



OPEN ACCESS

EDITED BY

Mengmeng Song,
University of California, San Francisco,
United States

REVIEWED BY

Wang Yao,
Shanghai Institute of Aviation Medicine, China
Haoyuan Wang,
Tongji University, China

*CORRESPONDENCE

Panpan Wang

✉ wangpp@jnu.edu.cn

Jieling Weng

✉ jieling_weng@163.com

[†]These authors have contributed equally to this work

RECEIVED 20 August 2025

ACCEPTED 06 October 2025

PUBLISHED 22 October 2025

CITATION

Liu H, Xu Z, Karsidag I, Wang P and Weng J (2025) Immune cell communication networks and memory CD8⁺ T cell signatures sustaining chronic inflammation in COVID-19 and Long COVID. *Front. Immunol.* 16:1689507. doi: 10.3389/fimmu.2025.1689507

COPYRIGHT

© 2025 Liu, Xu, Karsidag, Wang and Weng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](#). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Immune cell communication networks and memory CD8⁺ T cell signatures sustaining chronic inflammation in COVID-19 and Long COVID

Hengrui Liu^{1,2†}, Zewen Xu^{2†}, Ilayda Karsidag³, Panpan Wang^{2,4*} and Jieling Weng^{1*}

¹Department of Pathology, The Second Affiliated Hospital of Guangzhou Medical University, Guangzhou, China, ²San Diego School of Biological Sciences, University of California, San Diego, CA, United States, ³Guangdong Provincial Key Laboratory of Traditional Chinese Medicine Informatization, Guangzhou, China, ⁴Department of Traditional Chinese Medicine, The First Affiliated Hospital of Jinan University, Guangzhou, China

Background: COVID-19, including its post-acute sequelae (Long COVID), is increasingly recognized as involving persistent immune dysregulation and chronic inflammation. Severe and prolonged disease states are often accompanied by sustained cytokine release, immune cell exhaustion, and ongoing cell-cell communication that shapes the inflammatory milieu. Among immune subsets, CD8⁺ T cells play a central role in antiviral defense, yet the molecular mechanisms linking their dysfunction to prolonged inflammation remain incompletely understood.

Methods: We analyzed 73,110 peripheral blood mononuclear cells (PBMCs) from individuals across four disease states (Healthy, Exposed, Infected, and Hospitalized) using single-cell RNA sequencing. Immune cell subsets were annotated, and T cell heterogeneity was profiled. Cytokine and inflammatory scores were calculated to assess immune activation. Differentially expressed genes (DEGs) underwent Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Cell-cell communication was evaluated to map ligand-receptor networks. Additionally, nine machine learning models were trained on a bulk RNA-seq cohort, and the SHapley Additive exPlanations (SHAP) framework was applied to interpret key predictive genes.

Results: Progressive disease severity was associated with a decline in T cell proportions, enrichment of pro-inflammatory myeloid cells, and elevated cytokine expression, particularly IL-32. Memory CD8⁺ T cells showed increased exhaustion and inflammatory scores while maintaining a central position in MHC-I-mediated communication networks. Persistent activation of immune and metabolic pathways, including antigen presentation and oxidative phosphorylation, was observed in prolonged disease states. Seven genes (RPS26, RPS29, RPL36, RPL39, RPS28, RPS21, and CD3E) were identified as strong predictors of chronic immune dysregulation, with the XGBoost model achieving the highest AUC. SHAP analysis confirmed their contributions to disease classification.

Conclusion: This study maps the immune landscape of COVID-19 and Long COVID at single-cell resolution, revealing that persistent immune cell

communication, particularly involving memory CD8⁺ T cells, may sustain chronic inflammation beyond the acute phase. The identified molecular signatures offer potential biomarkers and therapeutic targets for mitigating post-viral inflammatory syndromes.

KEYWORDS

single-cell RNA sequencing, immune cell communication, chronic inflammation, COVID-19, long covid, machine learning, SHAP model

Introduction

Coronavirus Disease 2019 (COVID-19), caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has resulted in over 767 million confirmed cases and more than 6.9 million deaths worldwide as of June 2023 (WHO) (1). Although widespread vaccination has significantly reduced infection rates and mortality, the emergence of new variants with enhanced transmissibility and pathogenicity continues to pose serious challenges, making the pandemic far from fully contained (2–4). The impact of COVID-19 is profound, affecting multiple systems and contributing to a substantial disease burden (5). Its clinical manifestations are highly variable, ranging from asymptomatic or mild infections to severe cases and even death, depending on individual immune responses and other risk factors. Studies have identified age, sex, and comorbidities (e.g., hypertension and diabetes) as key determinants of disease progression (6–8). Common symptoms in asymptomatic or mild cases include fever, dry cough, and fatigue, with relatively short disease durations (9). However, in some patients, the condition may rapidly worsen, leading to severe pneumonia, acute respiratory distress syndrome (ARDS), or multi-organ failure (10, 11).

Based on the severity of the disease, COVID-19 can be classified into four states: Healthy (uninfected individuals), Exposed (contacts not yet diagnosed), Infected (confirmed cases without severe progression), and Hospitalized (severe cases) (12). Exposed individuals often exhibit mild or no symptoms but may carry the virus and contribute to its transmission (13, 14). Infected patients, however, experience more significant immune dysregulation, including elevated cytokine levels and mild lymphopenia (15). Hospitalized patients frequently present with a “cytokine storm”, characterized by marked increases in inflammatory cytokines (e.g., IL-6, IL-8, TNF- α), neutrophilia, lymphopenia, and severe immune dysfunction. This excessive and sustained inflammatory response is increasingly recognized as a key contributor to chronic immune dysregulation, driving disease progression and multi-organ damage in Hospitalized patients (16).

Single-cell RNA sequencing (scRNA-seq), a high-resolution gene expression analysis technology, provides a powerful tool for uncovering the immunological mechanisms of COVID-19 (17, 18). The heterogeneity and dynamic changes in immune cells induced by viral infection are difficult to capture using traditional bulk

analysis methods. By analyzing gene expression at the single-cell level, scRNA-seq allows for a comprehensive exploration of the functional states and molecular characteristics of specific cell subsets. COVID-19 severity is closely linked to immune dysregulation, particularly inflammation and CD8⁺ T cell dysfunction. Studies have shown that during the infection phase, the peripheral blood shows significant reductions in T and B cells, coupled with increases in highly inflammatory monocytes and neutrophils (15, 19, 20). Through scRNA-seq, researchers can analyze the transcriptional dynamics of these immune cells in detail, elucidating the mechanisms of cytokine storm and immune paralysis (21). Furthermore, scRNA-seq enables the identification of specific CD8⁺ T cell subsets, assessing their antiviral capacity and exhaustion status, and offering potential targets for immune restoration interventions (22).

Machine learning models are widely used to analyze complex datasets, particularly high-dimensional data in genomics and transcriptomics. By leveraging diverse algorithmic frameworks and mathematical approaches, these models uncover intricate relationships between patient characteristics and clinical outcomes (23). In clinical applications, models based on the SHAP (SHapley Additive exPlanations) framework are commonly used to quantify the contribution of individual features to predictive outcomes (24, 25). This excessive and sustained inflammatory response is increasingly recognized as a key contributor to chronic immune dysregulation, driving disease progression and multi-organ damage in Hospitalized patients. In this study, we applied SHAP-based models to identify and interpret key molecular markers associated with COVID-19-related immune dysfunction, offering novel insights into inflammatory mechanisms and improving the accuracy of predictive modeling (26).

Materials and methods

Collection of RNA sequencing data

Single-cell RNA sequencing (scRNA-seq) data (27) of PBMCs from COVID-19 patients with varying disease severities were retrieved from Data of 197 patients admitted to Yale New Haven Hospital with COVID-19 between 18 March and 5 May 2020 which

were previously described (28). This dataset includes samples from four groups: Healthy controls, Exposed individuals (close contacts not yet diagnosed), Outpatients (Infected), and Hospitalized patients (Severe), with two samples per group. The validation Bulk RNA-seq transcriptomic cohort was previously described and generated on the GPL24676 Illumina NovaSeq 6000 platform, encompassing 100 COVID-19 samples and 26 non-COVID-19 controls (29).

Unsupervised clustering and cell annotation of scRNA data

The raw unique molecular identifier (UMI) count matrix was processed using the R package Seurat (version 5.1.0) and converted into a Seurat object (30). Cells were filtered based on the following criteria: fewer than 1,000 detected RNA molecules, fewer than 200 or more than 10,000 detected gene features, mitochondrial gene proportions (percent.mt) exceeding 20%, and hemoglobin gene proportions (percent.hb) exceeding 90%. After filtering, 73,110 cells were retained for downstream analysis. The dataset was normalized using the ScaleData() function, and the top 2,000 highly variable genes were identified with FindVariableFeatures(). Principal component analysis (PCA) was performed using RunPCA(), selecting the top 10 principal components for dimensionality reduction. t-SNE and UMAP analyses were then conducted using RunTSNE() and RunUMAP(), respectively (31). Clustering was performed with FindNeighbours() and FindClusters(), with the resolution set to 0.3. Cluster-specific marker genes were identified using FindAllMarkers(), applying thresholds of $|\log_2\text{FC}| > 0.25$ and $p\text{-value} < 0.05$ (32). Finally, cluster marker genes were annotated using the scMayoMap package, integrating scMayoMap Database and lung tissue-specific data (tissue = 'lung') (33, 34).

