Discovering and validating the correlation between morphology and gene expression in bacterial sepsis CD8+ T cells using deep learning approach

Jong Hyun Kim

Department of Biomedical Systems Informatics
The Graduate School, Yonsei University

Discovering and validating the correlation between morphology and gene expression in bacterial sepsis CD8+ T cells using deep learning approach

Directed by Professor Yu Rang Park

The Master's Thesis submitted to the Department of Biomedical Systems Informatics, and the Graduate School of Yonsei University in partial fulfillment of the requirements for the degree of Master of Science

Jong Hyun Kim

December 2023

This certifies that the Master's Thesis of
Jong Hyun Kim is approved.

_____
Thesis Supervisor: Yu Rang Park

_____
Thesis Committee Member I: Dukyong Yoon

_____
Thesis Committee Member II: Hyun-Seok Min

The Graduate School
Yonsei University

December 2023

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Yu Rang Park, for her consistent support and guidance during this project. I am grateful to Professor Kyung Soo Chung and MinDong Sung in the Department of Pulmonary Medicine, for their hep with specimen collection and processing. I would also like to express my appreciation to team members Sooyoung Jang, Bo Kyu Choi, Chan Young Ko, Taewon Kim, and Min Kyoon Yoo, as well as the rest of the DHLab members.

Finally, I would like to thank my family for their continuous support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Discovering and validating the correlation between morphology and gene expression in bacterial sepsis CD8+ T cells using deep learning approach

Jong Hyun Kim

Department of Biomedical Systems Informatics
The Graduate School, Yonsei University
Directed by Professor by Yu Rang Park

The complex interplay between the morphology and the gene expression of T cells plays a key role in the immune response. However, a comprehensive understanding of these interactions remains elusive, particularly in the context of dynamic immune-related diseases like sepsis. Here, we investigate the association between T cell gene expression profiles through single-cell RNA sequencing, and three-dimensional cellular images obtained through holotomography. Sepsis is a dynamic immune-related disease with notable changes in CD8 T-cell morphology. This study examined the relationships in CD8 T-cells within a longitudinal cohort of sepsis patients using deep learning models to elucidate underlying patterns and relationships. We identified genes specific to morphology that exhibited a high association with the longitudinal morphological changes in CD8 T-cells. Additionally, these genes bear biological significance in relation to cellular structures, such as chromatin organization. The clinical relevance of the morphology-specific genes was validated by analyzing open sepsis and coronavirus

2019 (COVID-19) single-cell RNA sequencing datasets. The genes that consistently reflect disease severity were identified, thereby enabling the filtering of genes associated with disease severity. This approach deepens our understanding of the interrelationship between gene expression and cellular morphology and underscores the potential of cellular morphology as a target to advance new diagnostic and prognostic strategies in various immune-related diseases.

Keywords: deep learning, single cell RNA sequencing, 3D cell images, sepsis

Discovering and validating the correlation between morphology and gene expression in bacterial sepsis CD8+ T cells using deep learning approach

Jong Hyun Kim

Department of Biomedical Systems Informatics
The Graduate School, Yonsei University
Directed by Professor by Yu Rang Park

I. INTRODUCTION

T cell morphology undergoes significant alterations during activation characterized by changes in cell size, membrane fluidity, and cytoskeletal rearrangements. These changes are mediated by a complex series of internal events that are regulated by a variety of gene expression patterns to enhance immune response efficiency[1,2,3,4,5]. This relationship between cell morphology and gene expression is complex and bidirectional since changes in cell morphology can induce shifts in gene expression, just as variations in gene expression can affect cell morphology[6,7]. Morphological changes are involved in the formation of immune synapses, and these structural changes are essential for the interaction between T cells and antigen-presenting cells, and to enhance the immune response[8]. Moreover, variations in the morphological phenotypes of CD8 T cells were observed owing to genetic

deficiencies associated with immune functions[9]. Also, the chromatin structure of T cells undergoes dynamic shifts in response to gene expression and cellular signals. These changes regulate the accessibility of DNA through histone modification and chromatin remodeling[10]. This plays a central role in the immune response of T-cells and helps our immune system to work more effectively. These findings underscore the intricate relationship between cellular morphology and genetic regulation in T cell functions.

Advanced technologies such as single cell RNA sequencing (scRNA-seq)[11] and three-dimensional quantitative phase imaging (3D-QPI)[12] have led to advances in biology and many immune disease research fields. Single-cell RNA sequencing (scRNA-seq) has enabled the unbiased categorization of cell types from a wide range of samples, and driven advances in many research areas including immunology[13,14], developmental biology[15,16], and oncology[17,18]. It allows the comprehensive parallel analysis of thousands of cells at the transcriptomic level and has enabled us to understand complex patterns of gene expression that were previously undetectable.

Three-dimensional quantitative phase imaging is a significant development over traditional two-dimensional (2D) imaging technique. Unlike 2D imaging that provides a flat, limited view of cell morphology, 3D-QPI enables the imaging of living cells in three dimensions (3D) without the need for labelling at a single cell level[12]. The method does not involve staining or fixation procedures, thus preserving the intrinsic state of the cells and unaffecting their intracellular structure which offers a more detailed view of their morphological characteristics. This is essential to accurately capture the dynamic and complex morphological changes in

T cells during activation. In addition, 3D-QPI could provide quantitative information of cellular biochemical and biophysical properties such as cell volume and dry mass in high resolution[19,20]. Consequently, 3D-QPI was used to conduct research in a variety of fields, including immune cells[21,22], blood cells[23], and cancer cells[24].

Single cell RNA-seq provides deep insights into the gene expression profiles of individual cells, although these insights are limited to the gene expression perspective and have limitations in directly linking to cellular morphological changes. However, 3D-QPI cannot relate these changes directly to genetic factors or underlying gene expression. Therefore, understanding the changes in gene expression associated with T cell activation and the resulting morphological changes remains a challenge.

This can be addressed by integrating scRNA-seq with 3D-QPI techniques to deliver a comprehensive understanding of the relationship between gene expression and morphological changes at the cellular level. Single cell RNA-seq captures a single cell's gene expression in high resolution to understand its status and function, while 3D-QPI provides a detailed view of the cell's morphological characteristics. The combination of these two technologies may potentially identify specific genes responsible for morphological changes in cells and reveal the biological pathways involved in cellular structural changes. This approach may provide a new understanding of the interplay between gene expression and cellular morphology.

This study discovered the correlations between morphological changes in CD8 T cells and specific gene expression patterns through deep learning algorithms by

combining scRNA-seq and single cell 3D cell morphology image data. We developed a deep learning model that integrates single-cell transcriptomics and 3D QPI cell imaging dataset to identify genes associated with morphological changes. Our model was trained by collecting longitudinal data from sepsis patients, where the immune shift rapidly varies based on the severity[25,26,27]. Specifically, we focused on CD8 T cells since they are known for their rapid adaptability in immune responses. This allowed us to closely observe the rapid immune shifts that occur in sepsis patients and provide a more immediate insight into the dynamic changes[28,29,30]. Subsequently, we validated the identified gene expression patterns using published sepsis and COVID-19 scRNA-seq datasets. These findings revealed the biological understanding of a potential sepsis diagnostic biomarker based on cellular morphology previously discovered by our research team. Moreover, the patient's immune cell could reflect the patient's immune status. This enables the possibility of developing new treatments targeting cellular morphology.

II. RESULTS

Overview of the study workflow

The design of our study is shown in Figure 1. We enrolled patients with septic shock from the emergency room of a tertiary academic hospital. Demographic attributes of the patient cohort are detailed in Supplementary Table 1. Blood samples were collected at three specific time points: during the acute phase of septic shock (T1), after the resolution of the shock (T2), and immediately prior to hospital discharge (T3) (Fig. 1a). Single-cell RNA sequencing (scRNA-seq) and the gene selection pipeline was performed following the isolation of peripheral blood mononuclear cells (PBMCs) from these samples, (Fig. 1b). Additionally,

three-dimensional images were acquired after magnetic-activated cell sorting (MACS) of CD8 T cells using holotomography.

Gene expression levels were predicted based on a deep learning model that utilized 3D-QPI of cells as input data (Fig. 1c). Performance was evaluated using the mean absolute percentage error (MAPE), and 84 genes were selected with MAPE scores under 20%.

Initially, gradient-weighted class activation mapping (Grad-CAM) was applied to our morphology-gene prediction model to corroborate the association between gene expression and cellular morphology (Fig. 1d). Furthermore, the selected genes were cross-referenced with the Rosetta Project database[31], which specializes in gene-morphology associations. The selected morphology-specific genes underwent clustering and pathway analyses to assess their relevance to patient condition as a secondary validation step. Public datasets, including those with sepsis[32,33] and COVID-19[34] patient data were employed for further analyses. The relationship between these morphology-specific genes and patient severity was examined.

