IVYSPRING INTERNATIONAL PUBLISHER

Journal of Cancer

2026; 17(3): 469-482. doi: 10.7150/jca.121954

Research Paper

Mapping Myeloid Cell Diversity in Diffuse Large B-Cell Lymphoma: Impact on T Cell Exhaustion and Clinical Prognosis

Jingwen Wang#, Liangcheng Lv#, Zhenjun Li, Xiaoyu Yao, Ning Ding✉

Key laboratory of Carcinogenesis and Translational Research (Ministry of Education), Laboratory of Lymphoma Translational Research, Peking University Cancer Hospital & Institute. No. 52 Fucheng Rd, Haidian District, Beijing, 100142, China.

#These authors contributed equally: Jingwen Wang and Liangcheng Lv.

✉ Corresponding author: Ning Ding (ningding@bjmu.edu.cn).

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/). See https://ivyspring.com/terms for full terms and conditions.

Received: 2025.07.18; Accepted: 2025.12.01; Published: 2026.01.23

Abstract

Diffuse large B-cell lymphoma (DLBCL) is an aggressive subtype of B-cell lymphoma characterized by genetic variability and clinical heterogeneity. Single-cell sequencing technology enables mapping of intra-tumoral heterogeneity and microenvironment interactions. In this study, we analyzed single-cell and RNA expression microarray data from over 3,000 DLBCL patients to investigate the immune landscape of the tumor microenvironment and its association with clinical prognosis. Malignant B cells identified through B-cell receptor (BCR) clonal analysis and copy number variation (CNV) assessment exhibited enrichment in pathways related to the cell cycle, DNA replication and p53 signaling, which were closely related to adverse survival outcomes. Next, the myeloid cells derived from DLBCL tumor tissues could be further clustered into several distinct types, primarily comprising dendritic cells and macrophages. The increased prevalence of *SPP1*$^+$ macrophages within the tumor microenvironment was correlated with inferior overall survival. Additionally, CellChat analysis revealed that frequent interactions between *SPP1*$^+$ macrophages and CD8$^+$ T cells may contribute to T cell exhaustion and create an immunosuppressive microenvironment. Collectively, the diverse sub-populations, particularly the immunosuppressive *SPP1*$^+$ macrophages regulated immune suppression status within tumor microenvironment and represented a potential therapeutic target for DLBCL patients.

Keywords: diffuse large B-cell lymphoma, tumor microenvironment, immune cell heterogeneity

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the predominant subtype of B-cell lymphoma, accounting for approximately 30-40% of non-Hodgkin lymphomas (NHL)[1,2]. Despite advancements in clinical response and overall survival due to the incorporation of the anti-CD20 antibody rituximab alongside chemotherapy, approximately 30-40% of DLBCL patients continued to exhibit resistance to this combination immunotherapy[3]. Consequently, there is an urgent need to identify novel therapeutic strategies and molecular classifications of DLBCL to facilitate precise clinical management for these recurrent/ refractory patients.

DLBCL can be categorized into two molecular subgroups, germinal center B-cell-like (GCB) and activated B-cell-like (ABC), using mRNA expression profiling (GEP)[4]. These subgroups showed different prognoses and responses to standard R-CHOP immunotherapy[5]. GCB-DLBCL and ABC-DLBCL made up about 40% and 50% of DLBCL cases respectively, with a small unclassifiable group comprising 10-15%[4,6]. Based on omics data analysis, the novel molecular DLBCL classification had recently been developed. The four genetic categories had been identified as MCD, BN2, N1, and EZB subtypes in DLBCL patients by oncogenic mutations, including

MYD88, *CD79B*, *NOTCH1* and *NOTCH2* and *EZH2*, which improved the treatment strategy and prognosis of DLBCL patients[7].