Analysis of tissue preferences of cell types

To evaluate the distribution preferences of cell types across different tissue states (Healthy, Exposed, Infected, and Hospitalized), the calTissueDist() function from the sscVis package (version 0.1.0) was used to calculate the $R_{o/e}$ ratio, which quantifies differences in cell type distribution among tissues. The statistical significance of these differences was assessed using the chi-squared test (method = "chisq"), and a cell type ratio matrix associated with tissue states was extracted. The results were visualized as a heatmap created with the ComplexHeatmap package (version 2.18.0). Color gradients were used to intuitively represent changes in $R_{o/e}$ values, and symbolic annotations were added to heatmap cells to indicate the degree of preference: "+++" for $R_{o/e} > 1$, "++" for $0.8 < R_{o/e} \leq 1$, "+" for $0.2 \leq R_{o/e} \leq 0.8$, "+/-" for $0 < R_{o/e} < 0.2$, and "-" for $R_{o/e} = 0$. This approach provided a clear visualization of the significant differences in cell type distribution across tissue states (35).

Single-cell gene set scoring and visualization

To assess the activity of the Cytokine and Inflammatory gene sets at the single-cell level, the AddModuleScore() function was applied to calculate module scores for each cell, which were then mapped onto UMAP plots. Using the ggplot2 package (version 3.5.1), pie charts were generated to illustrate the activity proportions of these two gene sets across different cell types. To compare Cytokine Score and Inflammatory Score among different tissue states (Exposed, Infected, Hospitalized, and Healthy), boxplots were created, and statistical significance was evaluated using the Wilcoxon test. Additionally, T cell subpopulations were analyzed for their scores in Cytokine, Exhaustion, Inflammatory, and Regulatory Effector gene sets, with boxplots used to visualize score comparisons among different T cell subtypes (30).

T cell subset annotation and gene set scoring

T cell subpopulations were extracted using the subset function and reclustered into eight clusters. The sctype package (version 1.0) was used to annotate these clusters based on gene set scoring of specific T cell markers, identifying Effector CD8+ T cells, Memory CD4+ T cells, Memory CD8+ T cells, Naive CD4+ T cells, and Naive CD8+ T cells (36, 37).

Single-cell enrichment analysis

To explore the functional characteristics of different T cell subpopulations, differential expression analysis and functional enrichment analysis were performed based on single-cell transcriptomic data. Differentially expressed genes (DEGs) for each T cell subset were calculated using the FindAllMarkers function in the Seurat package, with the top 1,000 genes showing significant upregulation selected based on average fold-change values. The selected DEGs were then subjected to Gene Ontology (GO; biological processes and cellular components) and KEGG pathway enrichment analysis using the compareCluster function in the ClusterProfiler package (38). Gene names were converted to ENTREZ IDs using the org.Hs.eg.db database, and statistically significant enriched terms ($p\text{-value} < 0.05$) were identified. Finally, the enrichment results were visualized as dot plots using the enrichplot package.

Cell-cell communication analysis

This study employed the CellChat R package, a specialized tool for inferring, analyzing, and visualizing cell-cell communication from single-cell RNA sequencing data (39). The analysis focused on T cell subsets, including Effector CD8+ T cells, Memory CD8+ T

cells, Memory CD4⁺ T cells, and Naive CD8⁺ T cells. Overexpressed genes and ligand-receptor pairs were identified to calculate intercellular communication probabilities and infer signaling pathway networks. Visualizations, such as circular and bubble plots, were used to depict communication frequencies and strengths among subsets. Key pathways, particularly MHC-I, were analyzed in-depth, revealing the central role of Memory CD8⁺ T cells in immune modulation. These findings provide critical insights into the mechanisms of T cell-mediated immune regulation and their implications in COVID-19 progression (40).

Single-cell differential analysis

Differential expression analysis was conducted on Memory CD8⁺ T cells to investigate gene expression differences across various pathological states (Healthy, Exposed, Infected, and Hospitalized). The analysis employed the Seurat package and the MAST method, comparing each of the three experimental groups to the Healthy control group. The process involved extracting the Memory CD8⁺ T cell subset, identifying differentially expressed genes (DEGs) using the FindAllMarkers function in Seurat, and filtering genes with adjusted p-values (p_{val_adj}) less than 0.05 and average log fold-change (avg_log2FC) greater than 0. Results were visualized using the ggplot2 package, generating volcano plots, bubble plots, and lollipop plots to depict the distribution, expression proportions, and commonly upregulated genes among groups (41).

Machine learning screening

Nine commonly used machine learning methods were applied to model COVID-19-related data, including Linear Discriminant Analysis (LDA), Flexible Discriminant Analysis (FDA), Logistic Regression, Naive Bayes, Support Vector Machine (SVM), Random Forest, Gradient Boosting Machine (GBM), Mixture Discriminant Analysis (MDA), and XGBoost (42). Through multiple randomized experiments assessing AUC in both training and testing datasets, XGBoost demonstrated the best stability and performance. Consequently, XGBoost was selected as the final diagnostic model (43). The xgboost package was employed to train the model, using 70% of the data for training and 30% for testing. Additionally, the rmda package was utilized for Decision Curve Analysis (DCA) to evaluate the model's net benefits at varying risk thresholds, highlighting its clinical application potential.

SHAP-based diagnostic model interpretation

The shapviz package was employed to analyze SHAP values, providing interpretability for the XGBoost model. Key diagnostic

genes, such as RPS26, RPS29, and RPL41, were identified through feature importance bar plots and beeswarm plots, quantifying their contributions to the prediction results. Furthermore, a force plot was generated for the highest-scoring positive sample, visualizing the positive and negative contributions of individual genes to the prediction outcome.

Results

Single-cell transcriptomic analysis of COVID-19 patients

We obtained peripheral blood mononuclear cells (PBMCs) from eight individuals with confirmed COVID-19, covering a range of disease severities. After quality control procedures, including filtering based on mitochondrial gene content, transcript counts, and the number of detected genes, we retained 73,110 high-quality cells for downstream single-cell RNA sequencing (scRNA-seq) analysis (Figures 1A, B). Uniform Manifold Approximation and Projection (UMAP) embedding revealed 15 distinct cell clusters (Figure 1D). These clusters were assigned to 10 major immune and epithelial cell types based on canonical marker gene expression (Figures 1C, E), including T cells (CD3D, CD3G, CD3E), B cells (MS4A1, CD79A, CD79B), myeloid cells (CD14, LYZ, ITGAM), neutrophils (S100A8, S100A9, CSF3R), macrophages (CD68, CD163, MRC1), secretory cells (SCGB1A1, SCGB3A2), alveolar epithelial type I cells (AGER, HOPX, PDPN), alveolar epithelial type II cells (SFTPC, SFTPB, SFTPA1), plasma cells (MZB1, JCHAIN, TNFRSF17), and basal cells (KRT5, KRT14, TP63).

Characterization of cell types across disease states

UMAP analysis compared T cell distributions across four disease states (Healthy, Exposed, Infected, Hospitalized), revealing a progressive decline in T cell proportions with increasing disease severity (Figure 2A). Analysis of cell-type enrichment ratios ($R_{o/e}$) showed significant enrichment of macrophages, neutrophils, and secretory cells in severe disease states, while T cells were notably depleted in the Infected and Hospitalized groups (Figure 2B). T cell enrichment values exhibited a marked decline from Healthy to Hospitalized states (Figure 2C), underscoring their potential role in disease progression. Cytokine expression analysis identified IL32, LTB, and MIF among the top 10 cytokines highly expressed in pathological conditions, with IL32 showing the highest expression across all groups, suggesting its pivotal role in inflammatory responses (Figure 2D). These findings collectively demonstrate dynamic changes in cell-type composition and highlight their functional implications in disease progression.