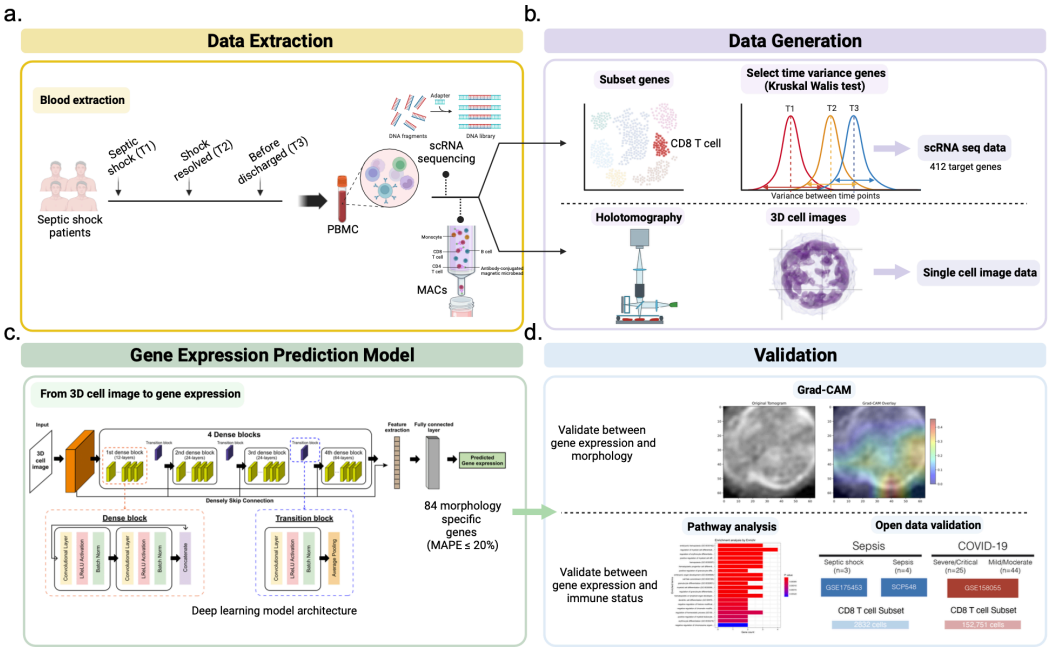


Figure 1. Schematic overview of the study workflow

a. Blood samples were collected from sepsis patients at three different time points based on severity. Peripheral blood mononuclear cells (PBMCs) were isolated from blood samples and part of the sample was used for scRNA-sequencing. The other part of the cells were subjected to single cell sorting by MACs from PBMCs and used to obtain 3D-QPI cell images. **b.** We then processed a subset of the scRNA-seq data to select target genes. Genes with low expression were filtered and genes showing a time-dependent expression pattern were identified using the Kruskal-Wallis test. A total of 412 target genes were selected. Holotomography imaging technology was used to capture 3D images of CD8 T cells. Approximately 100-200 cell images were obtained from each sample and a total of 1,639 images were retained for further study after a quality control pipeline. **c.** The deep learning model was developed to predict the expression levels of the selected genes using the 3D-QPI cell images as input and gene expression values as output. The model performance was evaluated using the mean absolute percentage error (MAPE), and 84 genes were selected with MAPE score of less than 20%. **d.** Grad-CAM was applied to the selected gene model and cross-referenced with the Rosetta project database. Further validation processes were done through scRNA-seq analysis using our data and public datasets (including sepsis and COVID-19 patient data) to assess their relevance to patient severity.

Morphology specific gene selection

A total of 75,297 genes and 33,521 cells from 11 samples were generated using the scRNA-seq preprocess and annotation pipeline (Fig. 2a and Supplementary Fig. 1). CD8 T cells from the scRNA-seq data, and only genes that expressed more than 10% of cells were used for analysis[35,36]. We then identified 3,453 genes that were common to all time points and selected 412 target genes according to the Kruskal-Wallis test ($P < 0.05$), which showed variations in gene expression patterns across different time points (Fig. 2b and Supplementary Fig. 2)

We developed a deep learning model using 3D cell morphology as input data to predict the gene expression values. We conducted a patient-level leave-one-out cross validation test among patients to account for variance among the 3D cell images. We computed the mean absolute percentage error (MAPE) between the predicted and observed expressions of each gene, and only selected genes with a MAPE of 20% or lower in each patient-level leave-one-out cross validation test[37,38]. A total of 84 genes consistently met the model's performance threshold across all three patient-level leave-one-out cross validation tests. The median MAPE values were between 2–18% for the selected genes according to the patient-level leave-one-out cross validation tests (Fig. 2c).

We also compared the results to the 3D-CNN-VIT model to further evaluate the performance and robustness of our current study model. The correlation of the MAPE and MSE from the two models demonstrated the consistency and reliability of our model (Supplementary Fig. 3). Our model identified 84 genes with MAPE values below 20%, while the 3D-CNN-VIT model identified 75 genes. This

comparison strengthened the validity of our model and revealed its robustness in gene identification in a research context.

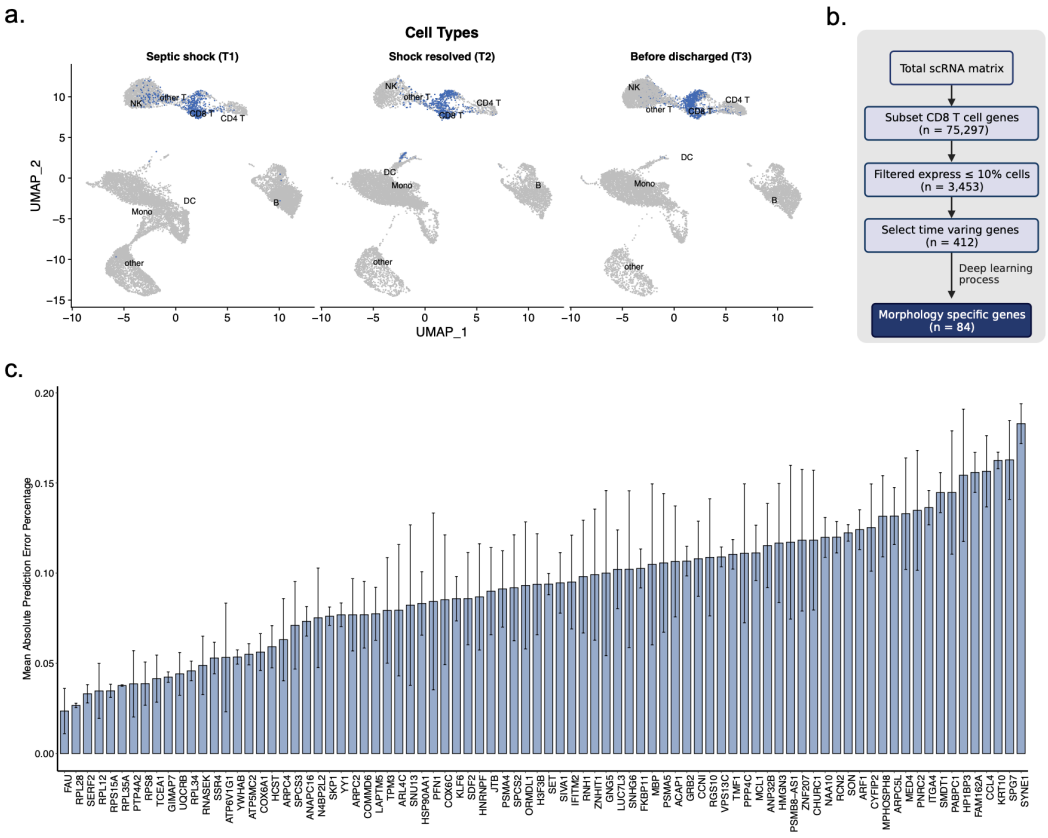


Figure 2. Identification of morphology-specific genes through deep learning

a. Uniform manifold approximation and projection (UMAP) of the septic shock (T1), shock resolved (T2), and before discharged (T3) condition. Each dot corresponds to a single cell, and only clusters of CD8 T cells are colored. **b.** Target gene selection pipeline, starting with a subset of CD8 T cells. Identification of 3,453 common genes at all time points, and further identification of 412 target genes with varying expression patterns over time (Kruskal-Wallis test, $P < 0.05$; detailed expression values over time points are shown in Supplementary Fig. 2) **c.** Bar plot showing the median MAPE percentage loss of 84 selected genes using leave-one-out cross-validation. Performance threshold with a MAPE of 20% or lower genes are selected, and genes are sorted in order of low MAPE percentage values.

Validating the association between morphology specific gene expression and cell morphology

We employed the Grad-CAM algorithm to elucidates the critical regions within the 3D images that the deep learning model prioritizes for prediction. We selected ARL4C, RNH1, and SKP1 as representative models, which overlapped with the genes listed in the Rosetta database31, with (MAPE) performances of 7.9%, 9.8%, and 7.6%, respectively. Fig. 3a depicts representative cellular images for each temporal phase accompanied by their corresponding importance heatmaps to describe the features critical for the model's predictions. Notably, the heatmaps highlighted the interior regions of the images. Additionally, we assessed the spatial distribution of gradient weights throughout all 3D-QPI cell images for a quantitative comprehension of the cellular regions. Gradient weights were significantly concentrated in the central region of the image for the ARL4C gene model, particularly in shells 1, 2, 3, and 4 (Fig. 3b). In contrast, gradient weights for the SKP1 and RNH1 gene models were mainly concentrated in shells 4, 5, and 6, and not in the center of the cell. This suggests that the model primarily emphasizes intracellular structures, such as chromatin density or organizational patterns in the nucleus and cytoplasm.