The biology and clinical behavior of DLBCL were influenced not only by the molecular alterations within the DLBCL tumor cells themselves but also by their interactions with the surrounding microenvironment. Evidence from both lymphoma patients and animal models suggested that within the lymphoma niche, external stimuli provided by microenvironmental cells and the extracellular matrix (ECM) played a significant role in disease development, progression, and response to treatment[8–10]. Consequently, elucidating the precise cellular and molecular mechanisms that facilitated tumor immune evasion remained a critical objective for enhancing current immunotherapies or developing novel therapeutic strategies. Based on the microenvironmental gene expression profiles, stromal-1 and stromal-2 signature had been identified to be related to the survival outcomes of DLBCL patients. The stromal-1 signature characterized by extracellular matrix deposition and histiocytic infiltration was associated with favorable survival in DLBCL. In contrast the stromal-2 signature marked by angiogenesis correlated with poor clinical outcomes[8]. Recently four distinct lymphoma microenvironment classification were found by transcription gene analysis, including "germinal center-like", "mesenchymal", "inflammatory" and "depleted" subtypes. Furthermore, patients with GC-like lymphoma microenvironment responded best to R-CHOP chemoimmunotherapy and had the highest survival rates, while those with "depleted" lymphoma microenvironment had the lowest response and survival rates, which indicated the significant prognostic value of tumor microenvronement[9]. Given the increasing evidence that the prognostic value of the TME was significant and independent of current standards, this area is of clear pathophysiological importance.

Given the controversies surrounding the latent role of the tumor microenvironment (TME) in the pathogenesis and progression as reported in several studies, recent research employing high-dimensional, single-cell analyses of primary tumors had significantly advanced our understanding of cancer biology[11]. The detailed characterization of cellular heterogeneity in DLBCL at the single-cell level hold promise for the development of more effective molecularly targeted therapies and prognostic biomarkers. In this study, we utilized single-cell RNA sequencing (scRNA-seq) and RNA expression microarray data from existing literature to construct a comprehensive cellular atlas of both malignant and non-malignant cells in DLBCL. This approach allowed us to investigate the impact of key signatures of malignant B cells and infiltrating immune cells on the pathogenesis and prognosis of DLBCL, potentially informing the development of novel molecularly targeted therapies and immunotherapeutic strategies.

Results

Generation of a single-cell atlas for DLBCL

The single-cell RNA sequencing data derived from tumor tissues of 12 samples diagnosed with DLBCL (GSA-Human HRA002297 and GEO GSE182436) were reclustered and analyzed. A total of 54,199 high-quality single-cell transcriptomes from an initial 59,066 cells were analyzed, with an average of 2,378 genes detected per cell. T-cell receptor (TCR) and B-cell receptor (BCR) sequences were identified in 15,037 and 12,428 cells, respectively. Nonlinear dimensionality reduction was conducted using the Uniform Manifold Approximation and Projection (UMAP) method. Based on canonical marker gene expression, we discovered 4 different major clusters: B lymphocytes, myeloid cells, $CD4^+$ T cells and $CD8^+$ T cells (Figure 1A). The violin plot illustrated the principal genes for distinguishing the four major clusters (Figure 1B-C). B cells displayed a higher expression of *CD19*, *MS4A1* and *CD79A*. A higher expression of *CD33*, *TYROBP* and *ITGAX* was showed in myeloid cells. The expression level of *CD3D*, *CD3E*, and *CD4* was found to be markedly elevated in $CD4^+$ T cells. *CD3D*, *CD3E* and *CD8A* genes were found to be highly expressed in $CD8^+$ T cells. As illustrated in Figure 1D, the stacked bar chart depicted the four cell clusters in 12 DLBCL samples. Significant variation was observed in the proportions of these cell clusters among samples. B cells were the most abundant cell type presented in the DLBCL tumor tissues, with infiltrating T cells and myeloid cells being the next-most abundant.