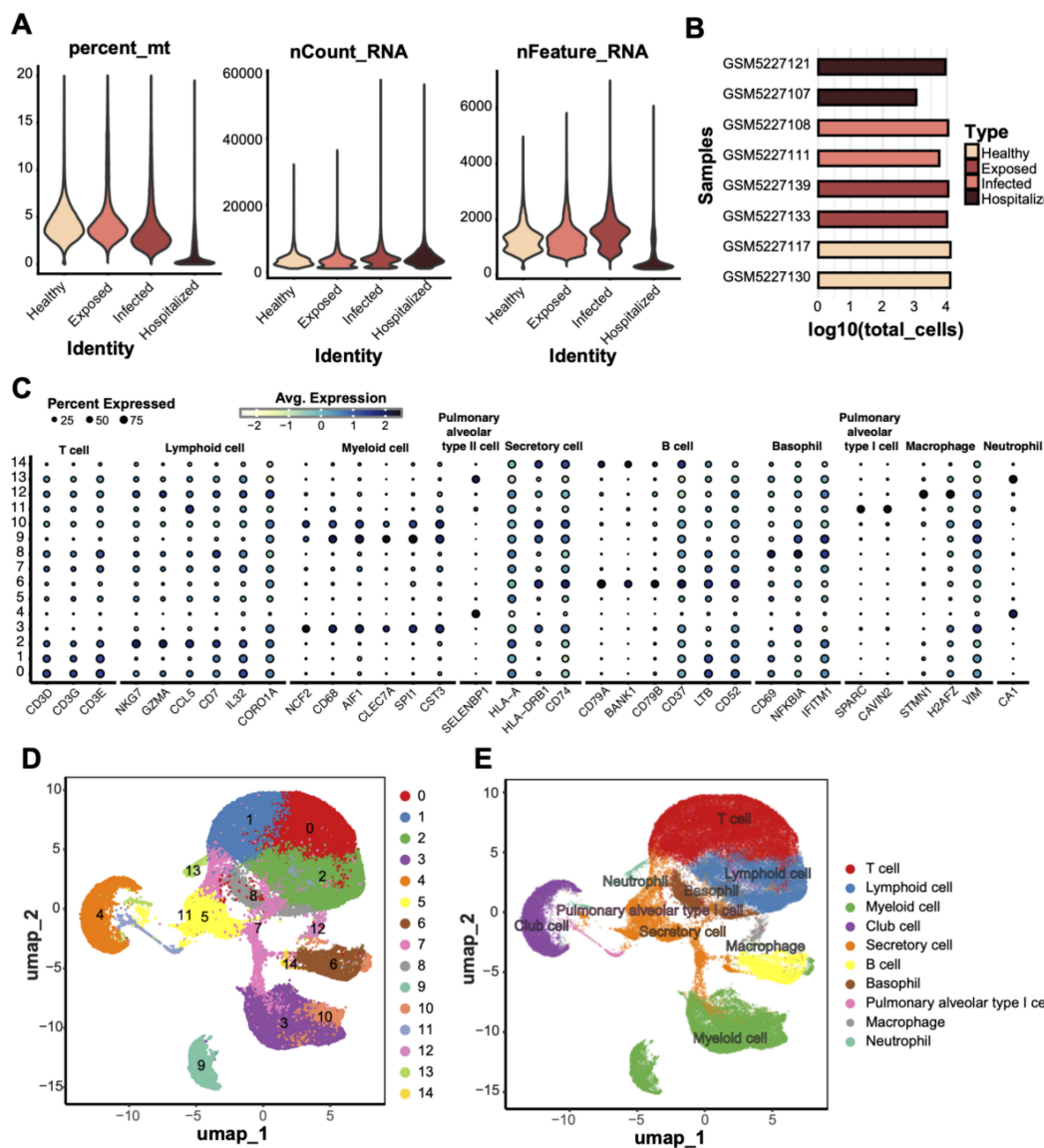


FIGURE 1

scRNA data quality control and cell type annotation. (A) Quality control metrics of scRNA data from four groups (Healthy, Exposed, Infected, Hospitalized), including mitochondrial gene ratio (percent_mt), transcript counts (nCount_RNA), and detected gene numbers (nFeature_RNA). (B) Bar plot displaying the total cell counts (log10 scale) per sample, grouped and compared across the four conditions. (C) Bubble plot showing the expression of characteristic genes across different cell types; bubble size represents the proportion of cells expressing the gene, and color indicates the average expression level. (D) UMAP dimensionality reduction of 14 unsupervised clusters, with each dot representing a cell and colored by cluster ID. (E) UMAP clustering annotated with cell type labels, identifying clusters as specific cell types (e.g., T cells, macrophages, and secretory cells).

Single-cell transcriptomic analysis of cytokine and inflammatory characteristics across cell types

To investigate the role of various cell types in immune regulation and inflammatory responses, we conducted an analysis of cytokine and inflammatory scores. T cells exhibited the highest cytokine scores, constituting 34.7% of the spatial distribution (Figure 3A), underscoring their pivotal role in immune responses. Myeloid cells (27.9%) and lymphocytes (18.3%) also displayed significant cytokine activity, indicating their critical involvement

in the pathological states. Inflammatory score analysis (Figure 3B) revealed that myeloid cells had the highest contribution (35.4%), followed by T cells (31.2%). Comparative analysis across different disease states (Figures 3C, D) showed that T cells' cytokine and inflammatory scores were significantly elevated compared to the Healthy group ($p < 0.05$), with the highest increase observed in the Hospitalized group. Additionally, neutrophils, macrophages, and secretory cells showed a progressive rise in scores from Exposed to Infected and Hospitalized states, suggesting their amplified activation in exacerbating inflammation. Conversely, alveolar epithelial and basal cells demonstrated minimal variations.

T cell subtype clustering and functional scoring

Through UMAP analysis, T cells were subdivided into eight clusters (Figure 4A) and annotated into classical subtypes based on specific markers (Figure 4C), including Effector CD8+ T cells, Memory CD8+ T cells, Naive CD8+ T cells, Memory CD4+ T cells, and Naive CD4+ T cells (Figure 4B). Enrichment analysis across disease groups (R_{o/e} values) revealed a higher proportion of naive T cells in Healthy individuals, while effector and memory T cells were significantly enriched in diseased groups, with Memory CD8+ T cells showing the strongest enrichment in the Hospitalized group (Figure 4D).

Functional scoring of gene sets further elucidated subtype-specific characteristics. Effector CD8+ T cells displayed the highest cytotoxic scores, reflecting their critical killing functions during disease progression (Figures 5A, E). Memory T cells showed elevated exhaustion and inflammatory scores, suggesting functional

impairment during prolonged immune responses (Figures 5B, F). Moreover, regulatory effector scores highlighted the role of Memory CD4+ T cells in immune modulation (Figures 5C, D, G, H). Collectively, these findings emphasize significant functional heterogeneity among T cell subtypes, particularly the roles of effector and memory T cells in immune responses and pathophysiology.

Core role of memory CD8+ T cells in cell communication networks

Interaction strength analysis revealed that Memory CD8+ T cells exhibited the highest interaction intensity with other subtypes, such as Effector CD8+ T cells and Memory CD4+ T cells (Figures 6A, B). Further exploration of MHC-I signaling pathways underscored the critical involvement of Memory CD8+ T cells (Figure 6C). Quantitative analysis of interaction strength

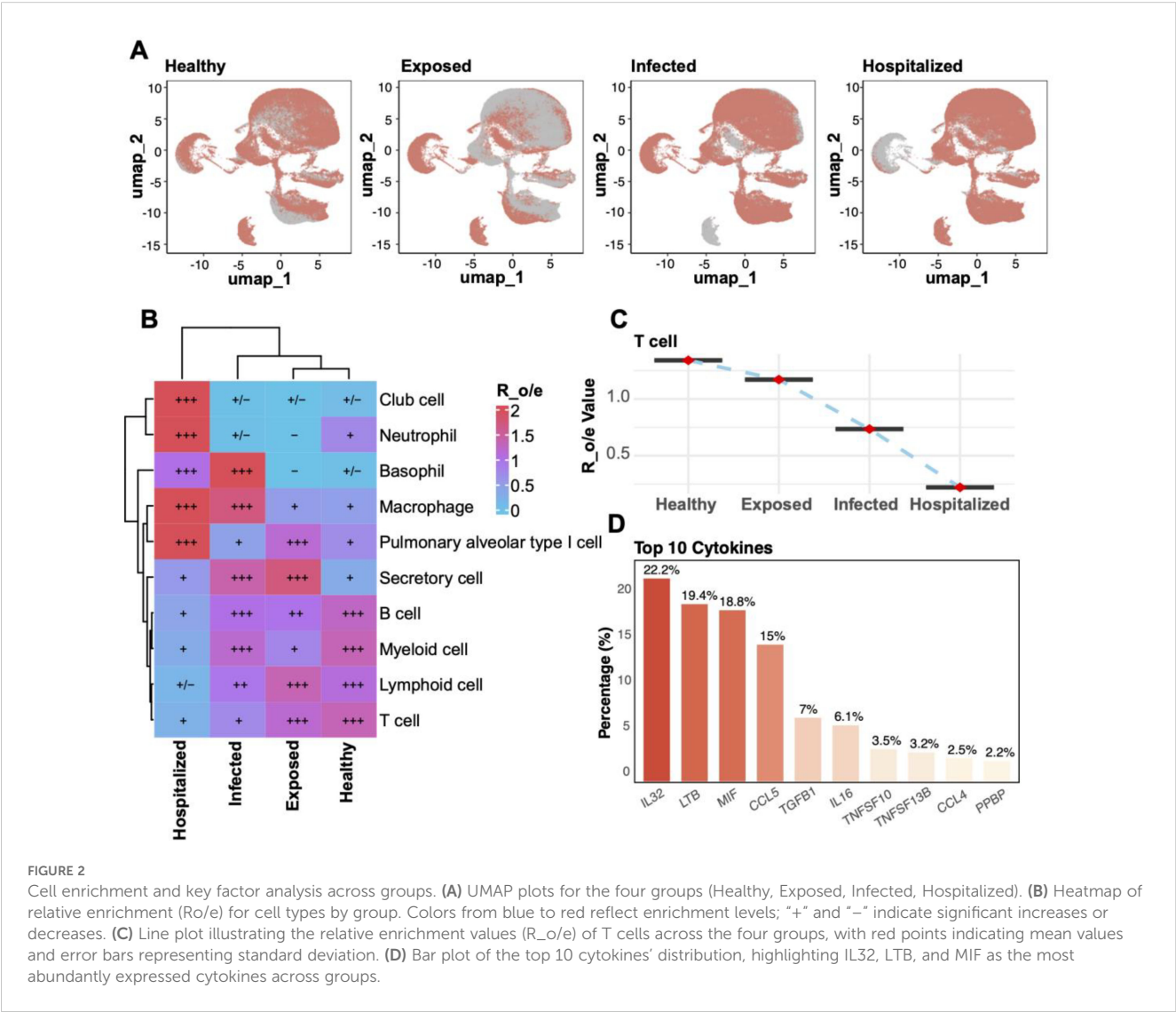


FIGURE 2 Cell enrichment and key factor analysis across groups. (A) UMAP plots for the four groups (Healthy, Exposed, Infected, Hospitalized). (B) Heatmap of relative enrichment (R_{o/e}) for cell types by group. Colors from blue to red reflect enrichment levels; “+” and “-” indicate significant increases or decreases. (C) Line plot illustrating the relative enrichment values (R_{o/e}) of T cells across the four groups, with red points indicating mean values and error bars representing standard deviation. (D) Bar plot of the top 10 cytokines’ distribution, highlighting IL32, LTB, and MIF as the most abundantly expressed cytokines across groups.