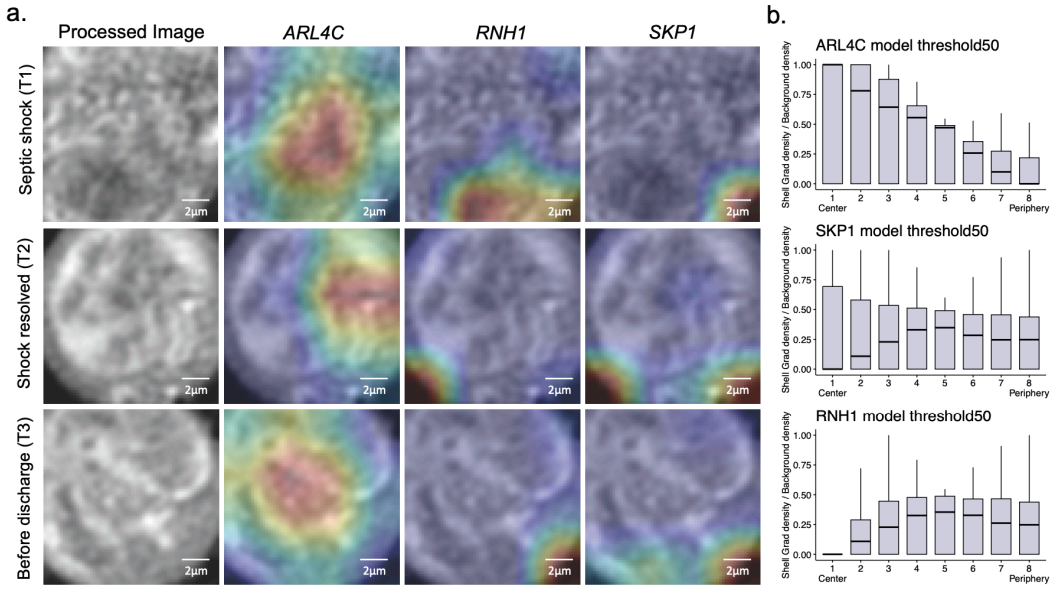


Figure 3. Differential morphological regions influencing gene expression predictions

a. Representative cell images at each time point and the corresponding Grad-CAM heatmap overlaid on the processed images highlighting crucial regions within the cell. Heatmaps indicate a concentration in the central regions of the cell in the case of the ARL4C gene model, emphasizing intracellular structures. In contrast, the RNH1 and SKP1 gene models show a focus on the peripheral regions of the cell that highlight external cellular structures. **b.** Boxplot illustrating the spatial distribution of gradient weights across all 3D cell images. The distribution of gradient weights indicates important cellular regions that may be associated with the expression of ARL4C, RNH1, and SKP1 genes.

Validating the association between morphology, specific genes, and patient status

We conducted scRNA-seq analysis to understand the results of our deep learning model through biological insights. The goal was to gain a deeper understanding of how the expression of each of the 84 identified genes affects cell morphology and structure. We employed hierarchical clustering to scrutinize their expression patterns across three distinct temporal phases and conducted an ANOVA test for each gene (Supplementary Table 3). Six discrete clusters emerged from this analysis, with statistically significant genes denoted accordingly (Fig. 4a). We specifically focused on clusters 2 and 6 were of particular interest in our study. Cluster 2 contained genes such as TMF1, SPG7, VPS13C, CCL4, CYFIP2, KLF6, N4BP2L2, HP1BP33, ZNF207, HMGN3, MPHOSPH8, SSR4, SYNE1, MBP, NAA10, while cluster 6 composed genes with ATP6V1G1, GNG5, ARPC4, RNH1, GIMAP7, HCST, YWHAB, PPP4C, ARPC5L, ATP5MC2, FAU, RPL12, SDF2, SNHG6, RPS15A, JTB, PSMA5, H3F3B, PFN1, SMDT1, RPL34, LAPTM5, RPL35A, RPS8, MED4, SKP1, SIVA1. These specific genes were found to be involved in influencing cell morphology during the recovery phase of septic shock. Our morphological assessment of cells throughout the septic shock recovery trajectory revealed pronounced differences between T1 and both T2 and T3. No appreciable morphological differences were noted between T2 and T3 (Supplementary Fig. 4). Consequently, we focused on clusters 2 and 6, which exhibited down- and up-regulated gene expression at T1, respectively.

Pathway enrichment analysis showed that cluster 2 prominently showed an upregulation in biological processes related to translation recovery, like

cytoplasmic translation and ribosomal assembly restoration (Fig. 4b). Also, cellular components such as the cytosolic small ribosomal subunit, large ribosomal subunit, and their respective counterparts were identified. This illustrates the interrelationship between these biological processes and structural components in shock recovery[39]. Cluster 6 was upregulated in the septic shock phase and highlighted the changes in chromatin organization. This suggests that cells reorganize their DNA structure to cope with cellular stress[40]. Changes in histone acetylation and DNA methylation further emphasize the changes of chromatin organization, and response to UV-C radiation suggests that cells may also be processing and repairing DNA damage[41,42]. The results from the cellular component also identified the nucleus, which is the cellular component that stores and organizes DNA.

In addition, we defined a score for each cell based on the expression of a set of relevant genes to characterize the translation and mRNA metabolic responses by time points. CD8 T cell septic shock patients decreased the translation process and upregulated the mRNA metabolic processes (Fig. 4c). This indicated an emphasis on activities associated with transcription Specifically, genes involved in translation, RPS8[43], and RPL35A[44] were expressed at slightly lower levels during septic shock phase (Fig. 4d). HP1BP3[45] and SET[46] are representative genes associated with chromatin structure that were both upregulated in septic shock patients (Fig. 4e). These biological findings indicate that the interplay between changes in chromatin structure and complex transcriptional regulation may be associated with changes in cell morphology during septic shock.

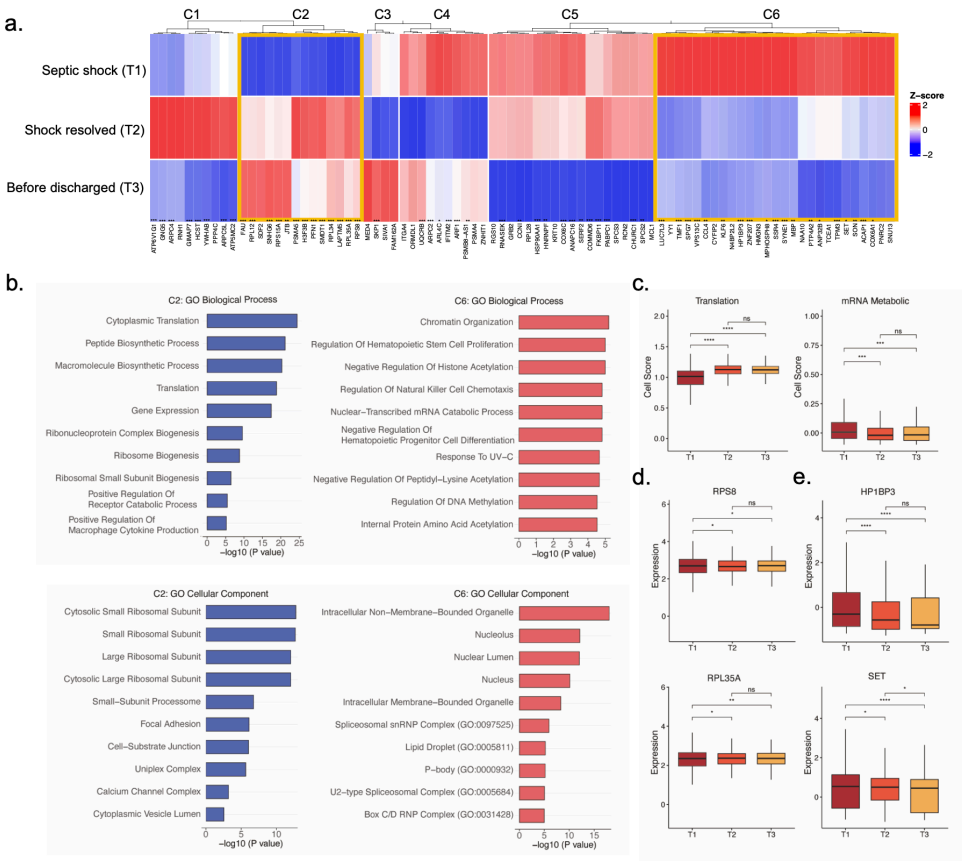


Figure 4. Genetic and morphological insights into CD8 T cells in septic shock

a. Hierarchical clustering of 86 genes selected by deep learning model between septic shock (T1), shock resolved (T2), and before discharged (T3). Six distinct clusters were identified during the different phases of septic shock. **b.** Bar plots showing pathway enrichment analyses associated with the genes in cluster 2 and 6 (cluster 2, n = 14, cluster 6, n = 27). The top 10 Gene Ontology (GO) terms of Biological Process and Cellular Component were enriched. **c.** Box plots of the median cell scores for two GO biological process terms of septic shock (T1), shock resolved (T2), and before discharged (T3). Horizontal lines represent median values of cell scores and all differences with $P < 0.05$ were indicated after t-test analysis. **d.** Box plots of RPS8 and RPL35A genes related to translation, and **e.** Box plots of HP1BP3 and SET structure mRNA expression (log-normalized) correlated with chromatin structure.

Comparative analysis of gene expression pattern across open sepsis and COVID-19 scRNA-seq data

We aimed to investigate whether the gene expression patterns during the recovery phase of septic shock persisted across different sepsis severity groups in severe and moderate cohorts. Moreover, we sought to determine if these gene expression changes were exclusive to septic shock or if they also presented in other inflammatory conditions, such as COVID-19. Lymphocytes from patients infected with COVID-19 also induced cellular morphological changes, including increased cell volume, and altered cytoplasmic structure[47,48]. We therefore assessed the transcriptional similarity of genes in our resulting cluster 2 and 6 in two published sepsis datasets[32,33], and one published dataset of COVID-19[34].

We integrated two distinct published datasets to compare gene expression patterns with published sepsis datasets. We also used published data for the COVID-19 comparison, and both published data were subtracted only on the gene matrix corresponding to CD8 T cells. These datasets were used to analyze the expression patterns of 14 genes from cluster 2 that were downregulated and 27 genes from cluster 6 that were upregulated during the septic shock (T1) phase. Within the published sepsis datasets, four genes from cluster 2 showed decreased expression patterns in the severe group, while 16 genes from cluster 6 showed similar patterns (Fig. 4a). In the published COVID-19 datasets, 12 genes showed decreased expression in the severe group, with 15 genes showing increased expression patterns. To further refine our analysis of gene expression patterns in sepsis and COVID-19 datasets, we utilized a more advanced statistical approach. We employed a linear regression model to assess the differences in gene expression

coefficients between our study datasets and the open public datasets. Our comparative analysis revealed that the gene expression patterns of PTP4A2, TCEA1, TPM3, and YY1 genes were statistically similar in the open sepsis datasets. Similarly, in the open COVID-19 datasets, ACAP1, N4BP2L2, SET, and SON genes exhibited comparable gene expression patterns. These findings highlight specific transcriptional similarities in the context of different inflammatory conditions (Supplementary Table 4,5).