Heterogeneity of malignant B cell predicted the DLBCL prognosis

Malignant B cells and normal B cells were further identified from 12 DLBCL samples to study tumor heterogeneity in DLBCL. UMAP analysis was conducted to display the clustering of B cells belonging to each individual patient (Figure 2A). By using BCR clonal analysis and inferCNV assessment, we finally classified B cells into two clusters, consisting of 25,419 malignant B cells and 4,013 tumor-infiltrating normal B cells (Figure 2B, Figure S1A-E). Next, we made a comparison between the malignant and normal B cells of each sample to find significantly differentially expressed genes (Figure

2C). KEGG pathway analysis further revealed the significantly enriched pathways, including cell cycle, DNA replication, and p53 related signal pathways (Figure 2D). We validated the related pathway signatures using GEO cohorts. The Kaplan-Meier plot showed higher expression levels of genes associated with the cell cycle, DNA replication and p53 signal pathway were correlated with survival disadvantage within GEO database GSE32918 (Figure 2E), GSE181063 and GSE31312 (Figure S1).

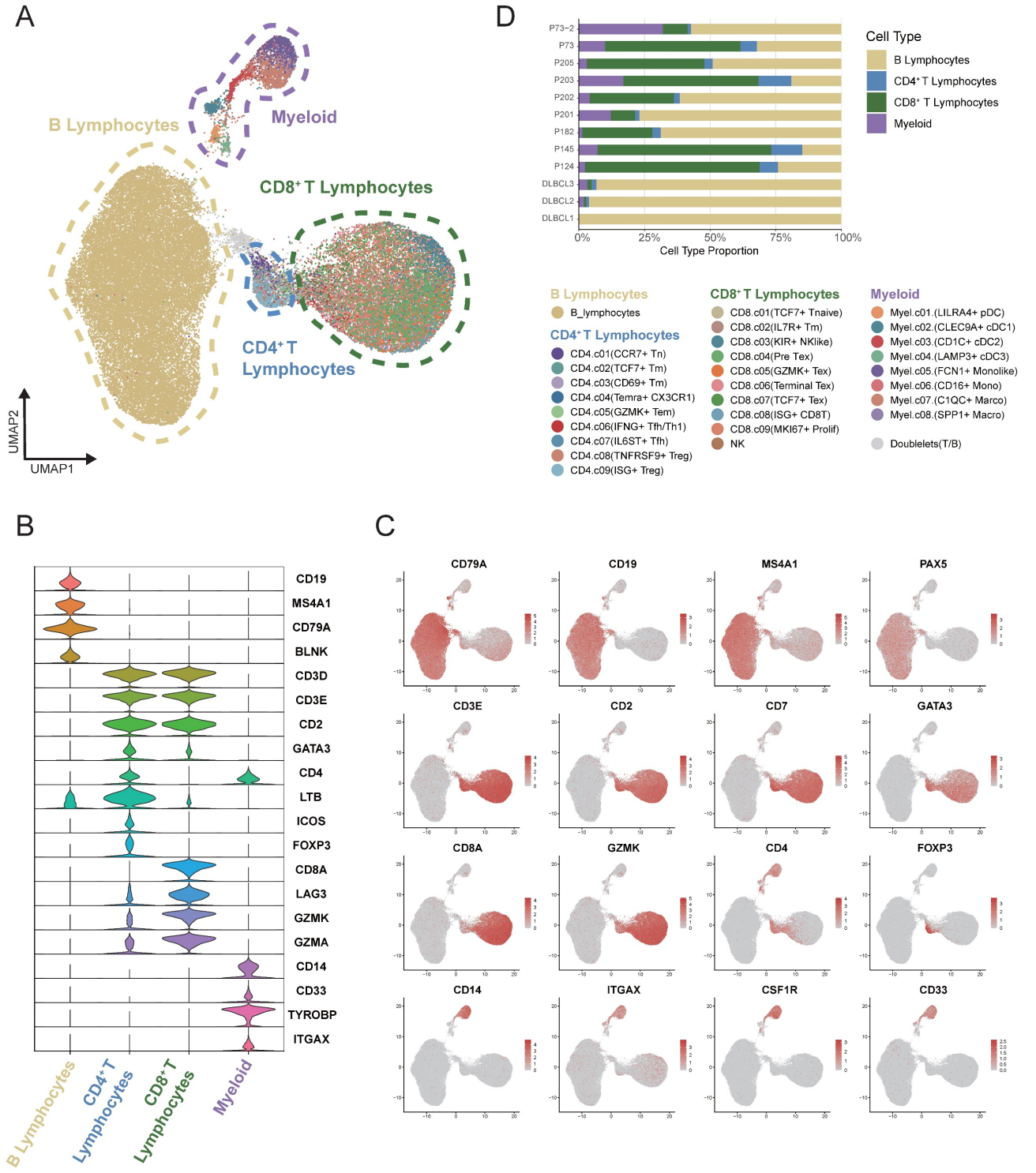

Figure 1. Single-cell analysis revealed the heterogeneity of the tumor microenvironment in the integrated DLBCL samples. (A) UMAP visualization of 54,199 single cells derived from DLBCL tumor tissues. UMAP plot showing the major lineages within tumor samples. Marker genes used for lineage definition were summarized in Supplementary Table S1. (B) Violin plots showing the expression of representative marker genes across major immune lineages. (C) UMAP plot showing the marker genes expression levels, with color intensity reflecting expression in individual cells. (D) Stacked bar plots showing the relative composition of immune cell subtypes across tumor samples from 12 patients, highlighting the inter-patient variability in microenvironmental cellular composition.