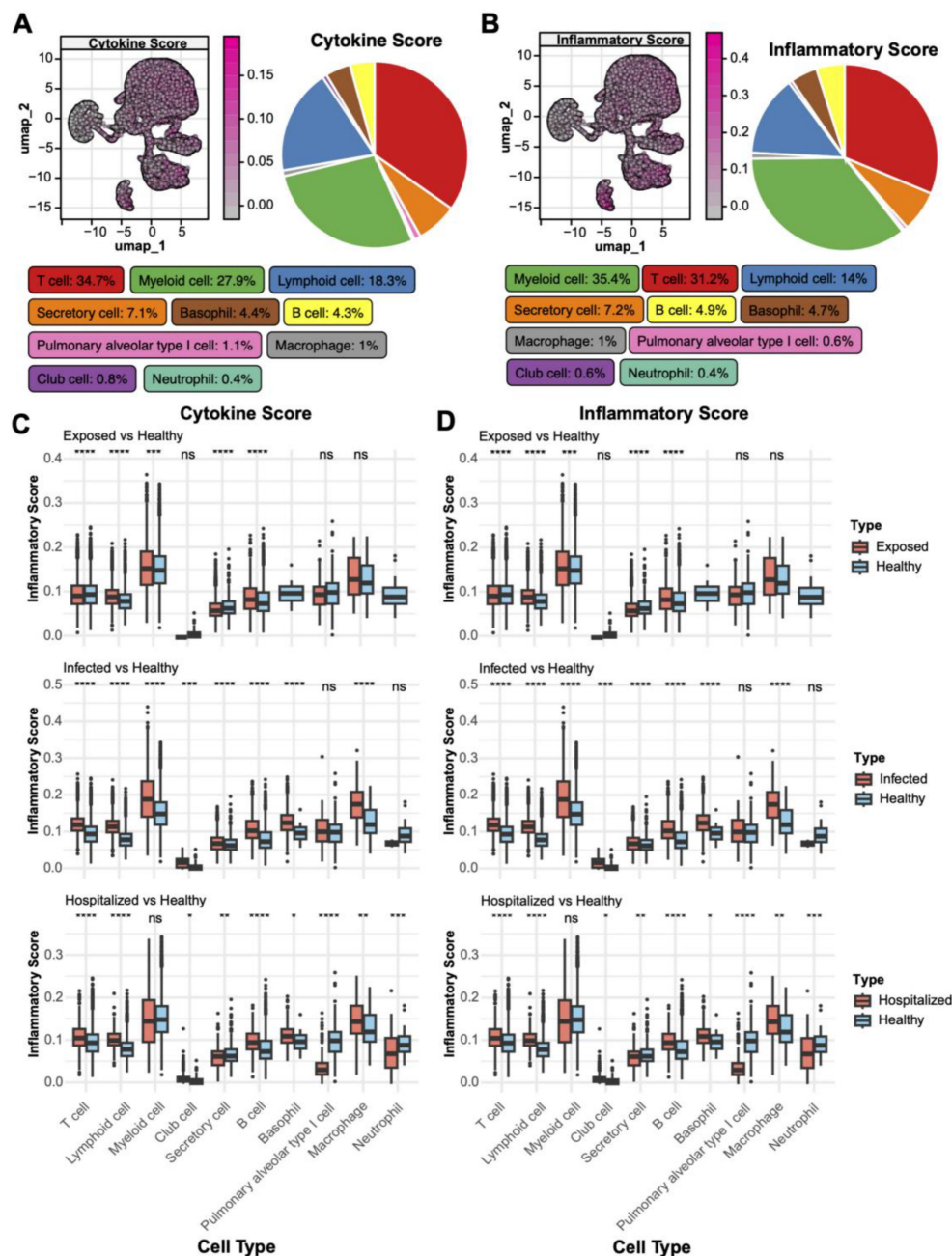
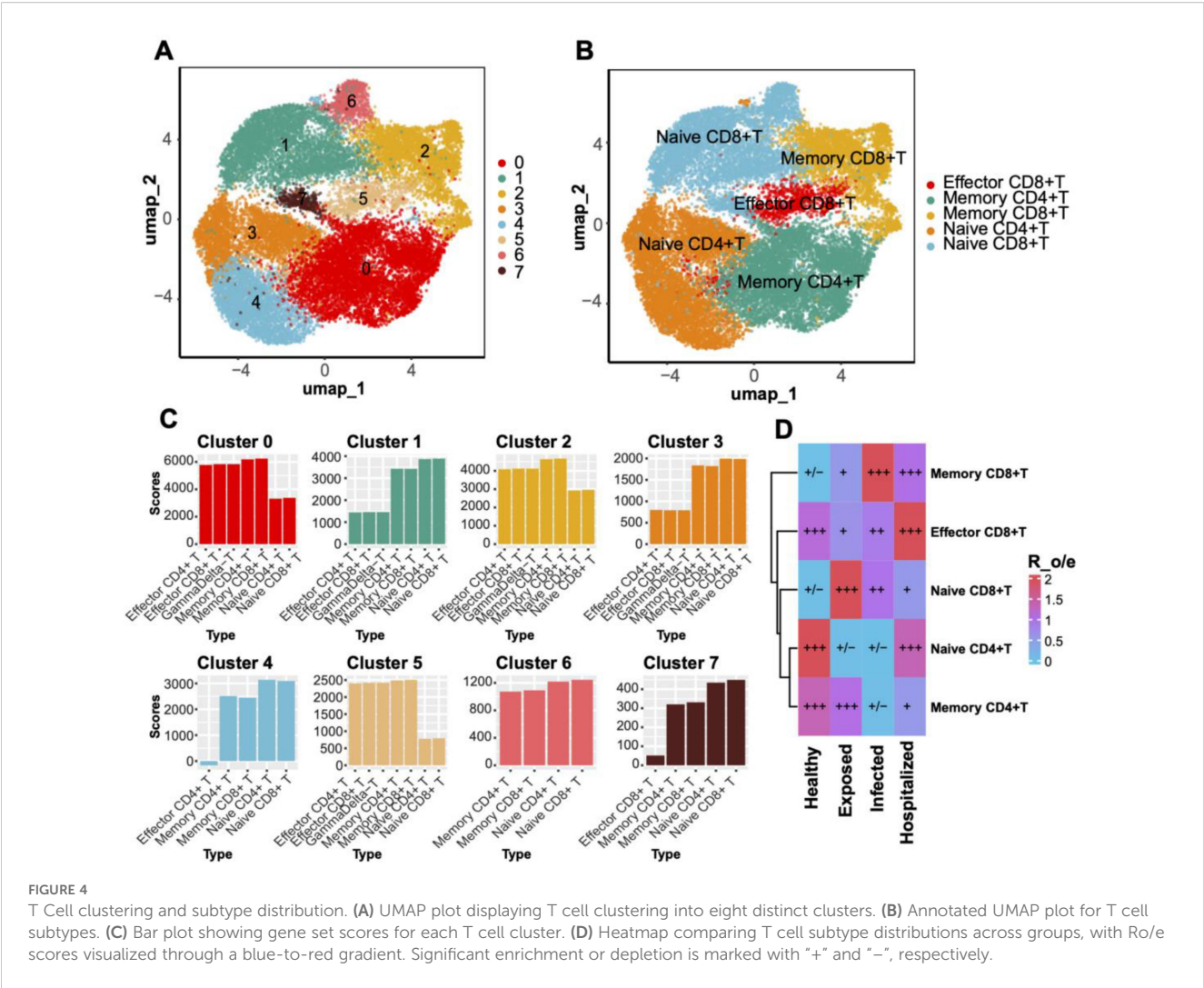


FIGURE 3
Distribution of cytokine and inflammatory scores. (A) UMAP plot of cytokine score spatial distribution, with color gradients indicating score intensity. The pie chart on the right shows the proportion of different cell types within the total population. (B) UMAP plot of inflammatory score distribution with a similar pie chart for inflammatory scores. (C) Boxplots comparing cytokine scores across cell types in Exposed, Infected, and Hospitalized groups versus Healthy controls (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$; ns, not significant). (D) Boxplots showing inflammatory score differences among cell types across the groups.