The upregulation of TCEA1, YY1 and SET in the context of infection reveals an adaptive genetic response of T cells, suggesting their potential role in modifying chromatin structure[49,50,51]. PTP4A2 and N4BP2L2, although not directly involved in chromatin structure modification, appear to influence it indirectly through mechanisms such as histone acetylation[52,53]. SON potentially coordinates pivotal pathways that can lead to changes in cell morphology and function during septic shock by regulating the assembly and function of intracellular structures such as centrosomes and the microtubule cytoskeleton[54]. TPM3, while not directly implicated in chromatin restructuring, plays an essential role in the formation of actin filaments[55]. Lastly, ACAP1 stands out for its direct impact on cell morphology through the induction of membrane protrusions[56].

Figure 5. Comparative transcriptional analysis in sepsis and COVID-19 conditions

a. Dot plots indicating the expression patterns of key genes from cluster 2 and 6 across sepsis (up) and COVID-19 (down) conditions. The size of the dot indicates the percentage of cell expression and color represents the average expression level of the gene in those cells. Genes with the same expression pattern between sepsis and COVID-19 are highlighted.

III. DISCUSSION

We ascertained that the morphology-specific genes correlated with CD8 T cell morphology but also hold associations with various facets of patient status, including the disease's temporal progression and severity. The deep learning techniques isolated genes that were intrinsically linked to alterations in cellular morphology. We visually demonstrated that these genes are associated with spatial variations in cellular morphology according to the Grad-CAM algorithm. This conclusion was further substantiated through quantitative analysis. Pathway analysis revealed that these genes also bear biological associations with cellular morphology and are directly linked to patient status, specifically concerning the severity of sepsis. Additionally, the expression patterns of these genes exhibited similarities in other sepsis datasets and in other infectious diseases such as COVID-19. Our results illuminate the role of cellular morphological changes in disease progression and offer novel insights into the linkage between these alterations and gene expression.

Predicting gene expression based on morphology is not a novel concept. It was performed in radiomics, which utilizes imaging techniques to forecast patient outcomes or gene expression, to recent studies in cancer research. Efforts are underway to predict gene expression by examining morphological characteristics of tissues under a microscope[57,58]. Furthermore, research is progressing to predict patient prognosis based on these tissue morphologies[59]. Such efforts are not confined to tissue-level analyses; they are also being extended to the single-cell level. Kerren et al. employed autoencoders to integrate disparate data modalities, such as RNA-seq and chromatin images[60] to provide a generalized framework. Our research is specifically tailored to understand sepsis at the single-cell level. We have

demonstrated the practical relevance of single cell morphology and elucidated its capability to mirror molecular variations and fluctuations in patient status.

Based on common expression patterns observed in published sepsis and COVID-19 data, we identified eight up-regulated genes and five genes among those were involved in chromatin mechanisms. Our study observed expansion in cell volume and surface area during the septic shock phase, which seems to be indirectly influenced by the increased activity of these chromatin-related genes. Over expression of these genes may indirectly affect cell morphology through changes in transcriptional activity and subsequent functional modifications during septic shock given the central role of chromatin dynamics in regulating cellular responses to stress. PTP4A2, TCEA1, YY1, N4BP2L2 and SET are involved in transcriptional regulation and chromatin remodelling[49,50,51,52,53]. The process of chromatin remodeling is regulated and leads to changes in transcriptional activity when these genes are activated during septic shock. These changes can affect cell morphology and function and lead to an amplified stress response.

Our study provides valuable insights into the relationship between cellular morphology and gene expression, particularly in the context of sepsis and its severity. However, there are some limitations. First, our study employed stringent inclusion and exclusion criteria (Supplementary Table 2), which enhanced the reliability of our findings, but limited the sample size and potentially restricting the generalizability of our results. This limitation was overcome by validating our findings using external datasets focused on sepsis and COVID-19, thereby reinforcing the robustness of our results despite the constrained sample size. However, we were unable to externally validate our findings in this specific domain

owing to the absence of publicly available datasets that combine cell morphology and gene expression. Nonetheless, we ascertained that the identified genes are reflective of patient status in existing sepsis and COVID-19 datasets. Lastly, we only focused on CD8 T cells. This narrow focus potentially overlooks the broader applicability of our findings across different cell types.

Our study expands the current understanding of cellular structure at the single cell level and provides a basis for new diagnostic and therapeutic approach. The study explored the relationship between CD8 T-cell morphology and gene expression, providing a biological interpretation of the results of previous studies and establishing the basis for a potential biomarker of sepsis severity. Our results shed new light on the intricate interplay between cellular structure and gene expression patterns and suggest that changes in cellular shape may serve as early indicators of disease progression and the efficacy of therapeutic treatments. This research represents the beginning of a diagnostic method based on cell morphology that will improve the accuracy of personalized treatment for sepsis and several related immune diseases.

IV. CONCLUSION

This study revealed a strong association between CD8 T cell morphology and gene expression. We highlighted genes and pathways that play an important role in cellular structural changes during sepsis and are associated with disease severity. A key aspect of our findings is the role of chromatin organization in driving these cellular changes. We discovered that changes in chromatin structure, particularly during septic shock, are crucial in determining the morphological changes observed in CD8 T cells. The reorganization of chromatin not only reflects the cell's adaptive

response to stress but also seems to be a critical factor in the maintenance of cellular morphological changes. This understanding has provided a fundamental framework for further exploration in the field of immune-related diseases. Moreover, these findings suggest a new approach to the development of future disease diagnostics and treatment strategies that focuses on cellular morphological changes, which may lead to the development of more effective and personalized treatment.

V. METHODS

Patient enrollment

The study enrolled septic shock patients in the Emergency Department of Severance Hospital, Seoul, Korea. Eligible participants were diagnosed with sepsis according to the consensus definition for sepsis[25]. Patients who were immunocompromised or were receiving immunosuppressive medications were excluded. Detailed inclusion and exclusion criteria for patient selection are provided in Supplementary Table 2. Clinical data were collected for each enrolled patient, and blood was sampled at three distinct time points: during septic shock (T1), after shock resolution (T2), and immediately prior to hospital discharge (T3).

Informed consent was obtained from all study participants. Ethical approval for this study was granted by the Institutional Review Board of Severance Hospital, Yonsei University Health System, Seoul, Korea (IRB numbers 4-2021-1236 and 4-2022-0317)

Peripheral blood mononuclear cell isolation and CD8 T cell sorting

Peripheral blood samples were collected in EDTA tubes and processed utilizing Ficoll-Paque Plus separation techniques, as provided by GE Healthcare (Barrington, IL, USA, Catalog No. 17144002). Initially, the blood was diluted with 5 mL of 2

mM EDTA-PBS obtained from Invitrogen (Carlsbad, CA, USA, Catalog No. 1555785-038). Subsequently, 10–20 mL of this diluted mixture was carefully layered over 15 mL of Ficoll in a 50 mL Falcon tube. Centrifugation was performed at 900 g for 30 min. The plasma layer was removed, and the PBMC layer was isolated. Subsequently, this layer was washed with EDTA-PBS and subjected to an additional centrifugation step at 500 g for 5 min. The resultant PBMC pellet was harvested, and its cell count, and viability were assessed using Trypan blue staining and a Countess II Automated Cell Counter (Thermo Fisher Scientific, Waltham, MA, USA).

Magnetic-activated cell sorting (MACS) technology sourced from Miltenyi Biotec (Bergisch Gladbach, Germany) was employed for the sorting of CD8 T cells. To preserve cell viability, the cells were maintained at 4°C for 15 min, followed by the addition of 80 μL of isolation buffer and 20 μL of magnetic microbeads. The cells were incubated at 4°C for an additional 15 min. A magnetic stand and column were prepared prior to the sorting procedure, and the column was equilibrated with 3 mL of isolation buffer. The column was removed, and the lymphocytes were collected into a conical tube. Finally, an additional 5 mL of isolation buffer was introduced, facilitating the specific extraction of CD8 T cells.

Quantitative phase image acquisition and preprocessing

Three-dimensional QPI of CD8 T cells was accomplished using 3D holotomography technology (HT-2H; Tomocube Inc., Daejeon, Republic of Korea)12. This technology generates a 3D refractive index (RI) image by amalgamating multiple 2D QPIs. The cellular RI serves as an intrinsic optical

parameter indicative of the way light traverses the cellular architecture. Importantly, the RI is associated with cellular mass and its spatial distribution.

A quality control process was initiated after image acquisition to eliminate low-quality, low-resolution, and noisy images through manual selection. Additionally, images featuring two or more closely juxtaposed cells were excluded to maintain focus on single-cell observations. Following the quality control phase, preprocessing was performed to prepare the images for utilization within the deep learning model. Specifically, the images were center cropped from dimensions of 210 x 276 x 276 pixels to a more concentrated 64 x 64 x 64 pixels, aimed at highlighting the internal cellular landscape. Lastly, min-max normalization was applied to each image[61].

Single-cell transcriptomics

Library preparation and sequencing

The prepared cell suspensions were used to generate a 10x Chromium Single Cell 3' library with Chromium Single Cell 30 v3 reagent (10x Genomics) according to the manufacturer's instructions. Approximately 10,000 cells were loaded per sample. The library was then sequenced using Illumina Nova 6000 according to the manufacturer's instructions.

