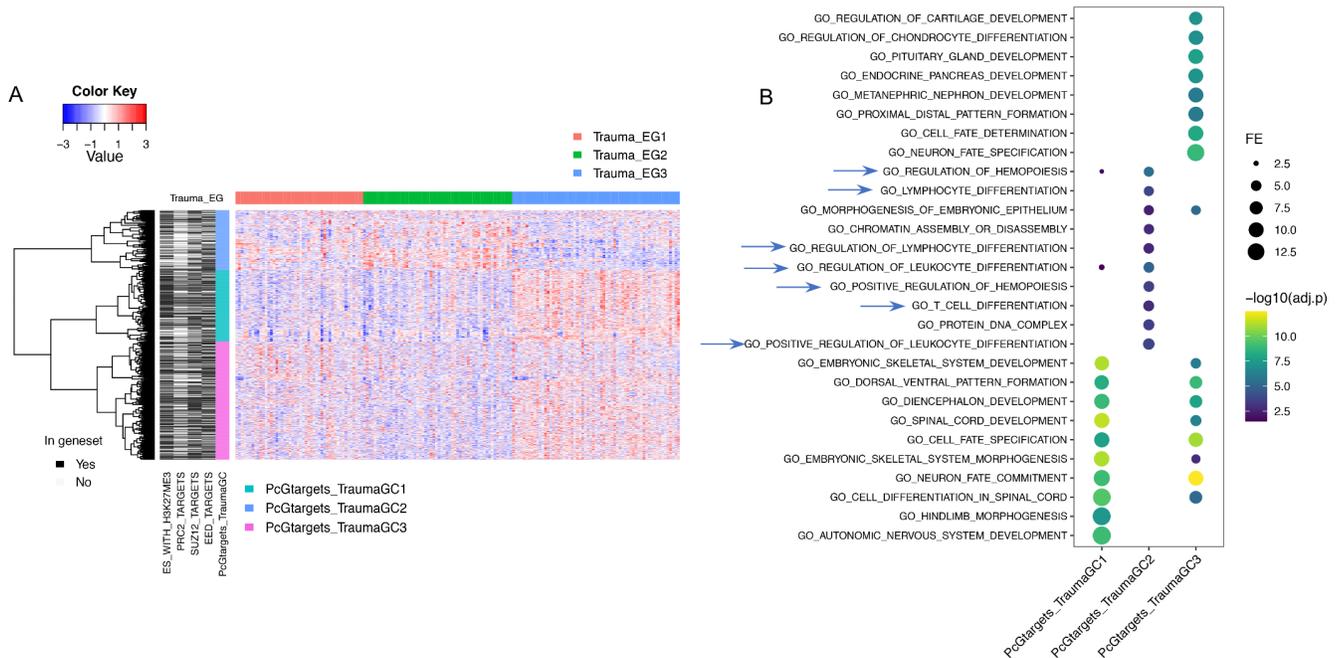


Figure 41 Gene expression profile of polycomb targets across trauma patients shown in (Figure 38A).

(A) Polycomb targets (rows) were collected as the union of 4 gene sets: BENPORATH_ES_WITH_H3K27ME3, BENPORATH_SUZ12_TARGETS, BENPORATH_PRC2_TARGETS, BENPORATH_EED_TARGETS (MSigDB gene sets v5.2). Patients (columns) were sorted by EG subtypes identified in (Figure 38A). Polycomb targets were clustered into 3 clusters, PcGtargets_TraumaGC1-GC3. (B) GO enrichment for the 3 gene clusters identified in (A). Top 10 enriched GO terms (sorted by fold enrichment (FE) decreasingly) for each gene cluster were shown (threshold for significance: $FDR < 0.05$ & the number of overlapped genes ≥ 10 & $FE > 2$). Hematopoietic relevant GO terms were pointed out by blue arrows.

4.3.8 Recapitulation of epigenetic profile in scRNA data of trauma patients

Since we had previously generated scRNA-seq data for these 16 samples used for ATAC-seq profiling, we wondered whether the global epigenetic alterations were also reflected in the scRNA-seq dataset. Thus, we used the top 500 up and top 500 down DEGs identified in EG3 trauma patients (817 of these genes were detectable in the scRNA data) to cluster the scRNA data from CD14+ monocytes (Figure 42A-42E). Consistently, the control samples mostly clustered together and were separated from trauma patients by these top DEGs. Furthermore, the DEGs that were up-regulated in EG3 trauma patients were largely separated from the down-regulated DEGs (Figure 42A). We noticed that the epigenetic DEGs were expressed at a much lower level and with less variation than the previously characterized genes associated with the SG signatures [18] (Figure 42F-42I). It is not surprising that genes associated with inflammation, IFN and MHC II signaling dominate the transcriptomic analysis [14, 15, 18, 58]. The full range of epigenetic-associated genes are likely to be more challenging to detect in scRNA datasets due to data sparsity and transcript abundance. Thus, scATAC-seq allowed us to extract genes directly associated with global epigenetic alterations that could not be recovered by scRNA analysis alone.

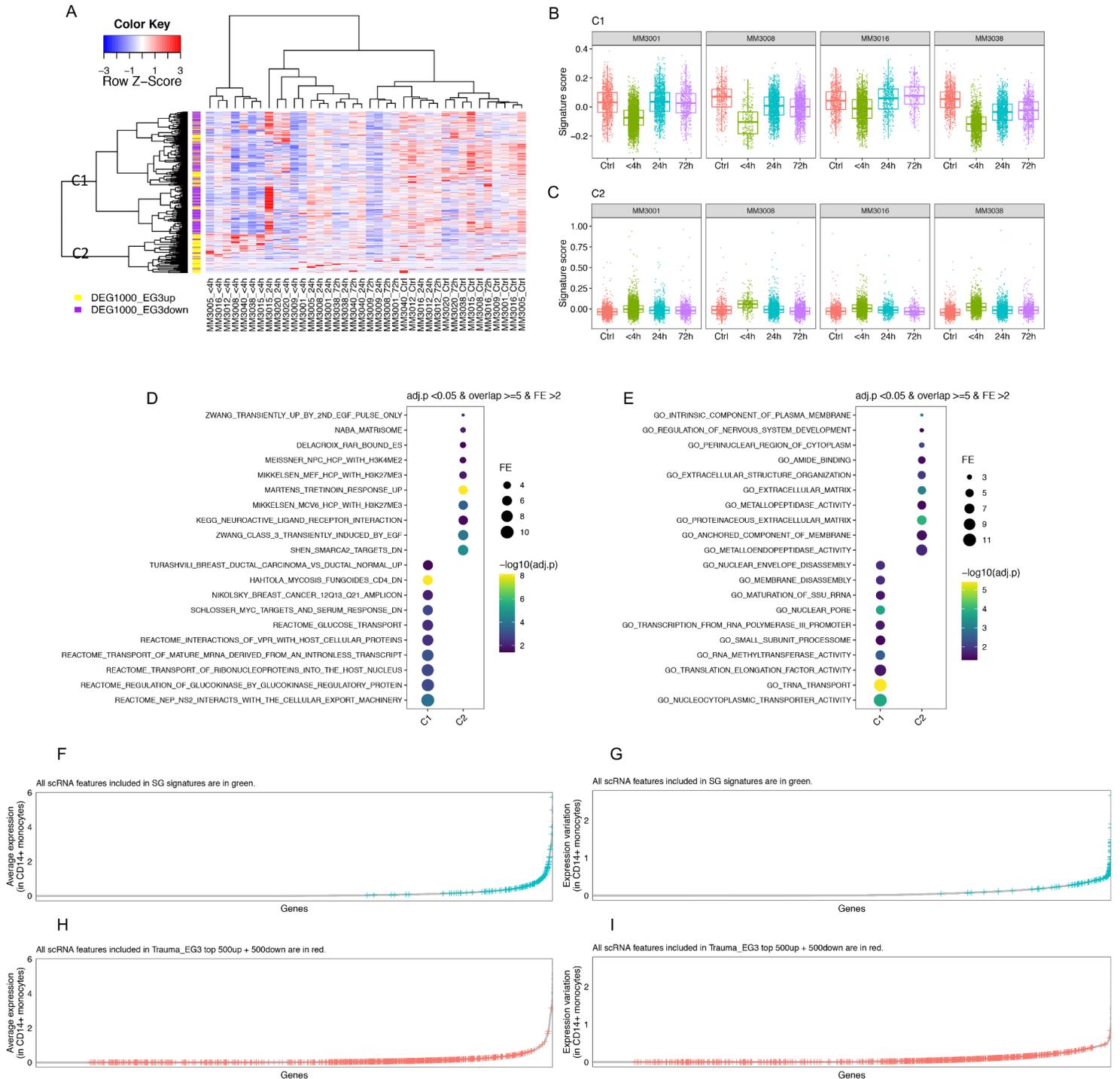


Figure 42 Recapitulate epigenetic profile in scRNA data.