(Figures 6D, E) highlighted their significant contribution to all pathways, particularly MHC-I signaling. At the transcriptional level, Memory CD8+ T cells demonstrated prominent expression of MHC-I-related genes (e.g., HLA-A, HLA-B, HLA-C) (Figure 6F), reinforcing their pivotal role in antigen presentation and immune response regulation.

Pathway enrichment analysis of memory CD8+ T cells

KEGG and GOBP analyses revealed the crucial functional roles of Memory CD8+ T cells under pathological conditions. In KEGG pathways, these cells were significantly enriched in immune-related



processes such as natural killer cell-mediated cytotoxicity, antigen processing and presentation, chemokine signaling, and COVID-19-related pathways (Figure 7A). Additionally, metabolic pathways like oxidative phosphorylation and ribosome-related pathways were prominently enriched, highlighting their potential role in energy metabolism under active immune states. GOBP analysis further emphasized their involvement in immune regulation, including leukocyte-mediated cytotoxicity, T cell activation, and immune response regulation (Figure 7B). These enriched pathways illustrate the essential functions of Memory CD8+ T cells in antigen recognition, immune responses, and inflammation control.

Differential analysis of single-cell groups

By comparing Exposed, Infected, and Hospitalized groups against the Healthy group, volcano plots depicted significant differential gene expression (Figures 8A–C). The Hospitalized group exhibited the highest number of upregulated genes, indicating that severe pathological states are associated with

substantial transcriptional changes. A cross-group comparison identified seven key genes (RPS26, RPS29, RPL36, RPL39, CD3E, RPS28, RPS21) (Figure 8E). Their expression patterns across groups (Figure 8D) revealed substantial upregulation in the Hospitalized group, suggesting their involvement in immune responses and protein synthesis during disease progression.

Multi-model diagnostics and SHAP-based interpretation

Among nine machine learning models, the XGBoost algorithm demonstrated superior performance in both training and testing datasets, achieving the highest average AUC and stable results (Figures 9A, B). Decision Curve Analysis (DCA) validated the model’s clinical utility, with standardized net benefits significantly surpassing baseline models (“All” and “None”) across various risk thresholds (Figure 9C). SHAP-based analysis identified key predictive genes (e.g., RPS26, RPS29, RPL36), highlighting their significant contributions to disease diagnostics (Figure 9D). Additionally, force

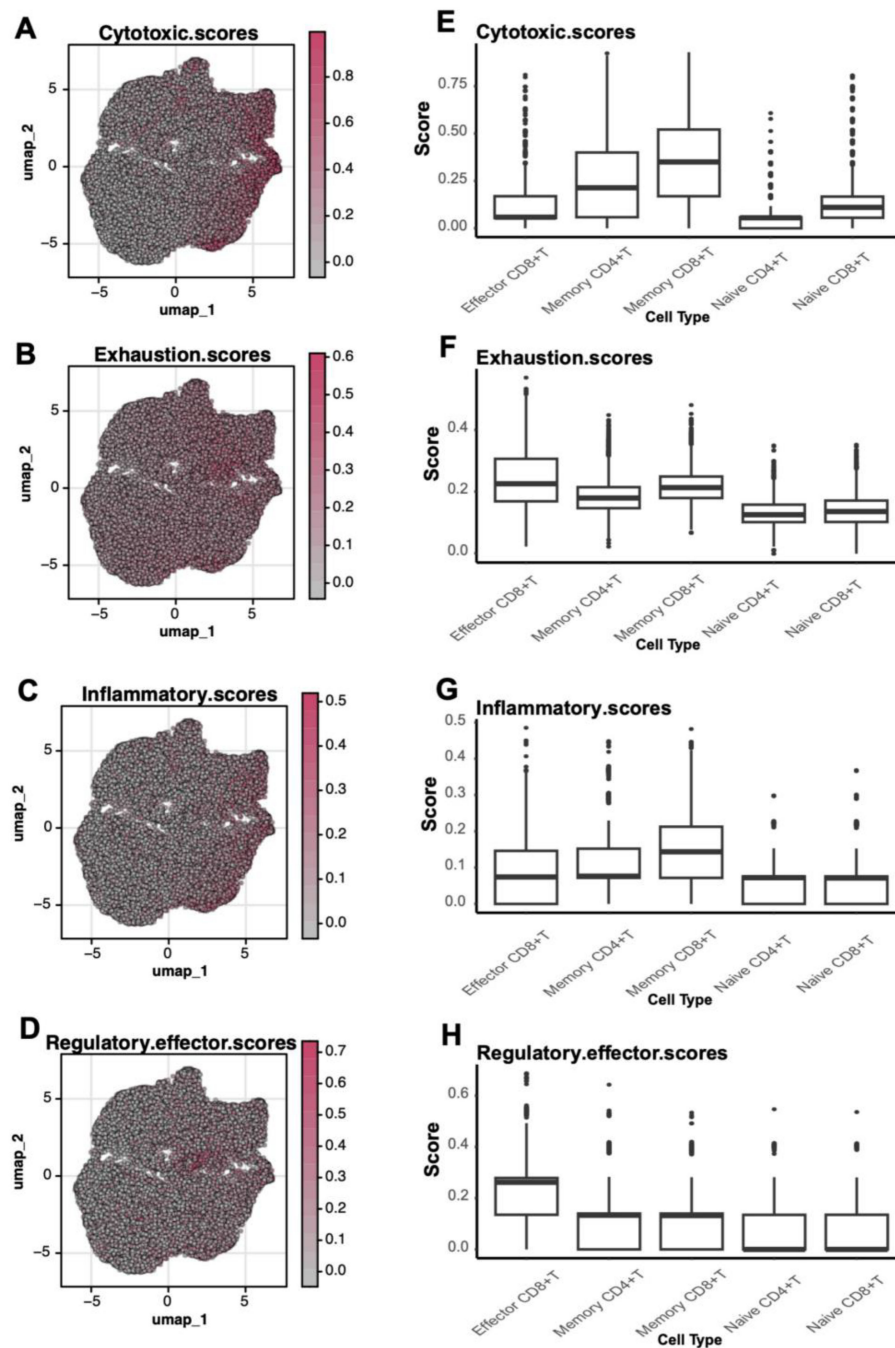


FIGURE 5

Functional scores of T Cells. (A–D) UMAP plots depicting the spatial distribution of cytotoxic (Cytotoxic Scores), exhaustion (Exhaustion Scores), inflammatory (Inflammatory Scores), and regulatory effector (Regulatory Effector Scores) scores in T cells. (E–H) Boxplots comparing functional scores across different T cell subtypes.

plots illustrated how these genes cumulatively influenced the prediction outcomes of specific positive samples (Figure 9E).

Discussion

In this study, we systematically characterized the composition and dynamic changes of peripheral immune cells in COVID-19

patients using single-cell transcriptomic analysis. We profiled 73,110 high-quality PBMCs from eight individuals, constructing immune landscapes across four clinical groups: Healthy, Exposed, Infected, and Hospitalized. UMAP clustering and canonical marker annotation identified ten major immune cell types. Among these, T cells were predominant in Healthy donors but exhibited a significant and progressive decline in Infected and Hospitalized patients, a pattern consistent with prior reports describing T cell