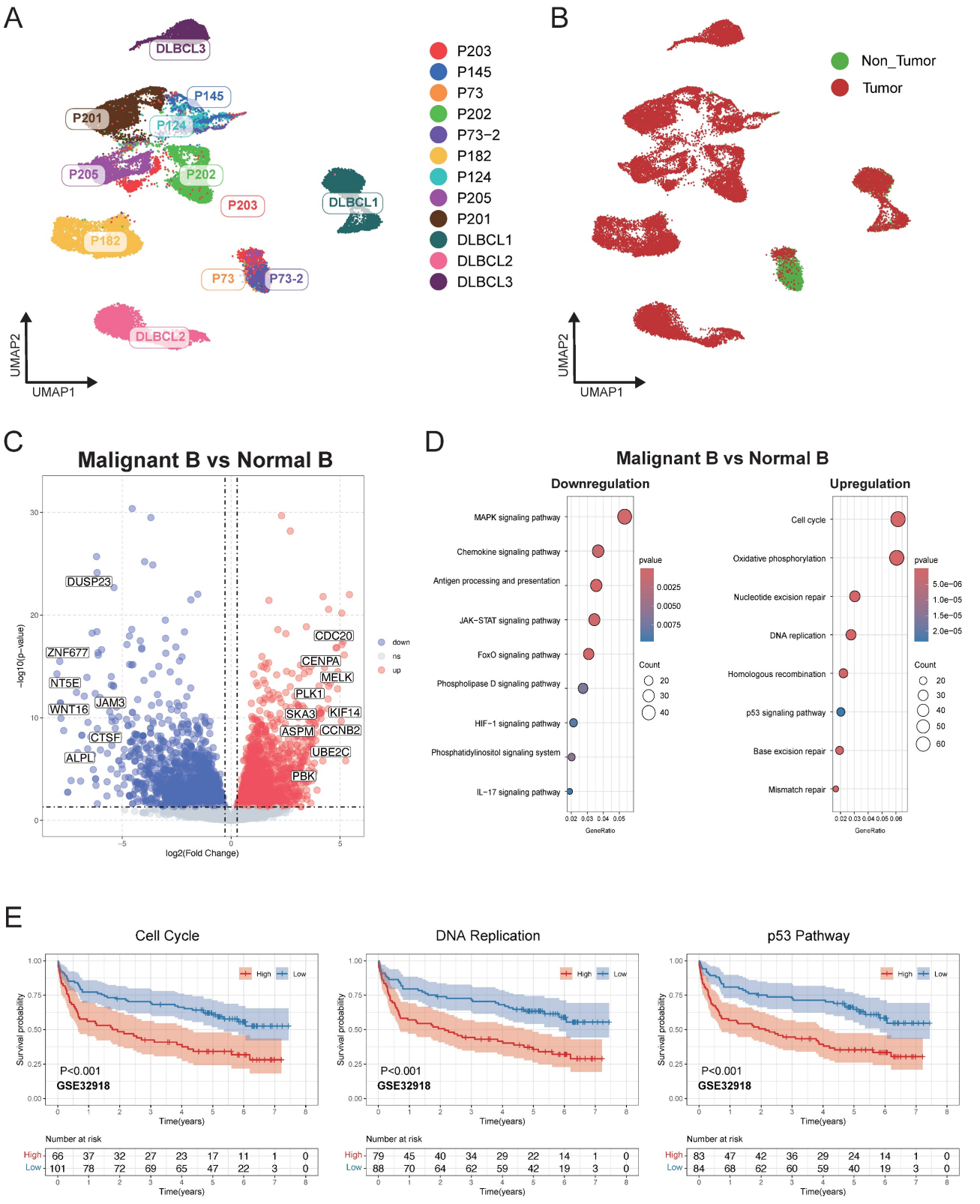

Figure 2. Single-cell analysis revealed distinct transcriptional characteristics of malignant and non-malignant infiltrating B cells in DLBCL. (A) UMAP plot showing B cells derived from 12 DLBCL patients. Each dot represents a single cell, colored by patient origin. (B) UMAP plot showing the malignant and non-malignant B cells. Malignant status was inferred using an integrated approach combining BCR clonality analysis and CNV estimation from scRNA-seq (inferCNV). (C) Volcano plot showing differentially expressed genes (DEGs) between malignant and non-malignant B cells. Significantly upregulated and downregulated genes were highlighted in red and blue, respectively ($|\log_2 FC| > 1.5$, $p < 0.05$). (D) Bubble plot presenting KEGG pathway enrichment analysis for the upregulated and downregulated DEGs. The size of the bubbles reflects the number of genes involved, while the color intensity represents the statistical significance of enrichment. (E) Kaplan–Meier survival curves illustrating the prognostic relevance of GSVA scores based on selected pathway signatures in external GEO RNA expression microarray datasets.

Clustering of myeloid cells from single-cell RNA sequencing

Eight myeloid cell subclusters were identified by single-cell RNA sequencing analysis within the integrated expression profiles of 12 DLBCL samples. This analysis characterized the transcriptional heterogeneity of infiltrating myeloid cells in DLBCL. The UMAP plot revealed distinct clustering of myeloid cells across patients based on their expression of immune markers (Figure 3A). We