Data preprocessing and quality control

Single-cell RNA (scRNA) sequencing data were processed using Cell Ranger Software (v6.1.3) to perform alignment, filtering, barcode separation, and unique molecular identifier (UMI) counting with default parameters. Raw reads were aligned to the human reference genome GRCh38 (GENCODE v32/Ensembl 98).

Feature-barcode matrices were generated for secondary analysis for each sample. We used CellBender software (v0.2.0) to remove background RNA contamination and barcode replacement from raw UMI-based scRNA-seq data. Subsequently, each sample was initially subjected to doublet removal using the scDblFinder R package (version 1.6.0)[62] to ensure the quality of the data.

The following criteria were used to filter the cells: (1) the number of genes sequenced per cell ranged between 200 to 4000, and (2) the percentage of mitochondrial RNA per cell was below 10%. After filtering, a total of 33,521 high-quality cells was obtained from 11 samples. These samples included three longitudinal samples from three distinct patients, and samples from two additional patients with septic shock.

Integration of scRNA-seq individual samples

A filtered gene-barcode matrix of all samples was integrated with the Seurat R package (version 4.3.0) using the integration pipeline to remove batch effects across different samples. We first normalized the Seurat objects and identified the top 2,000 variable features using the 'NormalizeData' and 'FindVariableFeatures' functions, respectively.

Integration features were selected using the 'SelectIntegrationFeatures' function, and 'anchors' were identified by the 'FindIntegrationAnchors' function. These anchors were used to integrate the datasets using the 'IntegrateData' function, resulting in a new 'integrated' assay which was set as the default for subsequent analysis. Then, we scaled the data using the 'ScaleData' function, performed Principal Component Analysis (PCA) using the 'RunPCA' function, followed by uniform manifold approximation and projection (UMAP) on the top 30 principal

components using the 'RunUMAP' function. Finally, we constructed a Shared Nearest Neighbor (SNN) graph with 'FindNeighbors' and identified clusters with the 'FindClusters' function at a resolution of 1.0.

Cell type identification and downstream CD8 T cell subpopulation analysis

Cell type identification was conducted through differential gene expression (DGE) analysis, using the 'FindAllMarkers' function in Seurat with a minimum cell fraction of 25% and a log fold-change threshold of 0.25. This analysis resulted in the annotation of cell clusters corresponding to monocytes, neutrophils, macrophages, B cells, T cells, natural killer (NK) cells, platelets, and hemoglobin populations. We validated the cell type annotations result using the Azimuth tool by comparing our annotation to a published PBMC dataset.

Downstream analysis was performed on the T cell population, which was divided into CD8 and CD4 T cell subsets using the 'subset' function in Seurat. This downstream analysis included a standard Seurat workflow: 'FindVariableFeatures', 'ScaleData', 'RunPCA,' 'FindNeighbours,' 'FindClusters,' and 'RunUMAP'. The 'FindClusters' function was performed with a resolution of 0.6 and UMAP was performed on the top 15 principal components.

Target gene selection and gene expression matrix normalization

A gene filtering strategy was employed to mitigate the influence of noise in the single-cell RNA sequencing (scRNA-seq) dataset. Specifically, only genes expressed in a minimum of 10% of the cells were selected, thereby eliminating genes with low expression values[35,36]. Following this criterion, 3,961 genes in CD8 T cells were identified for further analysis. A Kruskal-Wallis test was executed on the curated gene sets given that the primary objective was to discern genes

exhibiting temporal differences in expression during sepsis. A logarithmic transformation was applied to the gene matrix, utilizing a transformation of the form a→log10(1 + a) to address potential model bias favoring highly expressed genes in the deep learning regression analysis.

Developing a gene expression prediction deep learning model

We developed a deep learning model to predict gene expression levels from 3D QPI cell images. We use the architecture that integrated a Multi-Layer-Perceptron (MLP) with a 3D Dense Convolutional Network (DensetNet)[63,64]. The architecture is comprised of 82 dense layers, four dense blocks, and three transition layers. This architecture allowed for the extraction of features and optimized information between layers. The output of the model was compressed from high-dimensional feature maps to a 1-dimensional vector using adaptive average pooling. This compressed feature vector was then passed into the MLP for further processing. The MLP architecture was composed of three hidden layers, with sizes of 256, 128, and 64, respectively. Each layer consisted of a linear transformation, ReLU activation function, and dropout regularization. Finally, a MLP at the end of the network performed a regression task to predict the expression level of each gene.

The obtained images of CD8 T cells were divided into training, validation, and test sets at an 8:1:1 ratio. We implemented augmentation using random rotations, and horizontal and vertical flips to prevent overfitting and improve the generalization of the model. The model was trained for 1,000 epochs, with early stopping if there was no improvement in the validation set for 10 epochs. A MAPE loss and the Adam optimizer algorithm with 16 mini-batch sizes were applied to train the model. The learning rate was set to an initial step size of 0.001. We used a leave one out

test dataset approach, where individual patients were treated as separate test sets to evaluate the model's performance. This approach ensured an unbiased assessment of the model's generalization capability. All deep learning processes were performed using PyTorch (version 2.0.0) on a server that had two NVIDIA Tesla V100, 16 Gb memory, with CUDA version 11.1.

Validating the the association between cell morphology and gene expression

We used Grad-CAM to identify the regions within the cell that were most important in predicting gene expression[65]. We developed 3D Grad-Cam with a customized 3D convolutional neural network model for model interpretation. The gradients were combined in a weighted manner to give us an understanding of the most significant features. These significant features were then highlighted on a heatmap that can be mapped over the cell image, revealing areas of the cell that are key to predicting gene expression. This process was separately applied for each gene model of interest.

The Grad-CAM was quantified by projecting the Grad-CAM heatmap onto a 2D sliced cell images by focusing on the central z axis of the 3D cell image. We then divided this sliced image into concentric shells based on their Euclidean distance from the center. We divided it into eight distinct shells, with the central and outermost region termed shell 1 and shell 8, respectively. Each voxel within the slice was classified for density analysis.

All Grad-CAM weights were considered to calculate the density of individual regions within each shell. The gradient weight density of each shell was then calculated. This approach allowed us to identify the spatial distribution of Grad-CAM weights across the inner cell region.

Pathway enrichment analysis

Pathway enrichment analysis involves analyzing biological process and cellular components. The Enrichr platform provides a comprehensive gene enrichment analysis using databases with rich gene set annotation pathway information analysis[66]. The log (P value) was used as a parameter of enrichment. The top 10 enriched terms were selected for analysis. The results were obtained from the REST APIs provided by Enrichr.

Cell soring

We used the 'AddModuleScore' function from the Seurat package to compute cell scores. We focused on the gene sets associated with "Regulation of mRNA Metabolic Process (GO:1903311)" and "Translation (GO:0006412)" from the Gene Ontology database[67]. The average expression of genes within these gene sets were calculated for each cell.

Validating the association between morphology-specific genes and patient status

We used public datasets that closely mirrored the sepsis severity in our patient cohort to compare gene expression pattern using public data. We selected data from Dijoia et al[33] that focused on septic shock patients, and Miguel et al[32] that included patients with bacterial moderate sepsis to match the heterogeneity of sepsis observed in our study. We also used PBMC samples from 68 mild/moderate and 84 severe/critical COVID-19 patients without comorbidities from the data of Xianwen et al[34].

Statistical analysis

We used a comprehensive statistical approach to identify target genes for training our deep learning model. A statistical approach was applied after filtering and selecting genes that were commonly expressed at all time points. We selected genes that satisfied equality of variance using the Levene test ($p > 0.05$) and genes with differences in gene expression across time points using the Kruskal-Wallis test ($p < 0.05$). We found genes that significantly changed over time and had constant variance between groups by selecting genes that met both criteria.

Morphological features and gene expressions were compared using Student's t-test for each time point. We employed linear regression models to assess the differences in gene expression coefficients between our study datasets and the open public datasets. This comparison focused on the expression coefficients of different conditions in our study dataset and in the sepsis and COVID-19 groups. This test effectively identified statistically significant gene clusters, revealing key differences and similarities in gene expression between our data and open public datasets.

Statistical analyses were conducted using R software (version 4.1.0). Statistical significance is indicated as follows: * for $P < 0.05$, ** for $P < 0.01$, and *** for $P < 0.001$.

REFERENCES

1. Dustin, M. L. & Cooper, J. A. The immunological synapse and the actin cytoskeleton: Molecular hardware for T cell signaling. Nat Immunol 1, 23–29 (2000).

2. Iannicola, C. et al. Early alterations in gene expression and cell morphology in a mouse model of Huntington's disease. J Neurochem 75, 830–839 (2000).

3. Fooksman, D. R. et al. Functional anatomy of T cell activation and synapse formation. Annu Rev Immunol 28, 79–105 (2010).

4. Faure, S. et al. ERM proteins regulate cytoskeleton relaxation promoting T cell-APC conjugation. Nat Immunol 5, 272–279 (2004).

5. Junker, J. P. & Van Oudenaarden, A. Every cell is special: Genome-wide studies add a new dimension to single-cell biology. Cell 157, 8–11 (2014).

6. Esfahani, P. H. & Knoll, R. Cell shape: effects on gene expression and signaling Payam. Biophys Rev 12, 895–901 (2020).

7. Drareni, K., Gautier, J. F., Venteclef, N. & Alzaid, F. Transcriptional control of macrophage polarisation in type 2 diabetes. Semin Immunopathol 41, 515–529 (2019).

8. Lin, W. et al. Morphological Change of CD4+ T Cell during Contact with DC Modulates T-cell Activation by Accumulation of F-actin in the Immunology Synapse. BMC Immunol 16, 49 (2015).