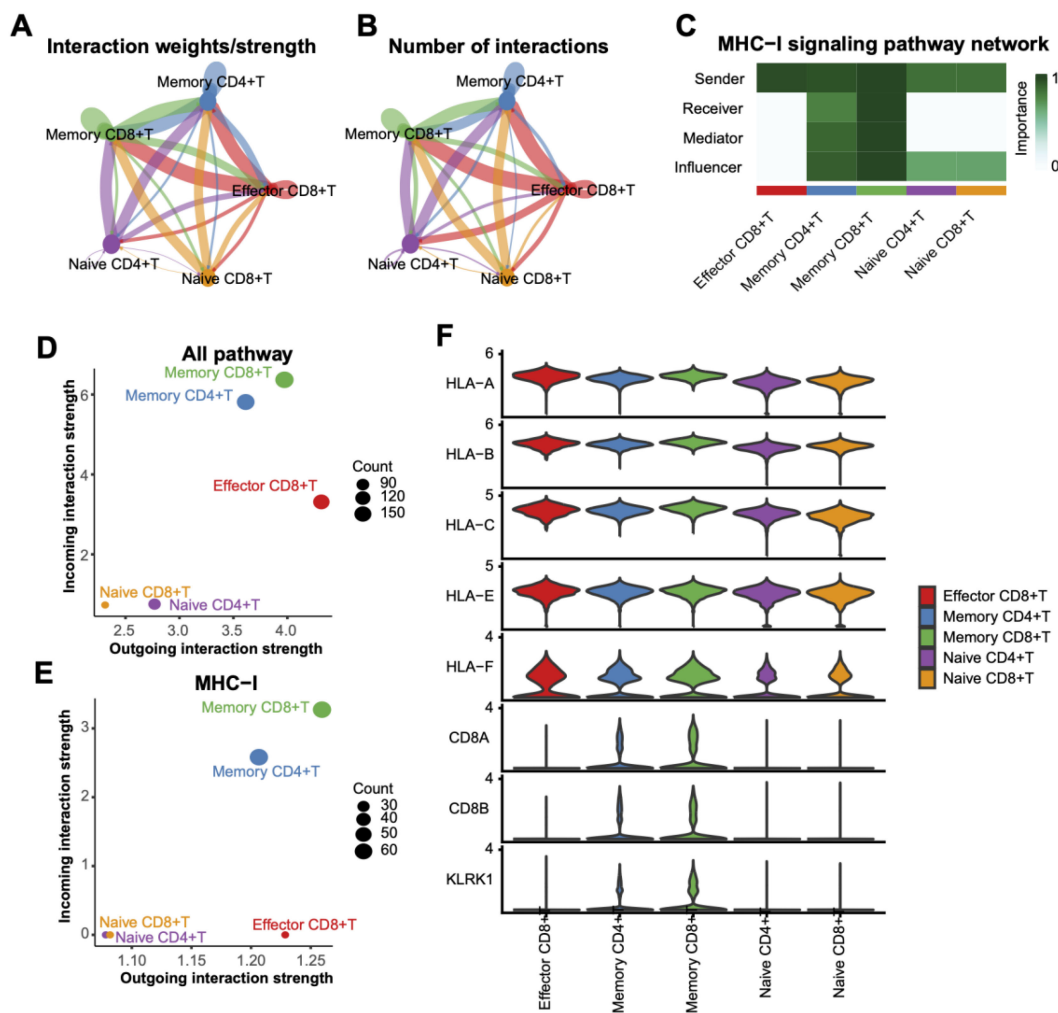


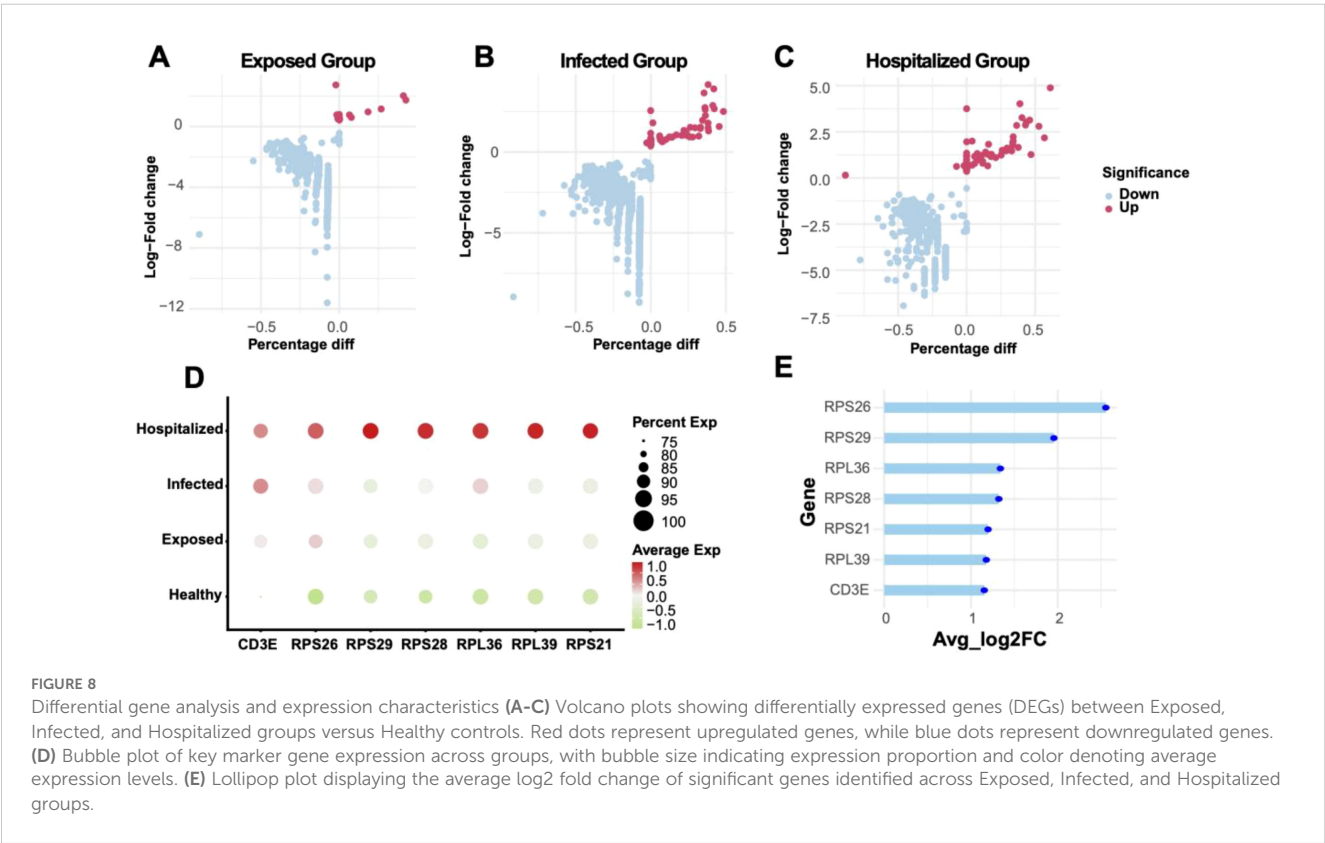
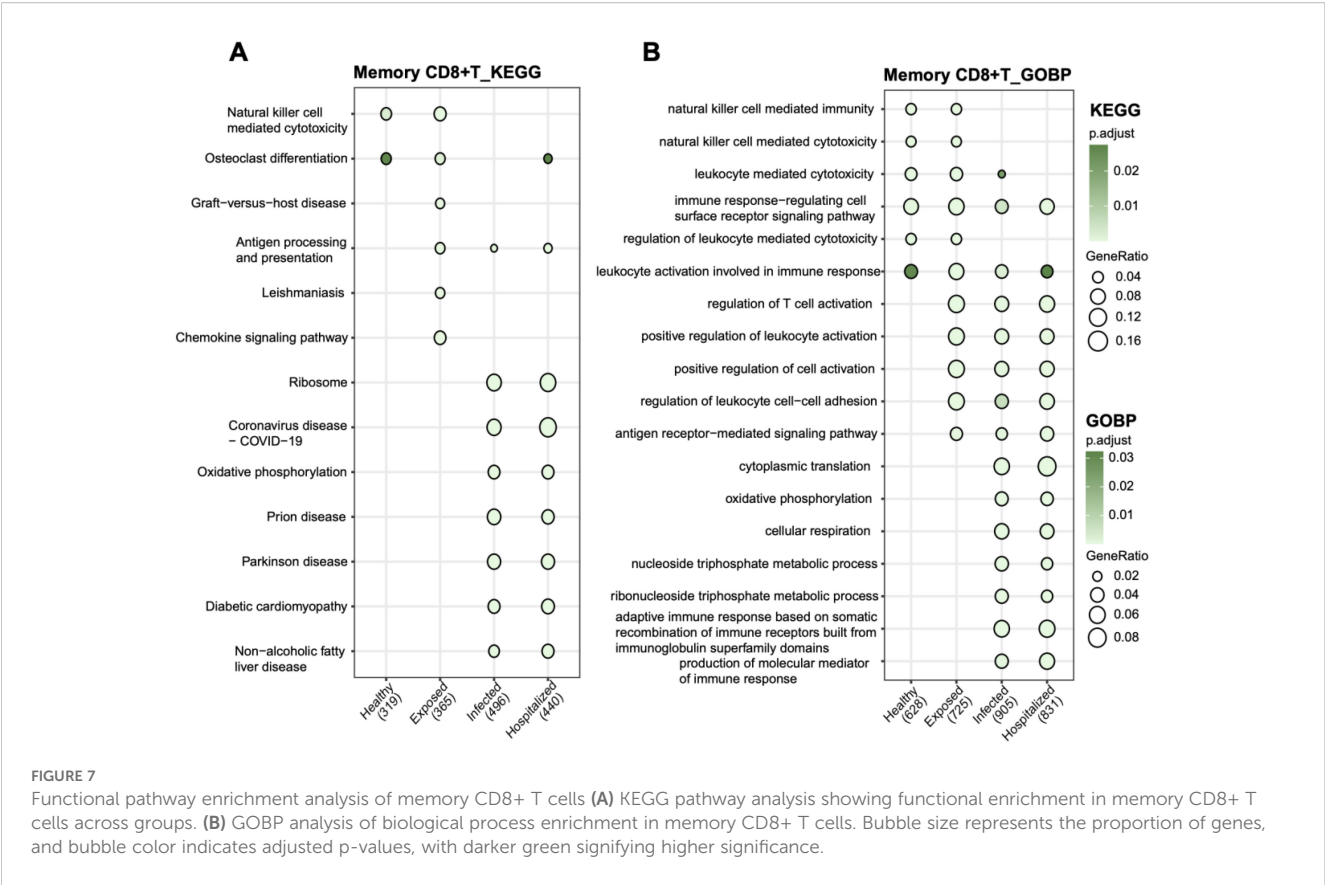
FIGURE 6

Interaction analysis between T Cell subtypes. (A) Interaction weight/intensity network based on all pathways, showing interaction strengths between T cell subtypes. Line color indicates interaction direction, and line thickness represents intensity. (B) Network plot of interaction counts between T cell subtypes. (C) Role-based heatmap of T cell involvement in MHC-I pathway signaling, with communication importance scaled from 0 to 1. (D) Scatterplot of interaction weights across all pathways, with dot sizes representing interaction counts. (E) Distribution of interaction weights based on the MHC-I pathway, highlighting the dominance of memory CD8+ T cells in interactions. (F) Violin plots of MHC-I pathway-associated gene expression levels across T cell subtypes.

lymphopenia and dysfunction in COVID-19 cases (44). Conversely, myeloid cells and neutrophils were significantly enriched with increasing disease severity, indicating a shift from adaptive to innate immune responses. Furthermore, expression levels of pro-inflammatory cytokines such as IL-32, LTB, and MIF were markedly elevated in the severe group, with IL-32 persistently upregulated across all disease stages, reinforcing its known pro-inflammatory role in viral infections and lung pathology (45, 46). Despite elevated cytokine and inflammation scores, T cells likely experience functional exhaustion in severe cases, while continuous myeloid cell accumulation may drive amplified inflammatory cascades characteristic of cytokine storms in critical COVID-19 patients (47, 48).