(A) Top 500 up-regulated and top 500 down-regulated DEGs (row colors) in trauma EG3 patients were used to cluster CD14+ monocytes scRNA data. Pseudobulk matrix averaged by different samples was shown. These DEGs were

clustered into two groups of genes, C1 and C2. C1 genes largely corresponded to the down-regulated DEGs in trauma EG3 patients, and C2 genes largely mapped to the up-regulated DEGs in trauma EG3 patients. (B-C) Visualization of C1 genes (B) and C2 genes (C) among the 16 samples subjected to scATAC-seq. (D-E) Top 10 enriched gene sets for C1 or C2 genes. (D) MSigDB C2 curated gene sets. (E) MSigDB Hallmark gene sets. (F-I) Differences in distribution of average expression and expression variation of SG signatures and trauma EG3 DEGs.

4.3.9 The independent prognostic value of global epigenetic changes in burn and sepsis patients.

We have previously shown that the SG subtypes and the association with differential prognosis can be recapitulated using whole-blood leukocyte transcriptomic datasets from burn and sepsis patients [18]. To determine if our findings on global epigenetic heterogeneity could also be identified in burn and sepsis patients, we sought to use C0C3C6_up signature genes to cluster burn and sepsis patients. However, the patient clusters based on C0C3C6_up signature genes did not show an obvious association with survival. This observation led us to hypothesize that while these other etiologies for acute critical illness are likely to be influenced by epigenetic processes, the gene co-expression patterns are likely to be different from trauma. Hence, the C0C3C6_up signature may be more specific to trauma.

Therefore, we took the following two steps to identify the potential epigenetic-associated genes for clustering burn or sepsis patients. In step 1, we extracted the top 2000 up-regulated and top 2000 down-regulated genes in Trauma_EG3 patients vs. Trauma_EG1+EG2 patients as the initial pool of features. These 4000 top DEGs generally included three gene clusters: Trauma_GC3 (highly expressed in EG3 trauma patients), Trauma_GC2 (markedly suppressed in EG3 trauma patients) and a small fraction of relatively less variable genes (Trauma_GC1) (Figure 40E). A

relatively low threshold for including DEGs was used to maximize the pool of DEGs that can be potentially influenced by global epigenetic alterations across different etiologies of critical illness. In Step 2, we used only the variable genes (standard deviation of scaled expression ≥ 0.5 among the analyzed samples) from the top 4000 DEGs to cluster the burn or sepsis patients.

The Burn dataset is a longitudinal dataset. We extracted the 1st time point from all adult burn patients (≥ 18 yr, $n = 121$). Based on the two steps described above, 1482 genes were used to cluster burn patients (Figure 43-44) leading to three epigenetic groups (Burn_EG1-EG3) and two gene clusters (Burn_GC1-GC2) (Figure 43A and Appendix Spreadsheet 11-12). EG2 burn patients had a higher expression of the genes in the C0C3C6_up signature (Figure 43B) and underwent a worse prognosis (K-M analysis: log-rank $p = 0.019$, Figure 43D). After adjusting for potential co-variants and SG classification, EG2 burn patients still showed a trend towards worse survival compared with the other burn patients ($p = 0.076$, Figure 43E-43F), similar to the clinical pattern seen in EG3 trauma patients. Furthermore, the gene expression profile of EG2 burn patients (gene cluster: Burn_GC2) can be generally mapped to EG3 trauma patients (gene cluster: Trauma_GC3) (Figure 43C). Unlike the adult blunt trauma group, the complete burn dataset contained a large number of young children and infants. The prognostic value of the EG subtypes showed a different pattern between the patients ≥ 18 yr vs. <18 yr (Figure 45). In contrast, the prognostic value of SG subtype designation exhibited a similar pattern between the two groups of ages of burn patients. This interesting finding suggests a lower influence of global epigenetic alterations on the outcomes of very young children after burns.

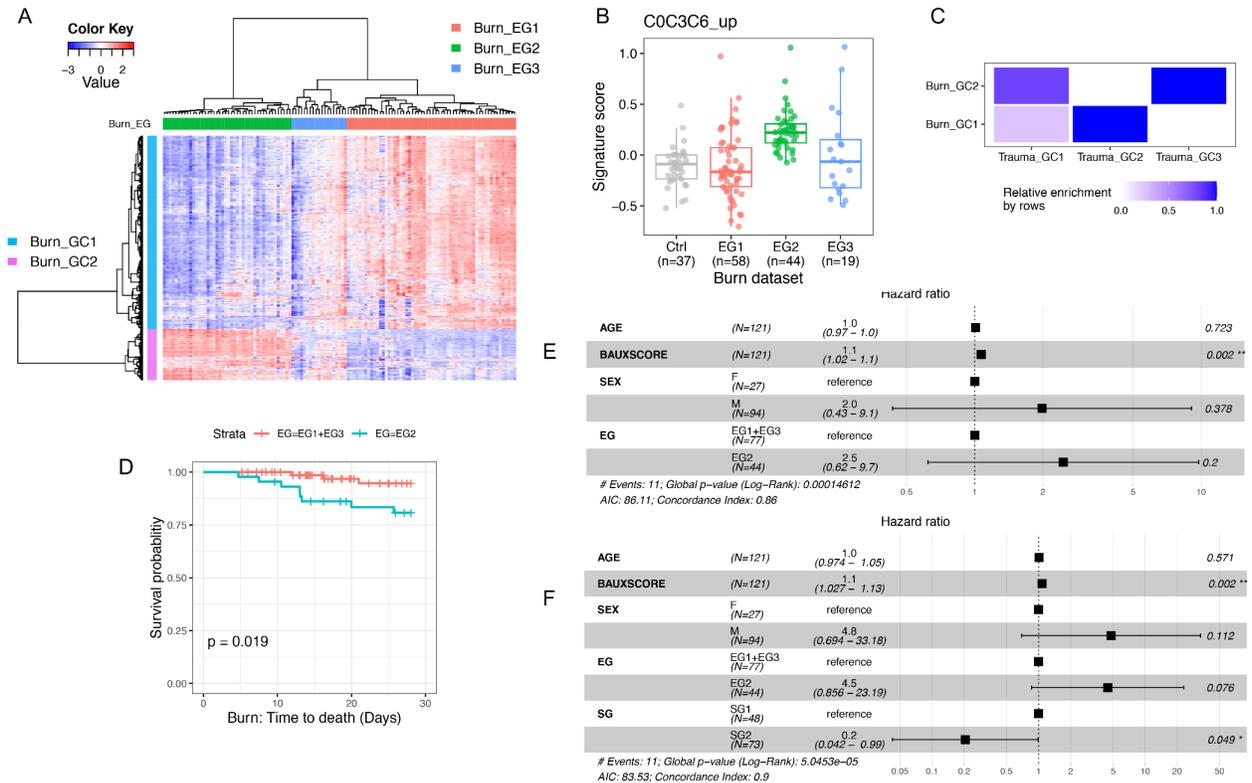


Figure 43 Epigenetic subtypes and prognostic value in burn patients.

(A) Epigenetic subtypes in adult burn patients. The 1st sampled time point for each patient was extracted for the clustering analysis. The top 2000 up and top 2000 down DEGs in trauma EG3 patients that showed standard deviation of scaled expression ≥ 0.5 among the 121 burn samples were used to cluster burn patients. These genes largely fell into two gene clusters Burn_GC1-GC2. (B) C0C3C6_up signature scores among burn EGs as shown in (A) and healthy controls. (C) Mapping of gene clusters derived from Burn patients (Burn_GC#) to those identified in Trauma patients (Trauma_GC#). Fold enrichment was computed between each Burn_GC# and each Trauma_GC# and then scaled between 0 to 1 for each Burn_GC#. (D-F) Survival analysis between Burn_EG2 (highest level of C0C3C6_up signature) versus the other Burn patients. (D) Univariate analysis by Kaplan-Meier estimate. (E) Multivariate analysis using Cox model to adjust potential co-variants. (F) Multivariate analysis using Cox model to further adjust SG subtype designations.