9. German, Y. et al. Morphological profiling of human T and NK lymphocytes by high-content cell imaging. Cell Rep 36, (2021).

10. Henning, A. N., Roychoudhuri, R. & Restifo, N. P. Epigenetic control of CD8+ T'cell differentiation. Nat Rev Immunol 18, 340–356 (2018).

11. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50, 1–14 (2018).

12. Park, Y. K., Depeursinge, C. & Popescu, G. Quantitative phase imaging in biomedicine. Nat Photonics 12, 578–589 (2018).

13. Zhang, J. Y. et al. Single-cell landscape of immunological responses in patients with COVID-19. Nat Immunol 21, 1107–1118 (2020).

14. Lee, J. S. et al. Immunophenotyping of covid-19 and influenza highlights the role of type i interferons in development of severe covid-19. Sci Immunol 5, (2020).

15. Cai, C., Yue, Y. & Yue, B. Single-cell RNA sequencing in skeletal muscle developmental biology. Biomedicine and Pharmacotherapy 162, 114631 (2023).

16. Peng, Y. & Qiao, H. The Application of Single-Cell RNA Sequencing in Mammalian Meiosis Studies. Front Cell Dev Biol 9, (2021).

17. Steele, N. G. et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. Nat Cancer 1, 1097–1112 (2020).

18. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat Commun 11, (2020).

19. Vasilenko, I., Metelin, V., Kuznetsov, A., Belyakov, V. & Yakushina, T. Opportunities of {{QPI}} in the Epigenetic Diagnostics and Assessment of Therapeutic Efficacy. in Quantitative {{Phase Imaging III}} vol. 10074 74–78 (SPIE, 2017).

20. Nguyen, T. L. et al. Quantitative Phase Imaging: Recent Advances and Expanding Potential in Biomedicine. ACS Nano 16, 11516–11544 (2022).

21. Lee, M. et al. Deep-learning based three-dimensional 1 label-free tracking and analysis of immunological synapses of car-t cells. Elife 9, 1–53 (2020).

22. Jung, Y., Wen, L., Altman, A. & Ley, K. CD45 pre-exclusion from the tips of T cell microvilli prior to antigen recognition. Nat Commun 12, 1–16 (2021).

23. Ryu, D. et al. Label-Free White Blood Cell Classification Using Refractive Index Tomography and Deep Learning. BME Front 2021, (2021).

24. Park, D. et al. Cryobiopsy: A Breakthrough Strategy for Clinical Utilization of Lung Cancer Organoids. Cells 12, (2023).

25. Singer, M. et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA - Journal of the American Medical Association 315, 801–810 (2016).

26. Boomer, J. S. et al. Immunosuppression in patients who die of sepsis and multiple organ failure. JAMA 306, 2594–2605 (2011).

27. Delano, M. J. & Ward, P. A. The immune system's role in sepsis progression, resolution, and long-term outcome. Immunol Rev 274, 330–353 (2016).

28. Kaech, S. M., Wherry, E. J. & Ahmed, R. Effector and memory T-cell differentiation: Implications for vaccine development. Nat Rev Immunol 2, 251–262 (2002).

29. Whitmire, J. K., Murali-Krishna, K., Altman, J. & Ahmed, R. Antiviral CD4 and CD8 T-cell memory: differences in the size of the response and activation requirements. Philos Trans R Soc Lond B Biol Sci 355, 373–379 (2000).

30. Shedlock, D. J. et al. Role of CD4 T Cell Help and Costimulation in CD8 T Cell Responses During Listeria monocytogenes Infection . The Journal of Immunology 170, 2053–2063 (2003).

31. Haghighi, M., Caicedo, J. C., Cimini, B. A., Carpenter, A. E. & Singh, S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. Nat Methods 19, 1550–1557 (2022).

32. Reyes, M. et al. An immune-cell signature of bacterial sepsis. Nat Med 26, 333–340 (2020).

33. Darden, D. B. et al. A Novel Single Cell RNA-seq Analysis of Non-Myeloid Circulating Cells in Late Sepsis. Front Immunol 12, 1–11 (2021).

34. Ren, X. et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. Cell 184, 1895-1913.e19 (2021).

35. Farhadian, M., Rafat, S. A., Panahi, B. & Mayack, C. Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. Sci Rep 11, 1–15 (2021).

36. Sha, Y., Phan, J. H. & Wang, M. D. Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data. Annu Int Conf IEEE Eng Med Biol Soc 2015, 6461–6464 (2015).

37. Zeng, Z. et al. CycleDNN-A Novel Deep Neural Network Model for CETSA Feature Prediction cross Cell Lines. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2022-July, 1647–1650 (2022).

38. Schumaker, G. et al. Optical Biopsy Using a Neural Network to Predict Gene Expression From Photos of Wounds. Journal of Surgical Research 270, 547–554 (2022).

39. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. Nat Rev Mol Cell Biol 11, 113–127 (2010).

40. Voss, T. C. & Hager, G. L. Dynamic regulation of transcriptional states by chromatin and transcription factors. Nat Rev Genet 15, 69–81 (2014).

41. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nat Rev Genet 13, 484–492 (2012).

42. Cadet, J., Sage, E. & Douki, T. Ultraviolet radiation-mediated damage to cellular DNA. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis 571, 3–17 (2005).

43. Hao, Y. et al. CDK11p46 and RPS8 associate with each other and suppress translation in a synergistic manner. Biochem Biophys Res Commun 407, 169–174 (2011).

44. Boyle, E. A. et al. Skipper analysis of eCLIP datasets enables sensitive detection of constrained translation factor binding sites. Cell Genomics 3, 100317 (2023).

45. Dutta, B. et al. Profiling of the chromatin-associated proteome identifies HP1BP3 as a novel regulator of cell cycle progression. Molecular and Cellular Proteomics 13, 2183–2197 (2014).

46. Muto, S. et al. Relationship between the structure of SET/TAF-Iβ/INHAT and its histone chaperone activity. Proc Natl Acad Sci U S A 104, 4285–4290 (2007).

47. Pozdnyakova, O. et al. Clinical Significance of CBC and WBC Morphology in the Diagnosis and Clinical Course of COVID-19 Infection. Am J Clin Pathol 155, 364–375 (2021).

48. Zeng, X. et al. Monocyte volumetric parameters and lymph index are increased in SARS-CoV-2 infection. Int J Lab Hematol 42, e266–e269 (2020).

49. Chen, Danlei, et al. "PTP4A2 regulates Sorafenib resistance via activating autophagy in hepatocellular carcinoma." (2020).

50. Kee, Anthony J., et al. "An actin filament population defined by the tropomyosin Tpm3. 1 regulates glucose uptake." Traffic 16.7 (2015): 691-711.

51. DiMarco, Stephana P., et al. "Transcription elongation factor SII (TCEA) maps to human chromosome 3p22→ p21. 3." Genomics 36.1 (1996): 185-188.

52. Gordon, S., Akopyan, G., Garban, H. & Bonavida, B. Transcription factor YY1: Structure, function, and therapeutic implications in cancer biology. Oncogene 25, 1125–1142 (2006).

53. Gasilina, Anjelika, et al. "The ArfGAP ASAP1 controls actin stress fiber organization via its N-BAR domain." IScience 22 (2019): 166-180.

54. Stelzer, Gil, et al. "The GeneCards suite: from gene data mining to disease genome sequence analyses." Current protocols in bioinformatics 54.1 (2016): 1-30.

55. Muto, S. et al. Relationship between the structure of SET/TAF-Iβ/INHAT and its histone chaperone activity. Proc Natl Acad Sci U S A 104, 4285–4290 (2007).

56. Stemm-Wolf, A. J., O'Toole, E. T., Sheridan, R. M., Morgan, J. T. & Pearson, C. G. The SON RNA splicing factor is required for intracellular trafficking structures that promote centriole assembly and ciliogenesis. Mol Biol Cell 32, 1–19 (2021).

57. Schmauch, B. et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun 11, 1–15 (2020).

58. He, B. et al. Integrating spatial gene expression and breast tumour morphology via deep learning. Nat Biomed Eng 4, 827–834 (2020).

59. Zhang, X., Wang, X., Shivashankar, G. V. & Uhler, C. Graph-based autoencoder integrates spatial transcriptomics with chromatin images and identifies joint biomarkers for Alzheimer's disease. Nat Commun 13, 1–17 (2022).

60. Yang, K. D. et al. Multi-domain translation between single-cell imaging and sequencing data using autoencoders. Nat Commun 12, (2021).

61. Onofrey, J. A. et al. Generalizable multi-site training and testing of deep neural networks using image normalization. Proceedings - International Symposium on Biomedical Imaging 2019-April, 348–351 (2019).

62. Germain, P. L., Robinson, M. D., Lun, A., Garcia Meixide, C. & Macnair, W. Doublet identification in single-cell sequencing data using scDblFinder. F1000Res 10, 1–26 (2022).

63. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 2017-Janua, 2261–2269 (2017).

64. Kim, G. et al. Rapid Species Identification of Pathogenic Bacteria from a Minute Quantity Exploiting Three-Dimensional Quantitative Phase Imaging and Artificial Neural Network. Light Sci Appl 11, 190 (2022).

65. Selvaraju, R. R. et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. Int J Comput Vis 128, 336–359 (2020).

66. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44, W90–W97 (2016).

67. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. Nat Genet 25, 25–29 (2000).