Detailed functional profiling of T cells revealed subtype-specific activity changes across disease stages. Through clustering and gene-set scoring, T cells were classified into eight subpopulations, including

effector CD8⁺, memory CD8⁺, naïve CD8⁺, and memory CD4⁺ T cells. Effector CD8⁺ T cells exhibited the highest cytotoxicity scores, whereas memory T cell subsets displayed elevated exhaustion and inflammatory scores, suggesting impaired activation during persistent antiviral response, mirroring T cell fatigue phenomena reported by Snyder et al. and others in COVID-19 immunotypes studies (49). Notably, memory CD4⁺ T cells demonstrated strong regulatory signatures, suggesting an immunomodulatory role. CellChat-based intercellular communication analysis identified memory CD8⁺ T cells as central hubs in the MHC-I-mediated antigen presentation network, with strong interactions involving memory CD4⁺ and effector CD8⁺ subsets. Pathway enrichment further revealed that memory CD8⁺ cells are engaged in both immune-related (e.g., antigen processing, NK-mediated cytotoxicity) and metabolic pathways (oxidative phosphorylation and ribosomal activity), supporting the concept of



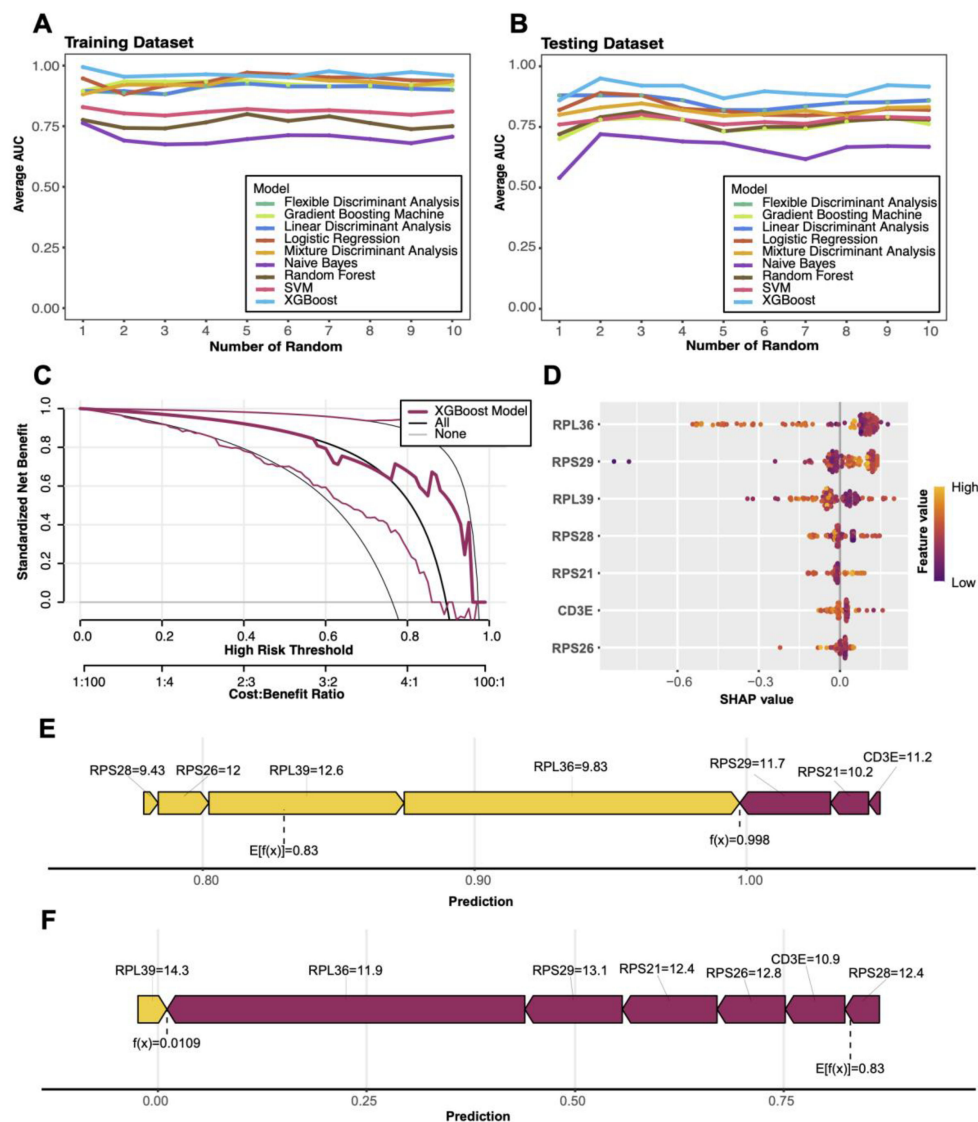


FIGURE 9

Machine learning model performance and SHAP visualizations (A, B) Comparison of average AUC (area under the curve) for nine machine learning models in training and test sets. (C) Decision curve analysis (DCA) evaluating the clinical net benefit of the XGBoost model across different risk thresholds. (D) SHAP-based feature importance plot highlighting key genes contributing to model predictions, with color indicating feature values. (E) Force plot illustrating the predictive mechanism for a positive sample, showing the cumulative contributions of key genes to the model's output. (F) Force plot illustrating the predictive mechanism for a negative sample, showing how key gene features contribute to driving the model output toward a low prediction probability.

immunometabolic coordination in activated T cells during viral infection (20).

In addition, through multi-omic integration and XGBoost-based diagnostic modeling, we identified several key genes, RPS26, RPS29, RPL36, RPL39, RPS28, RPS21, and CD3E, with high SHAP values, indicating strong associations with COVID-19 immune pathology. These genes predominantly encode ribosomal proteins involved in protein synthesis and cellular metabolism. Notably, RPS26 has been shown to regulate specific mRNA translation and maintain T cell development and homeostasis *in vivo*, with knockout leading to peripheral T cell deficits (50, 51). CD3E, a core component of the T cell receptor complex, is vital for T cell activation and signal transduction, and its downregulation may impair immune

recognition efficiency. These findings not only offer mechanistic insights but also suggest potential therapeutic targets aimed at enhancing T cell function via metabolic or ribosomal modulation. While our study is limited by sample size and lacks direct functional validation, it provides a robust framework for understanding immune reprogramming in COVID-19 and highlights critical cell subsets and molecular markers. Future investigations should prioritize larger cohorts and functional assays to elucidate causal mechanisms linking these genes to T cell dynamics.

Importantly, our findings have direct implications for understanding chronic inflammation in post-viral syndromes such as Long COVID. The persistence of elevated cytokine expression, enrichment of pro-inflammatory myeloid cells, and sustained MHC-I-

mediated interactions involving memory CD8⁺ T cells suggest that the immune system remains in a state of low-grade but continuous activation well beyond the acute infection. This prolonged immune communication network—centered on metabolically active yet functionally exhausted memory CD8⁺ T cells—may drive unresolved tissue inflammation. Notably, comorbidities such as diabetes or hypertension may further exacerbate this immune dysregulation, amplifying chronic inflammatory signaling. Our model aligns with emerging evidence that sustained antigen presentation and maladaptive cross-talk between adaptive and innate immune compartments underlie long-term symptoms. Therapeutically, targeting key communication nodes—through modulation of ligand-receptor pathways or restoration of T cell metabolic balance—may help disrupt this cycle and offer new strategies for managing Long COVID and related inflammatory comorbidities.

Conclusions

This study applied single-cell transcriptomics to map the immune landscape across COVID-19 disease stages, revealing that memory CD8⁺ T cells act as central hubs in sustained immune cell communication networks. These cells, despite signs of functional exhaustion, maintained strong MHC-I-mediated interactions that may perpetuate chronic inflammation, particularly in prolonged disease and Long COVID. Multi-model diagnostic analysis identified seven key genes linked to persistent immune dysregulation, offering potential biomarkers and therapeutic targets. Together, these findings provide mechanistic insight into the maintenance of post-viral inflammatory states and lay the groundwork for precision strategies aimed at restoring immune balance.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE171555>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE157103>.

Ethics statement

The study had been approved by the Yale Human Research Protection Program Institutional Review Board (FWA00002571, protocol ID 2000027690). Written informed consent was obtained from all participating patients and healthcare workers.

Author contributions

HL: Writing – review & editing, Writing – original draft. ZX: Writing – review & editing. IK: Writing – review & editing. PW: Writing – review & editing. JW: Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. PW received funding from the K. C. Wong Education Foundation, the Joint Funds of the National Natural Science Foundation of China (U24A6013), the National Natural Science Foundation of China (81603342), the Guangdong Basic and Applied Basic Research Foundation (2024A1515012948), the Guangzhou Science and Technology Project (2024A03J0154, 2023B01J1004), the Zhang Ronghua Guangdong famous traditional Chinese Medicine inheritance Studio Construction Project (2023).