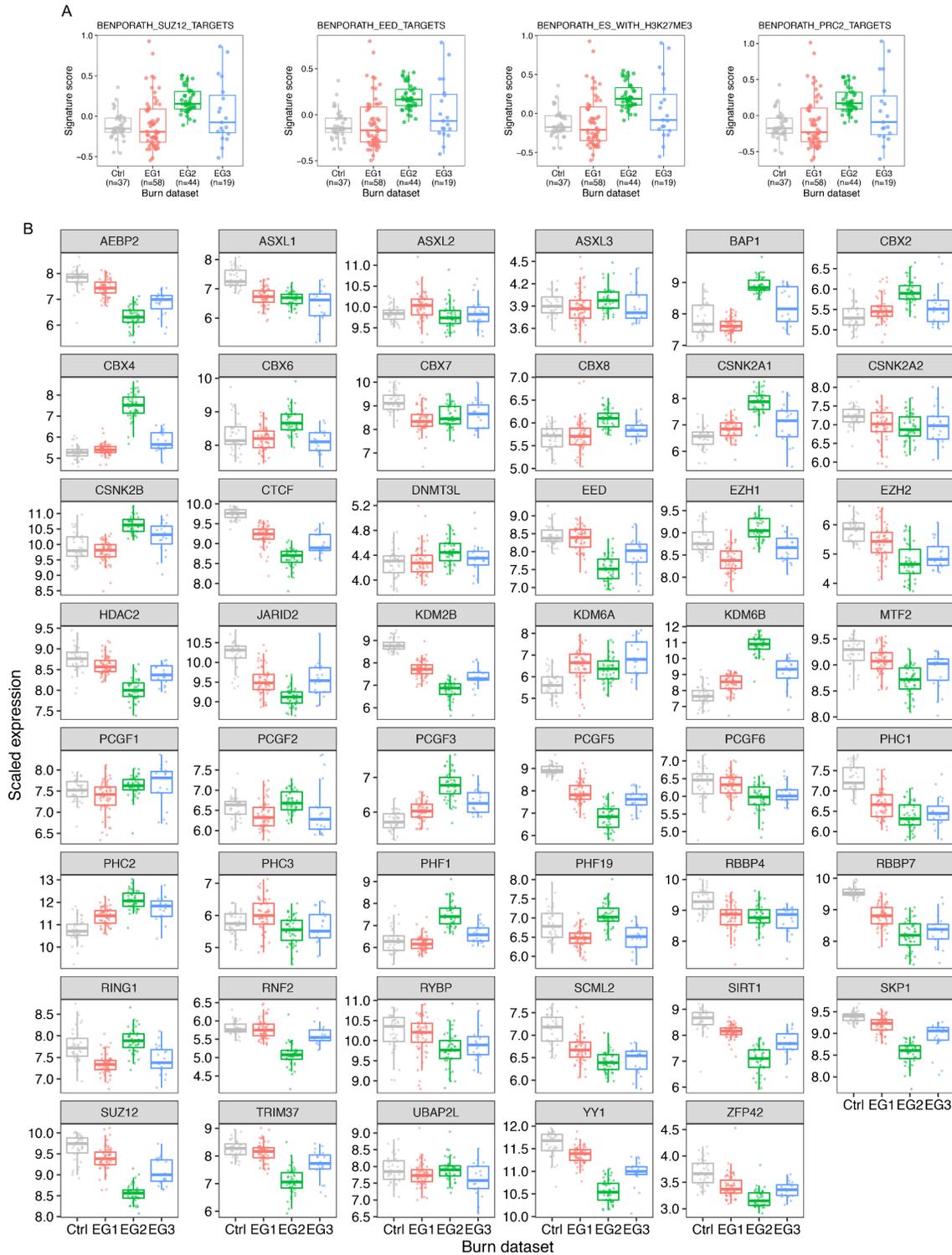


Figure 44 Expression of epigenetic signatures and genes among burn Epigenetic Groups (EGs) along with healthy controls (related to Figure 43).

(A) Signatures of polycomb targets. (B) Critical epigenetic regulators, including GO: PCG_PCG_PROTEIN_COMPLEX along with H3K27me3 demethylase <KDM6A, KDM6B> and CTCF.

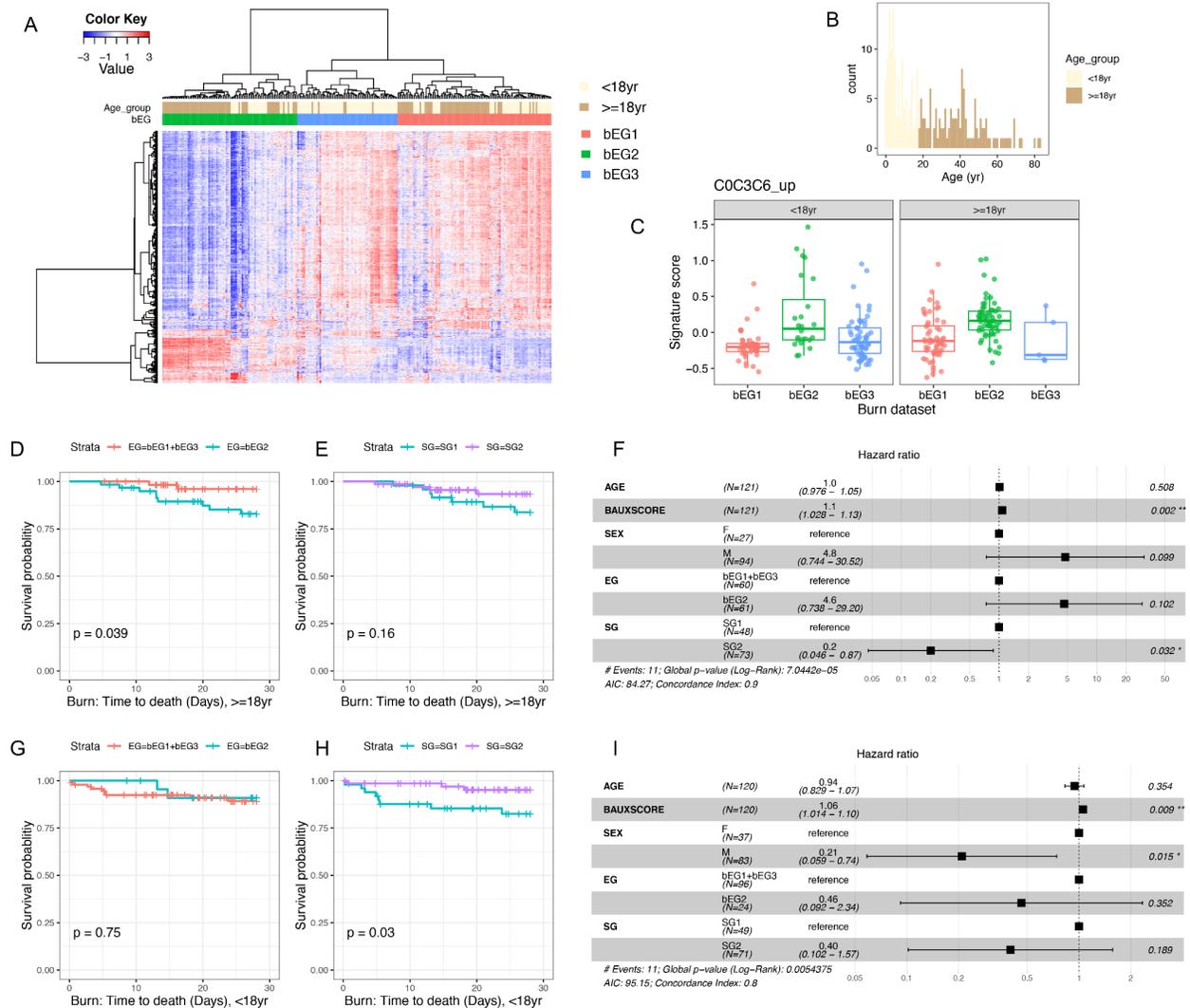


Figure 45 Different prognostic patterns of EGs in burn patients and impact of age on the patterns.

(A) Epigenetic subtypes in all burn patients. The 1st sampled time point for each patient was extracted for the clustering analysis. The top 2000 up and top 2000 down DEGs from trauma EG3 patients that also showed standard deviation of scaled expression ≥ 0.5 among the analyzed burn samples were used to cluster burn patients. To distinguish the epigenetic groups characterized in Figure 43, burn epigenetic groups (bEGs) were only used to specify the epigenetic groups identified by the clustering across all burn patients in this figure. (B) Distribution of ages across analyzed burn patients shown in (A). (C) C0C3C6_up signature scores among burn bEGs as shown in (A). (D-F) Survival analysis in burn patients with age ≥ 18 yr. (D) K-M analysis between bEG2 and bEG1+bEG3. (E) K-M analysis between SG1 and SG2. (F) Multi-variant survival analysis by Cox model. (G-I) Survival analysis in burn patients with age < 18yr.

(G) K-M analysis between bEG2 and bEG1+bEG3. (H) K-M analysis between SG1 and SG2. (I) Multi-variant survival analysis by Cox model.

The Sepsis dataset [15] contained a single sampled time point for each patient within 24h of ICU admission. All but one patient in this dataset was ≥ 18 yr (one patient was 17yr). Therefore, we included all patients from both the discovery and validation cohorts, a total of 479 patients (Figure 46-47). After applying our two-step process, 976 genes were used to cluster these 479 patients. Sepsis patients were generally clustered into three epigenetic groups (Sepsis_EG1-EG3) and the genes used for clustering formed four gene clusters (Sepsis_GC1-GC4) (Figure 46A and Appendix Spreadsheet 11-12). Of the three sepsis epigenetic groups, EG3 sepsis patients had the highest level of C0C3C6_up signature genes (Figure 46B) and a higher level of polycomb targets (Figure 47A). These results indicated that EG3 sepsis patients underwent a higher degree of global epigenetic alterations than the other sepsis EG groups, a pattern similar to EG3 trauma and EG2 burn patients. Consistent with the burn and trauma findings, EG3 sepsis patients also experienced worse survival compared with the other sepsis patients (K-M analysis: $p = 0.012$, Figure 46D). After adjusting for age, sex and SG classification by Cox model, the EG3 sepsis subtype still significantly associated with worse survival ($p = 0.013$, Figure 46E-46F).