SUPPLEMENTARY INFORMATION

Supplementary Table 1. Demographic features

	Sepsis (N = 5)
Age, year	75 (74, 88)
Male, n (%)	2 (40%)
Diagnosis	
Biliary septic shock, n (%)	3 (60%)
Pneumonia septic shock, n (%)	1 (20%)
Urinary septic shock, n (%)	1 (20%)
1st time point SOFA*	9.99 (8.00, 12.00)
2nd time point SOFA*	5.00 (1.00, 6.00)
3rd time point SOFA*	0.00 (0.00, 1.50)
Number of total cells	3,064 (CD8 T: 1,639)

* SOFA, Sequential Organ Failure Assessment

Supplementary Table 2. Patient inclusion and exclusion criteria

Criteria Type	No.	Description
Inclusion	1	Patients with suspected infections.
	2	Two or more criteria of the quick Sequential Organ Failure Assessment (qSOFA) were satisfied: - Respiratory rate ≥22/min - Altered mentation - Systolic blood pressure ≤100mmHg
	3	An acute change in total Sequential Organ Failure Assessment (SOFA) score ≥2 points due to the infection.
	4	Patients who were diagnosed with sepsis.
	5	Vasopressor requirement for a mean arterial pressure of 65 mm Hg or greater and a serum lactate level greater than 2 mmol/L in the absence of hypovolemia.
Exclusion	1	Age <19
	2	Pregnant or lactating
	3	Active cancer status
	4	Acute stroke
	5	Acute cardiovascular disease
	6	Acute burns
	7	Acute gastrointestinal bleeding or bleeding within the last three months
	8	Taking immunosuppressive drugs after organ transplantation
	9	Taking immunosuppressive drugs for autoimmune disease
	10	Previously diagnosed immunodeficiency conditions, or CD4 cell counts below 360 G/L
	11	Neutropenia (neutrophils < 500 G/L) or if the neutrophils were 500-1000 G/L due to chemotherapy and were expected to decrease.
	12	Diagnosed with adrenal dysfunction
	13	Prescribed a steroid equivalent to or greater than 0.5 mg/kg/day prednisone
	14	Active tuberculosis
	15	Cystic fibrosis
	16	Post-traumatic
	17	Patients who needed immediate surgery
	18	The state of Do-Not-Resuscitate (DNR)

Supplementary Table 3. ANOVA test results

Gene	Pr_F
N4BP2L2	0.129509937803094
NAA10	0.13515870581716
ACAP1	0.000743568577274568
ANAPC16	0.00028830568630829
ANP32B	0.0286376480494156
ARF1	0.00594224783218777
ORMDL1	0.0605985543234926
ARL4C	0.0123592833211802
ARPC2	3.99792760995731E-06
ARPC4	0.000407321513182995
PABPC1	0.000159561225068408
ARPC5L	0.000274762237626927
PFN1	3.6378469960392E-07
ATP5MC2	9.05061157780184E-05
PNRC2	0.200027135303339
ATP6V1G1	0.000492871018277362
PPP4C	0.0569278120864736
PSMA4	0.219514344439647
PSMA5	0.000167093622389519
PSMB8-AS1	0.266299386517054
CCL4	0.00324487857992835
CCNI	0.00488093087075932
PTP4A2	0.00510379904443045
RCN2	0.0666453126135416
RGS10	0.0614784933426084
CHURC1	0.000279958441199613
RNASEK	4.26076555531962E-05
COMMD6	2.05250954786738E-06
RNH1	0.0580756173405677
RPL12	1.14116276979324E-11
RPL28	0.154640712094851
RPL34	3.72727291315162E-07
COX6A1	0.0358454248538918
RPL35A	2.42629597034733E-10
COX6C	0.000933406866203093
RPS8	4.63196869273584E-10
RPS15A	4.17186333116153E-09

CYFIP2	0.279524841968575
SDF2	0.0576332732943598
SERF2	0.00103113674372893
SET	0.0109857997410685
SIVA1	0.0512051976999063
SKP1	3.71947469323142E-05
FAM162A	0.0982239294286712
FAU	6.97537237102098E-12
FKBP11	7.53715515811205E-05
SMDT1	0.000310074671209055
GIMAP7	9.21290679845216E-08
SNHG6	1.76227496032228E-05
GNG5	0.00030777531816448
GRB2	0.111247556113962
H3F3B	1.27542118083904E-08
SNU13	0.0752842534380501
SON	0.00164106857050651
HCST	0.000583612391180288
SPCS2	0.00943242684913351
SPCS3	0.0928473438347402
SPG7	4.17007389474475E-05
HMGN3	0.177328518985684
SSR4	0.000110938640320023
HNRNPF	0.00393016366985479
HP1BP3	4.71784584995712E-06
HSP90AA1	8.31720585346302E-05
IFITM2	0.000640733443372518
SYNE1	3.54867374568553E-07
TCEA1	0.161650663527425
ITGA4	0.116941888043846
JTB	0.00129525791577293
KLF6	0.00521520358126956
KRT10	0.053074312056001
TMF1	4.02132135831499E-05
LAPTM5	0.000269835992987571
TPM3	0.000748525725028675
LUC7L3	1.22196664086273E-05
MBP	0.0035492954337052
MCL1	0.0547591110035107
MED4	0.0614575198570707

MPHOSPH8	0.0170610342908387
UQCRB	6.38843928049698E-07
VPS13C	2.7621859328919E-05
YWHAB	3.68461133283568E-07
YY1	0.23633082432192
ZNF207	8.51105816950949E-07

Supplementary Table 4. Open sepsis coefficient difference test result

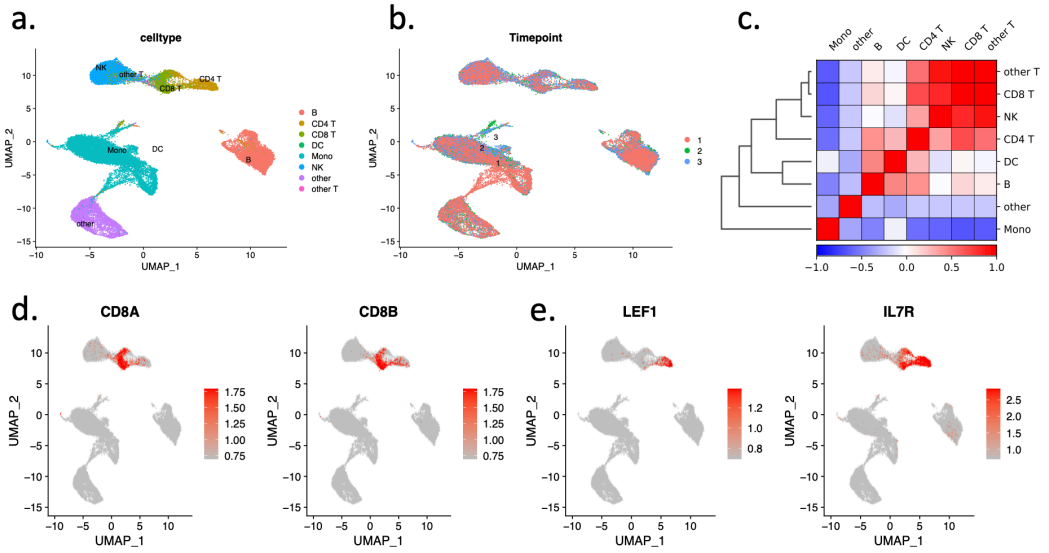
Gene	Cluster	CoefDifference	PValue
FAU	Cluster2	-0.380302136459969	1.37299225050404E-16
RPL12	Cluster2	-0.188913252413975	9.03878704002036E-05
SDF2	Cluster2	0.076172845176035	0.0259847233168203
RPS15A	Cluster2	-0.719141694727681	1.77624432827375E-62
JTB	Cluster2	0.206314086230256	3.92678898641599E-05
PSMA5	Cluster2	0.0751408894975305	0.111794820255433
H3F3B	Cluster2	0.00991695855161434	0.861383049109513
PFN1	Cluster2	-0.78065099771752	3.94556626236616E-49
SMDT1	Cluster2	0.464176013144302	5.13554066311331E-20
RPL34	Cluster2	-0.53282676683676	2.17005776232402E-33
LAPTM5	Cluster2	0.157807557682384	0.00365695395981103
RPL35A	Cluster2	-0.221319321778204	1.33230673971271E-06
RPS8	Cluster2	-0.47416887751202	1.46743015605905E-27
LUC7L3	Cluster6	-0.162367649202875	0.0017741664268277
YY1	Cluster6	-0.0238487061977216	0.608154343745023
TMF1	Cluster6	-0.193490240259582	4.30043931822743E-07
SPG7	Cluster6	-0.167640103584438	2.72988529404362E-07
VPS13C	Cluster6	-0.289171814223167	5.63161959024652E-10
CCL4	Cluster6	-0.34065879481393	3.55198131074374E-07
CYFIP2	Cluster6	-0.118974252907071	0.00573922785181948
KLF6	Cluster6	-0.14141633671375	0.00997839497204598
N4BP2L2	Cluster6	0.224691385992815	5.48652417339986E-05
HP1BP3	Cluster6	-0.378331787970582	5.2945005788179E-13
ZNF207	Cluster6	-0.144277844913994	0.00328153269673204
HMGN3	Cluster6	0.0367932235004628	0.426405211206594
MPHOSPH8	Cluster6	-0.228347330441571	1.54456994716596E-05
SSR4	Cluster6	0.222040259253636	8.99123100792678E-05

SYNE1	Cluster6	-0.455997134996688	4.43351930258497E-17
MBP	Cluster6	-0.267288137051014	1.57950285237774E-06
NAA10	Cluster6	0.0600207221584709	0.114830899445066
PTP4A2	Cluster6	-0.0925345303816842	0.100314896143556
ANP32B	Cluster6	0.107221444046102	0.0372122117495735
TCEA1	Cluster6	-0.0232212284181047	0.61280907913221
TPM3	Cluster6	-0.0479494395104375	0.40344165799353
SET	Cluster6	-0.222282571179688	0.000133564134198286
SON	Cluster6	-0.242001440990336	1.87814197154902E-05
ACAP1	Cluster6	0.0832709561833287	0.107570908058275
COX6A1	Cluster6	0.0346733452914451	0.527555622997396
PNRC2	Cluster6	0.249335272022174	3.24087344906198E-09
SNU13	Cluster6	0.509215707998228	1.82650213424101E-27