defined the identity of each cluster by evaluating subset-specific differentially expressed genes (Figure 3B-C). The *CLEC9A*$^+$ cDC1 cluster was characterized by the expression of *CLEC9A*, *BATF3*, and *ID2*, whereas the *CD1C*$^+$ cDC2 cluster was defined by *CD1C*, *FCGR2B*, and *CLEC10A*. The *LAMP*$^+$ cDC3 cluster was distinguished by high expression levels of *LAMP3*, *FSCN1*, and *CCR7*. Clusters *C1QC*$^+$ Macro and *SPP1*$^+$ Macro could be differentiated based on the expression levels of *C1QA/B/C* and *SPP1* respectively. A subset of cells with high expression of *FCN1* was identified distinct from macrophages, which contained *FCN1*$^+$ Monolike cells and *CD16*$^+$ Monocytes. The *LILRA4*$^+$ pDC cluster was defined by the expression of *LILRA4*, *IRF4*, and *IRF7*. Cluster distribution plot revealed that clusters *C1QC*$^+$ Macro, *FCN1*$^+$ Monolike, and *SPP1*$^+$ Macro constituted the predominant populations across all samples, a finding that was corroborated by the UMAP analysis.

Based on clustering results from scRNA-seq data, we defined transcriptional signatures representative of distinct cell subpopulations. Subsequently, we evaluated the prognostic relevance between the signatures and patient survival outcomes using a cell type deconvolution algorithm in independent RNA expression microarray data cohorts from GEO database. There was a consistent indication that elevated levels of *FCN1*$^+$ monolike cells and *SPP1*$^+$ macrophages were associated with poor prognosis across multiple DLBCL cohorts GSE31312 (Figure 3D), GSE10846, GSE11318 and GSE87371 (Figure S2).