Acknowledgments

We thank Weifen Chen, Zongxiong Liu, Yaqi Yang, and Bryan Liu for their support.

Conflict of interest

The authors declare that the paper was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2025.1689507/full#supplementary-material>

References

- WHO. T. Weekly epidemiological update on COVID-19–13 June 2023. (2023). <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19-1-september-2023>
- Polack Fernando P, Thomas Stephen J, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *New Engl J Med.* (2020) 383:2603–15. doi: 10.1056/NEJMoa2034577
- Rubin R. COVID-19 vaccines vs variants, determining how much immunity is enough. *JAMA.* (2021) 325:1241–3. doi: 10.1001/jama.2021.3370
- Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of SARS-CoV-2. *Nat Rev Microbiol.* (2023) 21:361–79. doi: 10.1038/s41579-023-00878-2
- Miller IF, Becker AD, Grenfell BT, Metcalf CJE. Disease and healthcare burden of COVID-19 in the United States. *Nat Med.* (2020) 26:1212–7. doi: 10.1038/s41591-020-0952-y
- Goodman KE, Magder LS, Baghdadi JD, Pineles L, Levine AR, Perencevich EN, et al. Impact of sex and metabolic comorbidities on coronavirus disease 2019 (COVID-19) mortality risk across age groups: 66–646 inpatients across 613 U.S. Hospitals. *Clin Infect Dis.* (2020) 73:e4113–23. doi: 10.1093/cid/ciaa1787
- Montani D, Certain MC, Weatherald J, Jais X, Bulifon S, Noel-Savina E, et al. COVID-19 in patients with pulmonary hypertension: A national prospective cohort study. *Am J Respir Crit Care Med.* (2022) 206:573–83. doi: 10.1164/rccm.202112-2761OC
- Wang Y, Guo H, Wang G, Zhai J, Du B. COVID-19 as a trigger for type 1 diabetes. *J Clin Endocrinol Metab.* (2023) 108:2176–83. doi: 10.1210/clinem/dgaa165
- Wen Z, Ablimit A. Comprehensive analysis of scRNA-Seq and bulk RNA-Seq reveals ubiquitin promotes pulmonary fibrosis in chronic pulmonary diseases. *Sci Rep.* (2024) 14:21195. doi: 10.1038/s41598-024-70659-1
- Lamperti M, Gattinoni L. Breathing face down. *Br J Anaesthesia.* (2022) 128:745–7. doi: 10.1016/j.bja.2022.01.024
- Williams GW, Berg NK, Reskallah A, Yuan X, Eltzschig HK. Acute respiratory distress syndrome. *Anesthesiology.* (2021) 134:270–82. doi: 10.1097/aln.0000000000003571
- Tay MZ, Poh CM, Rénia L, MacAry PA, Ng LFP. The trinity of COVID-19: immunity, inflammation and intervention. *Nat Rev Immunol.* (2020) 20:363–74. doi: 10.1038/s41577-020-0311-8
- Haddad A, Janda A, Renk H, Stich M, Frieh P, Kaier K, et al. Long COVID symptoms in exposed and infected children, adolescents and their parents one year after SARS-CoV-2 infection: A prospective observational cohort study. *EBioMedicine.* (2022) 84:104245. doi: 10.1016/j.ebiom.2022.104245
- He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat Med.* (2020) 26:672–5. doi: 10.1038/s41591-020-0869-5
- Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest.* (2020) 130:2620–9. doi: 10.1172/jci137244
- Mehta P, McAuley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ. COVID-19: consider cytokine storm syndromes and immunosuppression. *Lancet.* (2020) 395:1033–4. doi: 10.1016/S0140-6736(20)30628-0
- Zhang J-Y, Wang X-M, Xing X, Xu Z, Zhang C, Song J-W, et al. Single-cell landscape of immunological responses in patients with COVID-19. *Nat Immunol.* (2020) 21:1107–18. doi: 10.1038/s41590-020-0762-x
- Ren X, Wen W, Fan X, Hou W, Su B, Cai P, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell.* (2021) 184:5838. doi: 10.1016/j.cell.2021.10.023
- Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, et al. Reduction and functional exhaustion of T cells in patients with coronavirus disease 2019 (COVID-19). *Front Immunol.* (2020) 11:827. doi: 10.3389/fimmu.2020.00827
- Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat Med.* (2020) 26:1070–6. doi: 10.1038/s41591-020-0944-y
- Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.* (2020) 6:31. doi: 10.1038/s41421-020-0168-9
- Bange EM, Han NA, Wileyto P, Kim JY, Gouma S, Robinson J, et al. CD8+ T cells contribute to survival in patients with COVID-19 and hematologic cancer. *Nat Med.* (2021) 27:1280–9. doi: 10.1038/s41591-021-01386-7
- Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. *Nat Rev Genet.* (2015) 16:321–32. doi: 10.1038/nrg3920
- Lundberg S, Lee S-I. A unified approach to interpreting model predictions. Long Beach, CA, USA. (2017). doi: 10.48550/arXiv.1705.07874
- Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell.* (2020) 2:56–67. doi: 10.1038/s42256-019-0138-9
- Islam MM, Karray F, Alhaji R, & Zeng, J. A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19). *IEEE Access.* (2021) 9:30551–72. doi: 10.1109/access.2021.3058537
- Wang EY, Mao T, Klein J, Dai Y, Huck JD, Jaycox JR, et al. Diverse functional autoantibodies in patients with COVID-19. *Nature.* (2021) 595:283–8. doi: 10.1038/s41586-021-03631-y
- Lucas C, Wong P, Klein J, Castro TBR, Silva J, Sundaram M, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature.* (2020) 584:463–9. doi: 10.1038/s41586-020-2588-y
- Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al. Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst.* (2021) 12:23–40.e27. doi: 10.1016/j.cels.2020.10.003
- Hao Y, Hao S, Andersen-Nissen E, Mauck WM, III, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* (2021) 184:3573–3587.e3529. doi: 10.1016/j.cell.2021.04.048
- Lueck MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol.* (2019) 15:e8746. doi: 10.15252/msb.20188746
- Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, III, et al. Comprehensive integration of single-cell data. *Cell.* (2019) 177:1888–1902.e1821. doi: 10.1016/j.cell.2019.05.031
- Chen Y, Wang X. Evaluation of efficiency prediction algorithms and development of ensemble model for CRISPR/Cas9 gRNA selection. *Bioinformatics.* (2022) 38:5175–81. doi: 10.1093/bioinformatics/btac681
- Yang L, et al. Single-cell Mayo Map (scMayoMap): an easy-to-use tool for cell type annotation in single-cell RNA-sequencing data analysis. *BMC Biol.* (2023) 21:223. doi: 10.1186/s12915-023-01728-6
- Zheng L, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science.* (2021) 374:abe6474. doi: 10.1126/science.abe6474
- Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nat Commun.* (2022) 13:1246. doi: 10.1038/s41467-022-28803-w
- Szabo PA, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun.* (2019) 10:4706. doi: 10.1038/s41467-019-12464-3
- Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A J Integr Biol.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118
- Jin S, et al. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* (2021) 12:1088. doi: 10.1038/s41467-021-21246-9
- Neefjes J, Jongstra MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol.* (2011) 11:823–36. doi: 10.1038/nri3084
- Finak G, et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* (2015) 16:278. doi: 10.1186/s13059-015-0844-5
- Kuhn M. Building predictive models in R using the caret package. *J Stat Softw.* (2008) 28:1–26. doi: 10.18637/jss.v028.i05
- Chen T, Guestrin C. (2016)., in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Knowledge Discovery and Data Mining]*. pp. 785–94. San Francisco, California, USA: Association for Computing Machinery.
- Mathew D, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science.* (2020) 369:eabc8511. doi: 10.1126/science.abc8511
- Chen Z, John Wherry E. T cell responses in patients with COVID-19. *Nat Rev Immunol.* (2020) 20:529–36. doi: 10.1038/s41577-020-0402-6
- Bergantini L, et al. Cytokine profiles in the detection of severe lung involvement in hospitalized patients with COVID-19: The IL-8/IL-32 axis. *Cytokine.* (2022) 151:155804. doi: 10.1016/j.cyto.2022.155804
- Xu G, et al. The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. *Cell Discov.* (2020) 6:73. doi: 10.1038/s41421-020-00225-2
- Giamarellos-Bourboulis EJ, et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe.* (2020) 27:992–1000.e1003. doi: 10.1016/j.chom.2020.04.009
- Snyder TM, et al. Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. *medRxiv.* (2020) 15:1488860. doi: 10.1101/2020.07.31.20165647
- Chen C, et al. Ribosomal protein S26 serves as a checkpoint of T-cell survival and homeostasis in a p53-dependent manner. *Cell Mol Immunol.* (2021) 18:1844–6. doi: 10.1038/s41423-021-00699-4
- Piantanida N, et al. Deficiency of ribosomal protein S26, which is mutated in a subset of patients with Diamond Blackfan anemia, impairs erythroid differentiation. *Front Genet.* (2022) 13:1045236. doi: 10.3389/fgene.2022.1045236