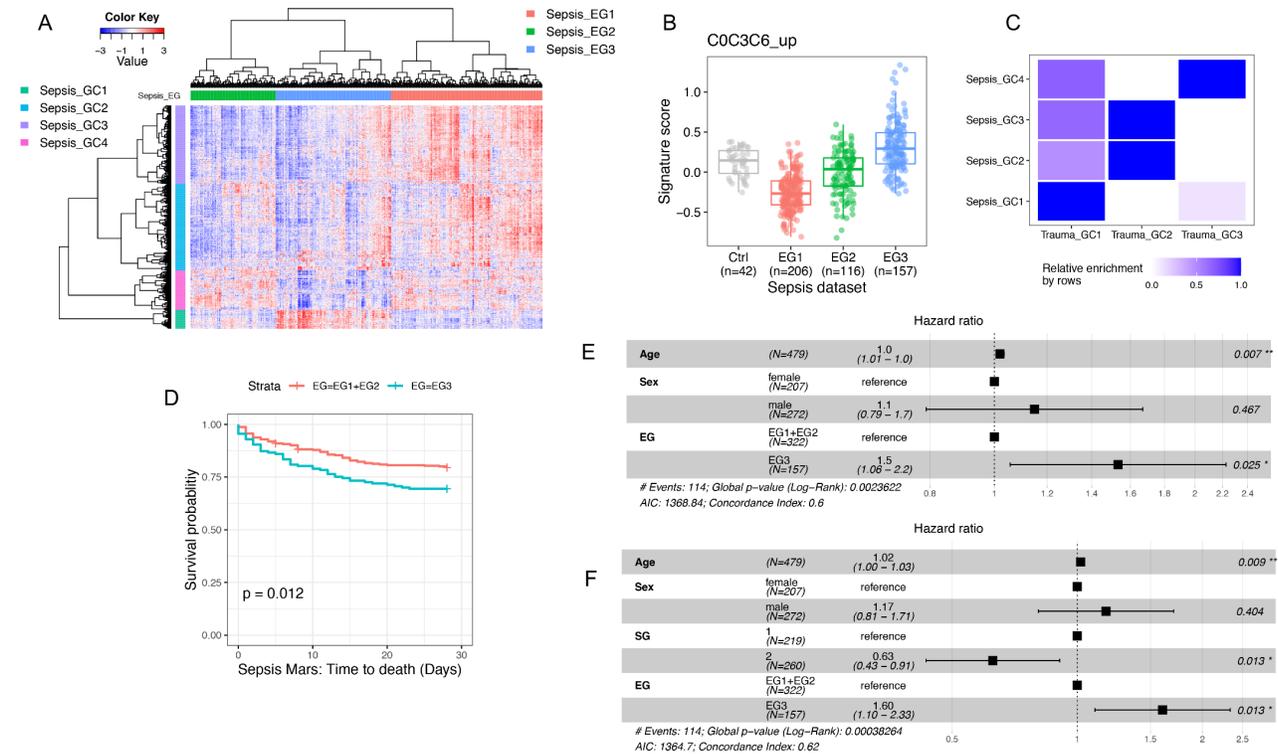


Figure 46 Epigenetic subtypes and prognostic value in sepsis patients.

(A) Epigenetic subtypes in sepsis patients. The top 2000 up and top 2000 down DEGs in trauma EG3 patients that also showed standard deviation of scaled expression ≥ 0.5 among the analyzed sepsis samples were used to cluster sepsis patients. These genes largely fell into four gene clusters Sepsis_GC1-GC4. (B) C0C3C6_up signature scores among Sepsis EGs as shown in (A) and healthy controls. (C) Mapping of gene clusters derived from Sepsis patients (Sepsis_GC#) to those identified in Trauma patients (Trauma_GC#). Fold enrichment was computed between each Sepsis_GC# and each Trauma_GC# and then scaled between 0 to 1 for each Sepsis_GC#. (D-F) Survival analysis between Sepsis_EG3 (highest level of C0C3C6_up signature) versus the other sepsis patients. (D) Univariate analysis by Kaplan-Meier estimate. (E) Multivariate analysis using Cox model to adjust potential co-variants. (F) Multivariate analysis using Cox model to further adjust SG subtype designations.

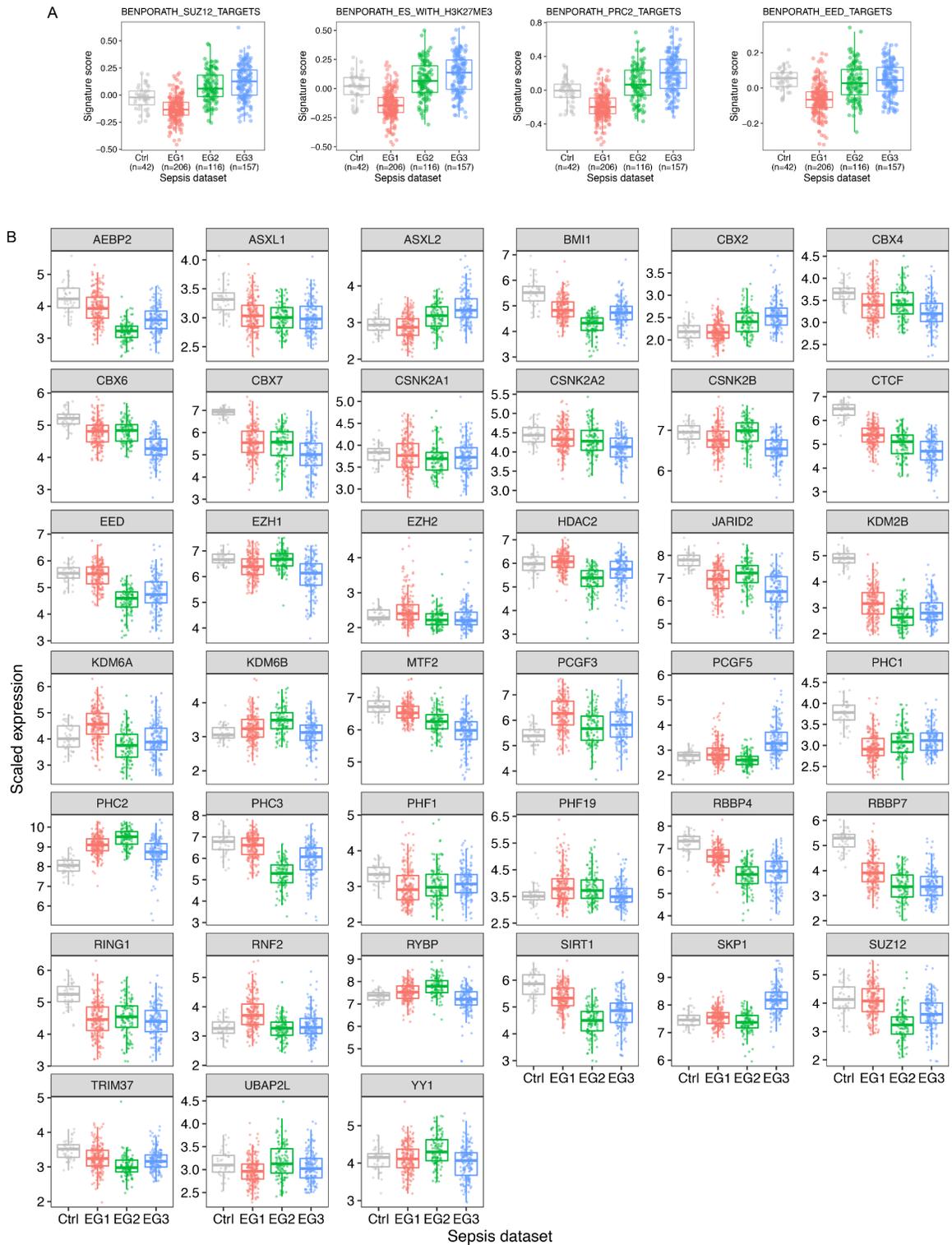


Figure 47 Expression of epigenetic signatures and genes among sepsis Epigenetic Groups (EGs) along with healthy controls (related to Figure 46).

(A) Signatures of polycomb targets. (B) Critical epigenetic regulators, including GO: PCG_PCG_PROTEIN_COMPLEX along with H3K27me3 demethylase <KDM6A, KDM6B> and CTCF.

EG3 sepsis patients highly expressed the gene cluster Sepsis_GC1. Surprisingly, Sepsis_GC1 was more highly enriched in genes found in Trauma_GC1 (relatively low variable DEGs in trauma) compared with Trauma_GC3 (highest expressed in the trauma patients with worse prognosis, Trauma_EG3) (Figure 46C). The expression of PRC2 components also showed a more complex pattern across Sepsis EG subtypes (Figure 47B). Some co-factors (JARID2 and MTFs) were lowest in EG3 sepsis patients, while most core components (EED, RBBP4/7, EZH2) were not. Thus, while the global epigenetic changes in sepsis exhibit characteristics that go beyond the SG classification, the suppression of PRC2 components may be more specific to trauma. The upstream epigenetic regulators that play a dominant role in sepsis may be different and possibly involve more PRC1 components (CBX4, CBX6 and CBX7, Figure 47B), leading to a different downstream co-expression profile. This suggests that a single-cell epigenomic analysis of sepsis is warranted to identify the sepsis-specific epigenetic changes.

Finally, we assessed the gene expression profile of polycomb targets (ES-derived) across sepsis patients. Compared with EG1 and EG2 sepsis patients, EG3 sepsis patients expressed markedly higher level of PcGtargets_SepsisGC1 genes, which were largely associated with non-hematopoietic developmental GO terms. In contrast, the GO terms associated with myeloid/leukocyte/hematopoietic progenitor cell differentiation, were only significantly enriched in PcGtargets_SepsisGC2 genes (Figure 48). Thus, in spite of some differences in the co-expression pattern of epigenetic associated genes between Trauma_EG3 and Sepsis_EG3 (Figure 46-47), the shared epigenetic features between these two groups of patients (both with worse prognosis) include a higher de-repression of non-hematopoietic developmental genes rather than

the hematopoietic developmental genes. The de-repression of developmental genes could potentially contribute dysfunctional states in immune cells.