Clustering of CD4$^+$ T cells from single-cell RNA sequencing

We characterized nine different CD4$^+$ T cell clusters within the TME of DLBCL from single-cell transcriptomes in 12 samples (Figure 4A). In order to verify the reliability of the basis of nine clusters of CD4$^+$ T cells, the expression patterns of representative marker genes in different cell clusters were analyzed. *CCR7*$^+$ Tn cells were defined by the expression of *CCR7*, *SELL* and *S1PR1*; *TCF7*$^+$ Tm cells by *TCF7*, *CCR7* and *IL7R*; *CD69*$^+$ Tm cells by *CD69*, *NR4A1/2* and *MYADM*. Temra$^+$ *CX3CR1* cells were characterized by *CX3CR1*, *KLRG1*, and *S1PR5*; *GZMK*$^+$ Tem cells by *GZMK*, *GNLY*, and *NKG7*. *IFNG*$^+$ Tfh/Th1 cells were marked by *IL21*, *IFNG*, and *CCL3*; whereas *IL6ST*$^+$ Tfh cells expressed *IL6ST*, *CD200* and *BCL6*. *TNFRSF9*$^+$ Treg cells were defined by *FOXP3* and *TNFRSF9*; ISG$^+$ Treg mainly by interferon-stimulated genes (ISGs) (Figure 4B-C). To investigate the association between CD4$^+$ T-cell subsets in tumor microenvironment and DLBCL patient's prognosis, we applied the same analytical framework used in myeloid cell analysis. Survival analyses revealed that an increased enrichment of *TNFRSF9*$^+$ Treg cells within the TME was significantly associated with poor survival. Conversely, enriched proportion of *IFNG*$^+$ Tfh/Th1 cells showed significant association with better outcomes across multiple independent GEO cohorts GSE32918 (Figure 4D) and GSE87371, GSE31312, GSE10846, GSE11318, GSE117556 (Figure S3).

Clustering of CD8$^+$ T cells from single-cell RNA sequencing

CD8$^+$ T cells represent the predominant infiltrating lymphocyte population in the DLBCL tumor microenvironment. CD8$^+$ T cells were analyzed and classified into nine distinct subpopulations based on their functional states and the expression profiles of lineage-defining genes across 12 samples (Figure 5A). To validate the reliability of this classification, we analyzed the expression patterns of representative marker genes across the identified clusters (Figure 5C-D). *TCF7*$^+$ T Naïve cells were characterized by high expression of *TCF7* and *LEF1*, and *IL7R*$^+$ Tm cells exhibited diminished *TCF7* expression alongside elevated *IL7R* and *ZNF683*, clearly distinguished from Naïve CD8$^+$ T cells. *KIR*$^+$ NK like T cells were transcriptionally similar to NK cells and expressed *KIR* genes and *KLR* family genes. In the trajectory of CD8$^+$ T cells, there was a transition from the pre-exhaustion T cells (Pre Tex), which expressed cytotoxic molecules, including *GZMK* and *GZMA* and low level of exhaustion markers, to cells *GZMK*$^+$ Tex, which expressed inhibitory receptors such as *PDCD1* and *LAG3*. Next these cells changed into terminally exhausted T cells (Terminal Tex), which were characterized by high expression of exhaustion-related genes and reduced effector function (Figure 5B). The progression from Pre Tex cells to *GZMK*$^+$ Tex, and finally to Terminal Tex cells indicated an important step of CD8$^+$ T cells dysfunction in TME. Moreover, we observed a distinct subset of exhausted T cells (*TCF7*$^+$ Tex) expressing exhaustion markers and stem-like gene *TCF7*, indicating a progenitor-exhausted signature.

ISG$^+$ CD8$^+$ T cells showed increased expression of interferon-stimulated genes (*IFIT1* and *ISG15*). MKI67$^+$ Prolif T cells showed increased *MKI67* and *NME1* expression.