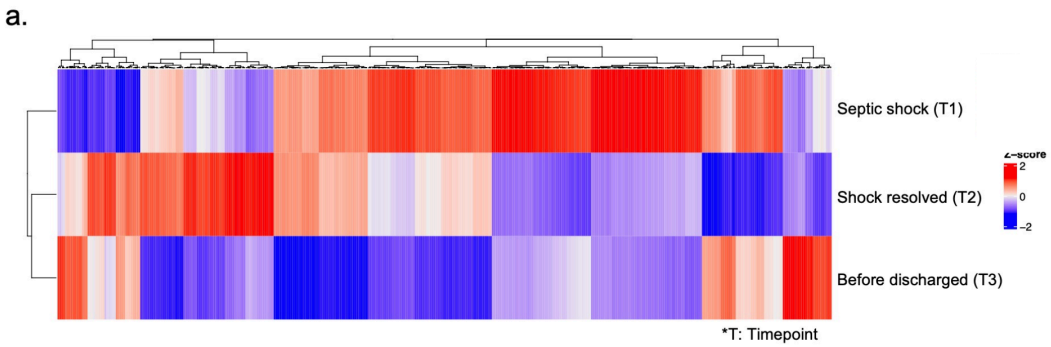
Supplementary Table 5. Open COVID-19 coefficient difference test result

Gene	Cluster	CoefDifference	PValue
FAU	Cluster2	0.299568896652964	3.69876633862359E-12
RPL12	Cluster2	0.315655005118925	2.11403897566331E-12
SDF2	Cluster2	0.0752801759312092	0.0162513217530862
RPS15A	Cluster2	0.312613359776427	1.87846046703982E-17
JTB	Cluster2	0.15825072911182	0.000583292995334726
PSMA5	Cluster2	0.196608511324452	7.06989994271224E-06
H3F3B	Cluster2	0.266516859409179	6.20782779232051E-07
PFN1	Cluster2	0.198132628799943	3.94196082204718E-05
SMDT1	Cluster2	0.191717505571616	3.27108908839848E-05
RPL34	Cluster2	0.298459643208007	9.72025715397895E-14
LAPTM5	Cluster2	0.303463464987329	1.02816023348578E-09
RPL35A	Cluster2	0.333254910677084	5.99423144389779E-15
RPS8	Cluster2	0.275958163341543	2.51684931517112E-12
LUC7L3	Cluster6	-0.242139495775	6.4210400333845E-07
YY1	Cluster6	0.0304587651798358	0.482558255374984
TMF1	Cluster6	-0.150855732846031	2.81091407305835E-05
SPG7	Cluster6	-0.136357729406268	8.92064536643044E-06
VPS13C	Cluster6	-0.210701412975234	1.67909283751104E-06
CCL4	Cluster6	-0.200322673585351	0.000779488908708566
CYFIP2	Cluster6	0.0928541907518546	0.020971592942182
KLF6	Cluster6	-0.218676777710223	1.68665320516071E-05
N4BP2L2	Cluster6	-0.0511740974207463	0.319477591304562
HP1BP3	Cluster6	-0.197880479585744	4.74899016881099E-05
ZNF207	Cluster6	-0.17640485227732	0.000115722831221811
HMGN3	Cluster6	0.129048859447876	0.00191278710848855
MPHOSPH8	Cluster6	-0.121899218329371	0.0138941869662596
SSR4	Cluster6	0.218842517923478	2.7392259968899E-05

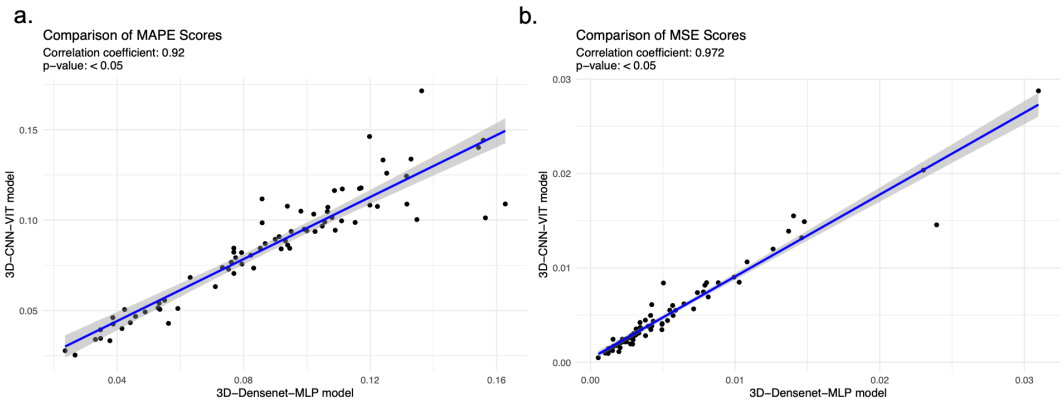
SYNE1	Cluster6	-0.2694062673161	9.67574125567468E-08
MBP	Cluster6	0.176000468324648	0.000522383518805063
NAA10	Cluster6	0.0635588351362411	0.0666056695465555
PTP4A2	Cluster6	0.172564338941341	0.000935195603043461
ANP32B	Cluster6	0.113427707325812	0.0176941934341044
TCEA1	Cluster6	0.0887325317261027	0.0337916548207776
TPM3	Cluster6	0.237459655216204	7.27716698591234E-06
SET	Cluster6	-0.0131731709323251	0.806319541021993
SON	Cluster6	-0.00769841289261294	0.882128953746524
ACAP1	Cluster6	-0.0126675849289535	0.789656916466872
COX6A1	Cluster6	0.152785770126046	0.00246977834878148
PNRC2	Cluster6	0.0742981304013084	0.0625402216368881
SNU13	Cluster6	0.113805857937406	0.00943755154115863



Supplementary Figure 1. Single-cell transcriptional profiling of PBMCs from sepsis patients. a. UMAP representation of cell-type annotated clusters integrated all samples. **b.** The longitudinal proportion of cell clusters within the UMAP. **c.** Dendrogram showing Pearson correlation coefficients between annotated cell types. **d, e.** Feature plot illustrating the expression of T cell specific markers. CD8A and CD8B are known markers of CD8 T cells, while LEF1 and IL7R are recognized as specific markers of CD4 T cells.

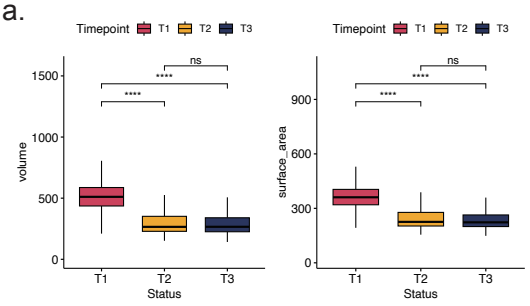


Supplementary Figure 2. Target gene expression pattern in sepsis progression. **a.** Heatmap illustrating the variable expression patterns of the selected 412 target genes in CD8 T cells across different sepsis severity.



Supplementary Figure 3. Comparison of MAPE and MSE score results between current study model and 3D CNN-VIT model. a. Comparison of MAPE and **b.** MSE score, respectively.

a.



Supplementary Figure 4. Cellular morphological features during shock recovery. a. Comparative morphological feature (Volume, surface area) analysis between time points T1, T2 and T3.

ABSTRACT IN KOREAN

딥러닝 접근 방식을 사용한 세균성 패혈증 CD8+ T 세포의 형태와 유전자 발현 간의 상관관계 발견 및 검증

연세대학교 일반대학원

의생명시스템정보학교실

김종현

T 세포의 형태학과 유전자 발현 간의 복잡한 상호작용은 면역 반응에서 중요한 역할을 합니다. 그러나 이러한 상호작용에ㄴ 대한 포괄적인 이해는 여전히 불확실하며, 특히 패혈증과 같은 역동적인 면역 관련 질병의 맥락에서 여전히 어려운 과제입니다. 여기서 우리는 단일 세포 RNA 시퀀싱을 통한 T 세포의 유전자 발현 프로파일과 홀로토모그래피를 통해 얻은 3차원 세포 이미지 간의 연관성을 조사합니다. 패혈증은 CD8 T 세포의 형태 변화가 두드러지는 역동적인 면역 관련 질환입니다. 이 연구는 딥러닝 모델을 활용하여 패혈증 환자의 종단 코호트 내 CD8 T 세포의 관계를 조사하여 근복적인 패턴과 관계를 규명했습니다. 그 결과 CD8 T 세포의 종단적 형태 변화와 높은 연관성을 보이는 형태 특이적 유전자를 확인하였습니다. 또한, 이러한 유전자들은 염색질 구성과 같은 세포 구조와 관련하여 생물학적으로 중요합니다. 형태 특이 유전자들의 임상적 중요성은 공개된 패혈증과 코로나바이러스 2019 (COVID-19) 단일 세포 RNA 시퀀싱 데이터셋을 분석함으로써 검증되었습니다. 질병의 중증도를 일관되게 반영하는 유전자가 확인되어 질병의 중증도와 더불어 형태 특이적 유전자를 필터링할 수 있게

되었습니다. 이러한 접근 방식은 유전자 발현과 세포 형태 사이의 상호 관계에 대한 이해를 깊게 하고, 새로운 진단을 발전시키기 위한 표적으로서 세포 형태의 잠재력을 강조합니다.