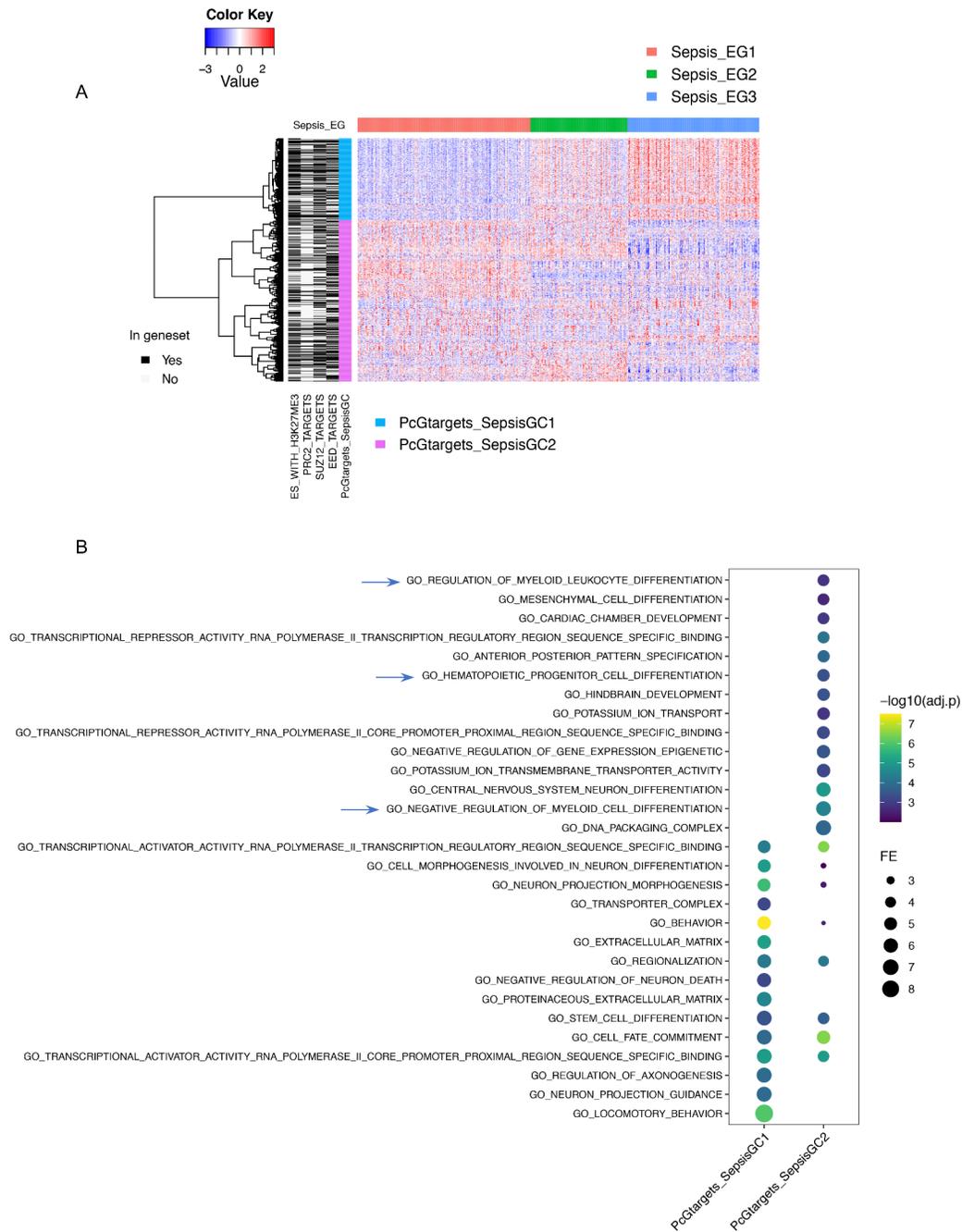


Figure 48 Gene expression profile of polycomb targets across sepsis patients.

(A) Polycomb targets (rows) were collected as the union of 4 gene sets: BENPORATH_ES_WITH_H3K27ME3, BENPORATH_SUZ12_TARGETS, BENPORATH_PRC2_TARGETS, BENPORATH_EED_TARGETS. Patients (columns) were sorted by EG subtypes identified in (Figure 46A). Polycomb targets were clustered into 2 clusters, PcTargets_SepsisGC1-GC2. (B) GO enrichment for the 2 gene clusters identified in (A). Top 15 enriched GO terms

(sorted by fold enrichment (FE) decreasingly) for each gene cluster were shown (threshold for significance: FDR < 0.05 & the number of overlapped genes ≥ 10 & FE >2). Hematopoietic relevant GO terms were pointed out by blue arrows.

4.4 Discussion

This study characterized the open chromatin patterns in PBMCs using scATAC-seq in humans subjected to an acute systemic insult that resulted in either a rapid recovery or a transition into a state of persistent critical illness. Using standard methods combined with HSMM 15-state epigenome model [64] within CD14⁺ monocytes, we show that known changes in transcriptomic patterns associated with the immune dysfunction of critical illness (excessive up-regulation of proinflammatory genes and suppression of MHC and interferon signaling) largely relate to the changes of active chromatin states (TssA/TssAFlnk/Enh) of reference epigenome. Unexpectedly, we also found that failure to resolve critical illness was associated with a higher degree of global epigenetic alterations, that include the de-repression of polycomb targets (associated with many developmental genes) not typically expressed in hematopoietic lineages. PcG is a well-characterized system essential for stable gene silencing [79]. We provide the first evidence that this system is impaired very rapidly (4-24h), along with the release of normally repressed targets during extreme stress caused by acute injury and other etiologies, in both myeloid and lymphoid lineages, and that this associates with worse clinical outcome.

Trauma and other etiologies leading to acute critical illness can be associated with immune dysfunction that manifests as early hyper-inflammation followed by a sustained immune-paralysis [1]. From our previous studies [18] and the studies of others [58, 59] using single-cell RNA

sequencing of PBMC from critically ill patients, a common finding across trauma and sepsis is the up-regulation of proinflammatory genes and suppression of genes associated with MHC and interferon signaling in myeloid cells. In the current study we go beyond the information captured in single-cell transcriptomics to identify epigenetic group (EG) subtypes in whole-blood leukocyte across three etiologies of critical illness. Here, we found that at the bulk level the signatures we used to define SG subtypes were not among the top genes highly correlated the pathologic EG subtype, the C0C3C6_up gene signature. This further supports the notion the factors that drive the pathologic SG and EG gene signatures are part of different processes. Nonetheless, patients that were defined as SG1 and had EG subtypes with the highest representation of the C0C3C6_up signature were consistently and independently associated with the worse prognosis across three etiologies of critical illness. This suggests that both processes contribute to the dysfunctional immune responses that associated with critical illness.

Intriguingly, a high dose of Lipopolysaccharide (LPS) exposure in multiple mouse strains, resulting in >75% morbidity within 48h, has been shown to result in alterations in expression of hedgehog signaling components and the developmental regulators Pax4, Hoxa4 and Cdx2 [80]. In keeping with these observations, macrophages deficient in one of the hedgehog receptors (Ptch1^{+/-}) were resistant to LPS induced inflammatory cytokine production. These findings indicate that some developmental regulatory genes also modulate inflammatory responses. We speculate that the molecular chaos of high and sustained stress signaling, along with other inputs (e.g., hypoxia, reactive oxygen species) drive rapid and global epigenetic alterations which result in mis-expression of developmental regulators that impact inflammatory cytokine signaling. This could, in turn, contribute to the wide-spread immune dysfunction of critical illness.

There were some paradoxes based on the current sepsis transcriptomic classification in circulation. The Mars [15] and SRS [14] sepsis classifications were based on gene array studies of whole-blood leukocytes from sepsis patients obtained at a single time point. The Mars1 and SRS1 subtypes have the worse outcomes in both of the studies. However, Mars1 was found to be largely mapped to SRS2 rather than SRS1 [15]. We had demonstrated in our previous paper (also discussion in Chapter 2) that SG1 and SRS1 shared common transcriptomic and outcome patterns, involving the higher degree of both up-regulated inflammation and down-regulated MHC II and IFN signaling pathways in the myeloid compartment, but this was not the case for Mars1, which also had worse outcomes [18]. Mars1 patients exhibited a pronounced decrease in both innate and adaptive immune response but only displayed an increase in metabolic pathways (mainly heme metabolism). The information above is far from sufficient for us to understand how immune response are dysfunctional in Mars1 patients. These led us to examine the relationship between Mars endotypes and the epigenetic subtypes we defined in our study. It turned out that the Mars1 endotype was significantly enriched in EG3 sepsis patients (Figure 49).

Thus, Mars study captured more features associated with global epigenetic alterations, while SRS study captured more local features (eg. inflammatory genes). Failure to distinguish the two different sources of features from each other led to the paradoxical but both reasonable conclusions of the two publications. This provides further support for the separation of biologic process identified by transcriptomic analysis alone and epigenomic analysis of the same conditions. This could also explain why anti-inflammatory therapies had marginal effects in trials of human sepsis [81]. Using single-cell sequencing data for feature selection helped us to better link patient heterogeneity within patient populations to specific cellular changes, and provided a more straightforward interpretation for the identified subtypes.

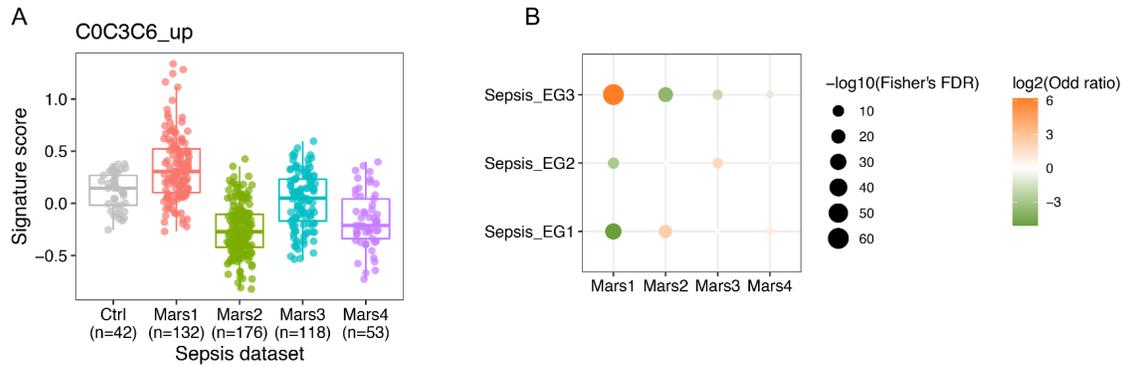


Figure 49 Relationship between Mars and EG classifications.

(A) C0C3C6_up signature scores among Mars endotypes and healthy controls. (B) Fisher's Exact test between Sepsis_EGs and Mars endotypes. p values were corrected by Benjamini-Hochberg method for multiple testing.

We also provide an analytic workflow incorporating the findings scATAC-seq data and bulk transcriptomic data to explore the global epigenetic heterogeneity in patient populations. Most importantly, we provide a strategy to directly extract the signature associated with global epigenetic changes in specific cell types using scATAC-seq data. In the future, this workflow can be used to establish a more burn or sepsis-specific epigenetic classification by directly scATAC sequencing burn or sepsis patients. The workflow can also be applied to other diseases suspected to undergo global epigenetic changes, such as aging [82].

While our studies using epigenetic analysis provides new insights into the processes involved gene regulation in immune cells during acute critical illness, we also provide some clarity on the path forward to better understand the mechanisms of immune dysfunction in this challenging area of research. Specifically, we need to understand where (cell compartments and cell types) and under what forms of cell stress and activation the de-repression takes place. What are the most proximal signaling events involved in the epigenomic reprogramming? What is the range of functional consequences stemming from the de-repression of polycomb targets in immune

cells and how long do these changes last? Finally, are these changes restricted to critical illness or will these changes be part of other more chronic immune and inflammatory processes?

5.0 Conclusions and Future Work

This collection of studies was undertaken to deconvolute the immune response to systemic injury. We began with scRNA-seq followed by scATAC-seq to characterize the transcriptomic and epigenomic changes in human PBMCs and extended this work to mouse PBMCs and BMMCs as well.

Using scRNA-seq, we found that myeloid cells (esp. CD14⁺ monocytes) underwent dramatic changes, including up-regulation of inflammation and suppression of specific immune pathways (genes associated with MHC II and IFN signaling), along with the changes of several critical transcriptional factors. These changes were largely consistent between humans and mice, and we were able to in silico track these changes back to myeloid progenitors in a mouse T/HS model. In addition, these single-cell derived features were used to define two transcriptional subtypes in the whole leukocyte transcriptome of trauma patients, which we called SG subtypes. The SG1 trauma patients, which exhibited higher levels of inflammation and lower levels of MHC II and IFN signaling pathways, displayed a slower recovery.

Using scATAC-seq, we were first able to recapitulate the transcriptomic changes we characterized using scRNA-seq (local features). More importantly, we revealed that trauma induced global epigenetic alterations across the major immune cell types that pointed to global increased accessibility in chromatin structure and de-repression of polycomb targets, which should be normally suppressed in hematopoietic lineages. The features associated with the global epigenetic alterations were used to define epigenetic subgroups (EG subtypes) in the whole blood leukocyte transcriptome of trauma patients. The EG3 trauma patients, who displayed a higher degree of global epigenetic alterations than other patients, suffered worse outcomes. To highlight,

the gene signatures associated with global epigenetic alterations were distinct from the transcriptomic features used to define SG subtypes (Figure 36C-36E). Furthermore, both SG and EG subtypes were significantly associated with differential prognosis, and were independent of each other. Thus, we provided two biologically driven trauma patient classifications that associated with distinct cellular mechanisms (Table 8, Figure 50).

Table 8 The comparison between SG and EG classification

	SG subtypes	EG subtypes
Dataset where the features derived from	scRNA-seq	scATAC-seq
Extracted signature(s)	Six signatures (C1-C6), Table 4	C0C3C6_up signature (83 peaks/ 65 genes), Table 7
Biological interpretation	Up-regulation of inflammation and suppression of MHCII and IFN signaling	Global epigenetic alterations
Involved immune cell types	Myeloid, mainly CD14+ monocytes	Myeloid + lymphoid
The subtype associated with worse prognosis	SG1 (higher inflammation & lower MHC II and IFN signaling)	The EG subtype with higher levels of global epigenetic alterations (Trauma: EG3)
BM-derived	Yes	Uncertain, considering that lymphocytes are also involved.

More broadly, the SG and EG classifications were largely recapitulated in the patients with critical illness spanning different etiologies, including burn and sepsis, and demonstrated independent prognostic values in these patient groups as well. Thus, our findings reveal immune response patterns that are shared across several causes of acute critical illness. Trauma, sepsis and burns represent states of extreme stress that are often lethal even in the setting of advanced medical care. Hence, we would expect these common causes of major systemic stress in humans to manifest aberrant response patterns at the molecular level. We summarize these patterns as follows:

(1) Transcriptomic features: At the level of an excessive adaptive response, SG subtypes reveal the adaptive and mal-adaptive relationships with the safe (early resolution) and excessive (slow resolution) ranges detected in the deviation in gene subsets in myeloid cells. These changes in transcription are initiated in the BM and seen in the periphery as the extent of deviation from steady state --- SG1 (excessive/ maladaptive) and SG2 (non-excessive/ adaptive).

(2) Epigenomic features: In parallel, the extreme stress states also result in global alterations in the chromatin accessibility driven by processes mostly independent from the adaptive changes as revealed by the EG patterns.

When the wider global epigenetic alterations (Trauma_EG3, Burn_EG2, Sepsis_EG3) occur on top of the changes seen in SG1, immune dysfunction may become severe and sustained. The EG changes are more global and could explain the dysfunction seen across multiple cell types.

Admittedly, both single-cell studies (scRNA or scATAC) were based on a small number of patients. However, we were able to validate the single-cell derived signatures in whole blood leukocyte transcriptomes including hundreds of patients across three etiologies of critical illness, demonstrating that our conclusions are valid and reproducible. More importantly, the large-scale datasets with sufficient number of patients allowed us to characterize patient heterogeneity and provided sufficient statistical power to explore the association with clinical outcomes.

To reveal the epigenetic heterogeneity, we used global epigenetic associated genes to explore the transcriptomic data. It will be interesting to directly explore patient heterogeneity using large-scale epigenetic datasets, such as CpG methylation or H3K27me3 ChIP-seq. Our scATAC-seq study highlights a new and important feature when analyzing large-scale epigenetic datasets in the future; the importance of keeping in mind global vs. local epigenetic modifications and

analyzing these modifications separately. This approach is likely to give us clearer overview and better interpretation of the regulatory mechanisms immune mediated diseases.

In addition, more work will be needed to specifically align the transcriptomic and epigenomic findings with cell functional changes. Whether or how these two mechanisms interact with each other also needs to be elucidated. This dissertation provides a roadmap from single-cell transcriptomics and single-cell epigenomics to two distinct patient classification systems related to patient outcomes. In the short term, rapid identification of EG and SG subtypes early after admission could prognostication for outcomes. In the long term, the findings establish the foundations for promoting precision medicine in critical illness, by identification of patient subsets most likely to respond to targeted therapies (eg. anti-IL-6) and/or through the development of therapies aimed at preventing the global epigenetic alterations in appropriate subset of patients. Thus, this work has the potential to open new pre-clinical research directions and more sophisticated design of clinical trials in the future.

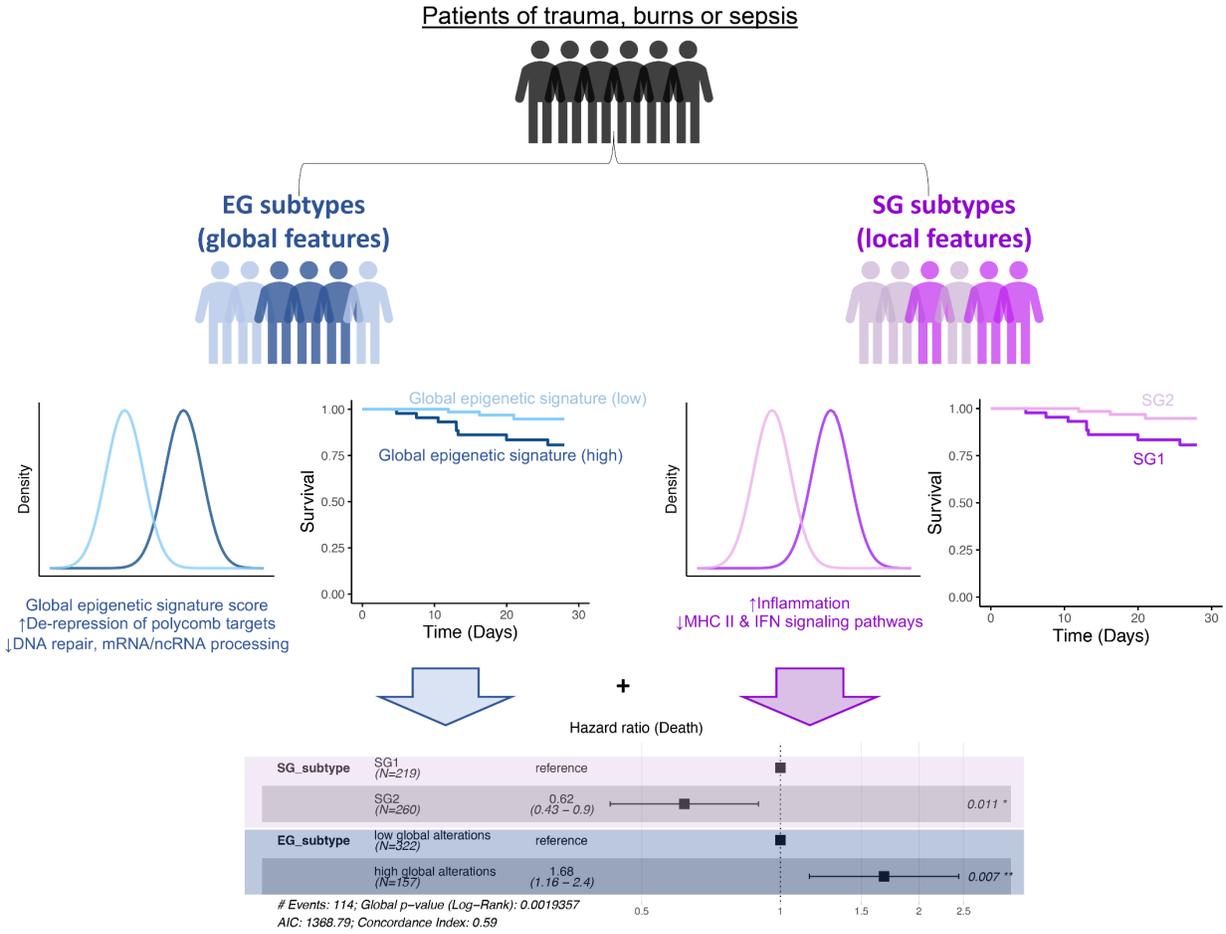


Figure 50 New diagram of patient classifications in critical illness: Two different methods of classifications with independent prognostic value beyond each other.

Appendix Spreadsheets

Appendix Spreadsheet 1: Subgroup analysis for outcome parameters based on SG status at 12h.

<https://d-scholarship.pitt.edu/41614/1/STable1.xlsx>

Appendix Spreadsheet 2: Subgroup analysis for outcome parameters based on SG status at 1d.

<https://d-scholarship.pitt.edu/41614/2/STable2.xlsx>

Appendix Spreadsheet 3: Subgroup analysis for associated factors based on SG status at 12h.

<https://d-scholarship.pitt.edu/41614/3/STable3.xlsx>

Appendix Spreadsheet 4: Subgroup analysis for associated factors based on SG status at 1d.

<https://d-scholarship.pitt.edu/41614/4/STable4.xlsx>

Appendix Spreadsheet 5: Association between SG status at 1d and known prognostic factors.

<https://d-scholarship.pitt.edu/41614/5/STable5.xlsx>

Appendix Spreadsheet 6: C0C3C6_up signature information and results gene set over-representation. <https://d-scholarship.pitt.edu/41614/6/STable6.xlsx>

Appendix Spreadsheet 7: C0C3C6_up highly correlated genes and GSEA results. <https://d-scholarship.pitt.edu/41614/7/STable7.xlsx>

Appendix Spreadsheet 8: Trauma patient classifications. <https://d-scholarship.pitt.edu/41614/8/STable8.xlsx>

Appendix Spreadsheet 9: Demographic comparison between EG3 trauma patients versus EG1+EG2. <https://d-scholarship.pitt.edu/41614/9/STable9.xlsx>

Appendix Spreadsheet 10: DEGs between Trauma_EG3 vs EG1+EG2 and GSEA results. <https://d-scholarship.pitt.edu/41614/10/STable10.xlsx>

Appendix Spreadsheet 11: Burn and sepsis patient classifications. <https://d-scholarship.pitt.edu/41614/11/STable11.xlsx>

Appendix Spreadsheet 12: Gene clusters identified from trauma, burn or sepsis patients. <https://d-scholarship.pitt.edu/41614/12/STable12.xlsx>

Bibliography

1. Lord, J.M., et al., *The systemic immune response to trauma: an overview of pathophysiology and treatment*. Lancet, 2014. **384**(9952): p. 1455-65.
2. Huber-Lang, M., J.D. Lambris, and P.A. Ward, *Innate immune responses to trauma*. Nat Immunol, 2018. **19**(4): p. 327-341.
3. Billiar, T.R. and Y. Vodovotz, *Time for trauma immunology*. PLoS Med, 2017. **14**(7): p. e1002342.
4. Gentile, L.F., et al., *Persistent inflammation and immunosuppression: a common syndrome and new horizon for surgical intensive care*. J Trauma Acute Care Surg, 2012. **72**(6): p. 1491-501.
5. Cole, E., et al., *Multiple organ dysfunction after trauma*. Br J Surg, 2020. **107**(4): p. 402-412.
6. Xiao, W., et al., *A genomic storm in critically injured humans*. J Exp Med, 2011. **208**(13): p. 2581-90.
7. Cabrera, C.P., et al., *Signatures of inflammation and impending multiple organ dysfunction in the hyperacute phase of trauma: A prospective cohort study*. PLoS Med, 2017. **14**(7): p. e1002352.
8. Liu, D., et al., *Unsupervised Clustering Analysis Based on MODS Severity Identifies Four Distinct Organ Dysfunction Patterns in Severely Injured Blunt Trauma Patients*. Front Med (Lausanne), 2020. **7**: p. 46.
9. Schimunek, L., et al., *Computational Derivation of Core, Dynamic Human Blunt Trauma Inflammatory Endotypes*. Front Immunol, 2020. **11**: p. 589304.
10. Seshadri, A., et al., *Phenotyping the Immune Response to Trauma: A Multiparametric Systems Immunology Approach*. Crit Care Med, 2017. **45**(9): p. 1523-1530.
11. Hotchkiss, R.S., G. Monneret, and D. Payen, *Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy*. Nat Rev Immunol, 2013. **13**(12): p. 862-74.
12. Mann, E.A., et al., *Comparison of mortality associated with sepsis in the burn, trauma, and general intensive care unit patient: a systematic review of the literature*. Shock, 2012. **37**(1): p. 4-16.

13. Takao, K. and T. Miyakawa, *Genomic responses in mouse models greatly mimic human inflammatory diseases*. Proc Natl Acad Sci U S A, 2015. **112**(4): p. 1167-72.
14. Davenport, E.E., et al., *Genomic landscape of the individual host response and outcomes in sepsis: a prospective cohort study*. Lancet Respir Med, 2016. **4**(4): p. 259-71.
15. Scicluna, B.P., et al., *Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study*. Lancet Respir Med, 2017. **5**(10): p. 816-826.
16. Marshall, J.C., *Why have clinical trials in sepsis failed?* Trends Mol Med, 2014. **20**(4): p. 195-203.
17. Seok, J., et al., *Genomic responses in mouse models poorly mimic human inflammatory diseases*. Proc Natl Acad Sci U S A, 2013. **110**(9): p. 3507-12.
18. Chen, T., et al., *A road map from single-cell transcriptome to patient classification for the immune response to trauma*. JCI Insight, 2021. **6**(2).
19. Darwiche, S.S., et al., *Pseudofracture: an acute peripheral tissue trauma model*. J Vis Exp, 2011(50).
20. Kohut, L.K., et al., *Fixed volume or fixed pressure: a murine model of hemorrhagic shock*. J Vis Exp, 2011(52).
21. Kochanek, K.D., et al., *Deaths: Final Data for 2017*. National Vital Statistics Reports, 2019. **68**(9).
22. Zheng, G.X., et al., *Massively parallel digital transcriptional profiling of single cells*. Nat Commun, 2017. **8**: p. 14049.
23. Butler, A., et al., *Integrating single-cell transcriptomic data across different conditions, technologies, and species*. Nat Biotechnol, 2018. **36**(5): p. 411-420.
24. Stuart, T., et al., *Comprehensive Integration of Single-Cell Data*. Cell, 2019. **177**(7): p. 1888-1902 e21.
25. Amir el, A.D., et al., *viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia*. Nat Biotechnol, 2013. **31**(6): p. 545-52.
26. Qiu, X., et al., *Reversed graph embedding resolves complex single-cell trajectories*. Nat Methods, 2017. **14**(10): p. 979-982.
27. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering*. Nat Methods, 2017. **14**(11): p. 1083-1086.

28. Singer, M., et al., *A Distinct Gene Module for Dysfunction Uncoupled from Activation in Tumor-Infiltrating T Cells*. Cell, 2016. **166**(6): p. 1500-1511 e9.
29. Gaublot, J.T., et al., *Single-Cell Genomics Unveils Critical Regulators of Th17 Cell Pathogenicity*. Cell, 2015. **163**(6): p. 1400-12.
30. La Manno, G., et al., *RNA velocity of single cells*. Nature, 2018. **560**(7719): p. 494-498.
31. Subramanian, A., 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): p. 15545-50.
32. Murtagh, F. and P.J.J.o.C. Legendre, *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* 2014. **31**(3): p. 274-295.
33. Mildner, A., et al., *Genomic Characterization of Murine Monocytes Reveals C/EBPbeta Transcription Factor Dependence of Ly6C(-) Cells*. Immunity, 2017. **46**(5): p. 849-862 e7.
34. Marigo, I., et al., *Tumor-induced tolerance and immune suppression depend on the C/EBPbeta transcription factor*. Immunity, 2010. **32**(6): p. 790-802.
35. Guo, X., et al., *Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing*. Nat Med, 2018. **24**(7): p. 978-985.
36. van der Maaten, L. and G. Hinton, *Visualizing Data using t-SNE*. Journal of Machine Learning Research, 2008. **9**: p. 2579-2605.
37. Gill, R., et al., *Systemic inflammation and liver injury following hemorrhagic shock and peripheral tissue trauma involve functional TLR9 signaling on bone marrow-derived cells and parenchymal cells*. Shock, 2011. **35**(2): p. 164-70.
38. Geissmann, F., S. Jung, and D.R. Littman, *Blood monocytes consist of two principal subsets with distinct migratory properties*. Immunity, 2003. **19**(1): p. 71-82.
39. Olsson, A., et al., *Single-cell analysis of mixed-lineage states leading to a binary cell fate choice*. Nature, 2016. **537**(7622): p. 698-702.
40. Deneault, E., et al., *A functional screen to identify novel effectors of hematopoietic stem cell activity*. Cell, 2009. **137**(2): p. 369-79.
41. McInnes, L., J. Healy, and J. Melville, *UMAP: uniform manifold approximation and projection for dimension reduction*. [preprint] arXiv:1802.03426, 2018.
42. Liberzon, A., et al., *Molecular signatures database (MSigDB) 3.0*. Bioinformatics, 2011. **27**(12): p. 1739-40.

43. Hirai, H., et al., *C/EBPbeta is required for 'emergency' granulopoiesis*. Nat Immunol, 2006. **7**(7): p. 732-9.
44. Hirai, H., et al., *Non-steady-state hematopoiesis regulated by the C/EBPbeta transcription factor*. Cancer Sci, 2015. **106**(7): p. 797-802.
45. Kurotaki, D., et al., *Essential role of the IRF8-KLF4 transcription factor cascade in murine monocyte differentiation*. Blood, 2013. **121**(10): p. 1839-49.
46. Kleiveland, C.R., *Peripheral Blood Mononuclear Cells*, in *The Impact of Food Bioactives on Health: in vitro and ex vivo models*, K. Verhoeckx, et al., Editors. 2015, Springer International Publishing: Cham. p. 161-167.
47. Guilliams, M., A. Mildner, and S. Yona, *Developmental and Functional Heterogeneity of Monocytes*. Immunity, 2018. **49**(4): p. 595-613.
48. Honda, K., A. Takaoka, and T. Taniguchi, *Type I interferon [corrected] gene induction by the interferon regulatory factor family of transcription factors*. Immunity, 2006. **25**(3): p. 349-60.
49. Peretti, M., et al., *Expression of the three human major histocompatibility complex class II isotypes exhibits a differential dependence on the transcription factor RFXAP*. Mol Cell Biol, 2001. **21**(17): p. 5699-709.
50. Champion, H.R., *Trauma scoring*. Scand J Surg, 2002. **91**(1): p. 12-22.
51. Hoffmann, M., et al., *Pupil evaluation in addition to Glasgow Coma Scale components in prediction of traumatic brain injury and mortality*. Br J Surg, 2012. **99** Suppl 1: p. 122-30.
52. Odom, S.R., et al., *Lactate clearance as a predictor of mortality in trauma patients*. J Trauma Acute Care Surg, 2013. **74**(4): p. 999-1004.
53. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles*. Nat Methods, 2015. **12**(5): p. 453-7.
54. Calvano, S.E., et al., *A network-based analysis of systemic inflammation in humans*. Nature, 2005. **437**(7061): p. 1032-7.
55. Amit, I., et al., *Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses*. Science, 2009. **326**(5950): p. 257-63.
56. John, S.P., et al., *IFIT1 Exerts Opposing Regulatory Effects on the Inflammatory and Interferon Gene Programs in LPS-Activated Human Macrophages*. Cell Rep, 2018. **25**(1): p. 95-106 e6.

57. Yanez, A., et al., *Granulocyte-Monocyte Progenitors and Monocyte-Dendritic Cell Progenitors Independently Produce Functionally Distinct Monocytes*. *Immunity*, 2017. **47**(5): p. 890-902 e4.
58. Reyes, M., et al., *An immune-cell signature of bacterial sepsis*. *Nat Med*, 2020.
59. Wilk, A.J., et al., *A single-cell atlas of the peripheral immune response in patients with severe COVID-19*. *Nat Med*, 2020.
60. Seymour, C.W., et al., *Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis*. *JAMA*, 2019. **321**(20): p. 2003-2017.
61. Blow, O., et al., *The golden hour and the silver day: detection and correction of occult hypoperfusion within 24 hours improves outcome from major trauma*. *J Trauma*, 1999. **47**(5): p. 964-9.
62. Claridge, J.A., et al., *Persistent occult hypoperfusion is associated with a significant increase in infection rate and mortality in major trauma patients*. *J Trauma*, 2000. **48**(1): p. 8-14; discussion 14-5.
63. Granja, J.M., et al., *Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia*. *Nat Biotechnol*, 2019. **37**(12): p. 1458-1465.
64. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes*. *Nature*, 2015. **518**(7539): p. 317-30.
65. Stuart, T., et al., *Multimodal single-cell chromatin analysis with Signac*. 2020: p. 2020.11.09.373613.
66. Srivatsan, S.R., et al., *Massively multiplex chemical transcriptomics at single-cell resolution*. *Science*, 2020. **367**(6473): p. 45-51.
67. Schug, J., et al., *Promoter features related to tissue specificity as measured by Shannon entropy*. *Genome Biol*, 2005. **6**(4): p. R33.
68. Murtagh, F. and P. Legendre, *Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?* *Journal of Classification*, 2014. **31**(3): p. 274-295.
69. Marshall, J.C., et al., *Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome*. *Crit Care Med*, 1995. **23**(10): p. 1638-52.
70. Wong, K.L., et al., *Gene expression profiling reveals the defining features of the classical, intermediate, and nonclassical human monocyte subsets*. *Blood*, 2011. **118**(5): p. e16-31.

71. Kohyama, M., et al., *Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis*. Nature, 2009. **457**(7227): p. 318-21.
72. Coccia, E.M., et al., *STAT1 activation during monocyte to macrophage maturation: role of adhesion molecules*. Int Immunol, 1999. **11**(7): p. 1075-83.
73. Wagner, E.F., *Bone development and inflammatory disease is regulated by AP-1 (Fos/Jun)*. Ann Rheum Dis, 2010. **69 Suppl 1**: p. i86-88.
74. Mark, M., F.M. Rijli, and P. Chambon, *Homeobox genes in embryogenesis and pathogenesis*. Pediatr Res, 1997. **42**(4): p. 421-9.
75. Voigt, P., W.W. Tee, and D. Reinberg, *A double take on bivalent promoters*. Genes Dev, 2013. **27**(12): p. 1318-38.
76. Pruitt, K., et al., *Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation*. PLoS Genet, 2006. **2**(3): p. e40.
77. Oberdoerffer, P., et al., *SIRT1 redistribution on chromatin promotes genomic stability but alters gene expression during aging*. Cell, 2008. **135**(5): p. 907-18.
78. Lee, T.I., et al., *Control of developmental regulators by Polycomb in human embryonic stem cells*. Cell, 2006. **125**(2): p. 301-13.
79. Lanzuolo, C. and V. Orlando, *Memories from the polycomb group proteins*. Annu Rev Genet, 2012. **46**: p. 561-89.
80. Yang, I.V., et al., *Novel regulators of the systemic response to lipopolysaccharide*. Am J Respir Cell Mol Biol, 2011. **45**(2): p. 393-402.
81. Freeman, B.D. and C. Natanson, *Anti-inflammatory therapies in sepsis and septic shock*. Expert Opin Investig Drugs, 2000. **9**(7): p. 1651-63.
82. Sen, P., et al., *Epigenetic Mechanisms of Longevity and Aging*. Cell, 2016. **166**(4): p. 822-839.