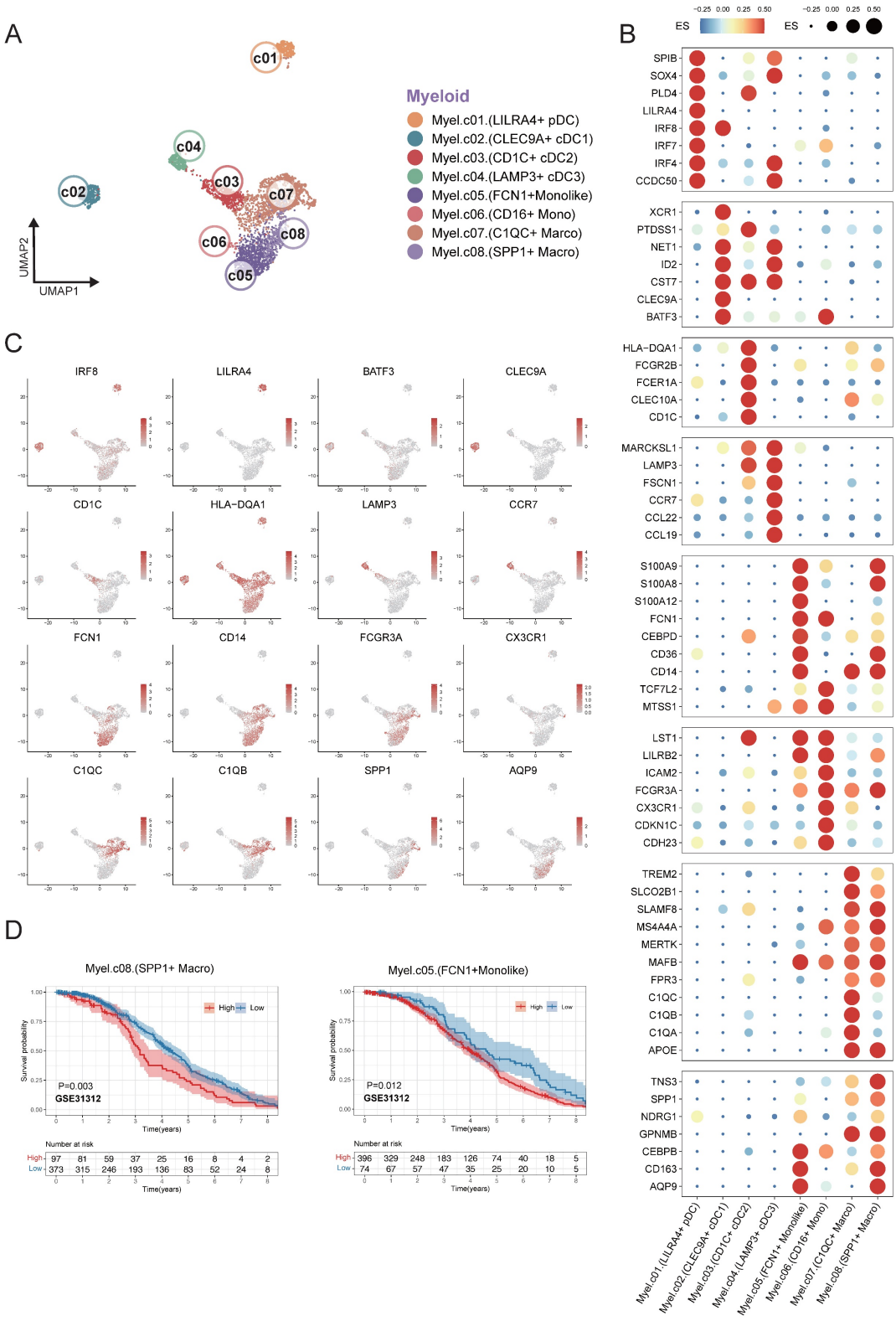

Figure 3. Fine-grained characterization of myeloid cell subpopulations in DLBCL. (A) UMAP plot showing the major lineages of myeloid cells. Marker genes used for lineage definition were summarized in Supplementary Table S1. (B) Dot plot showing the expression of marker genes associated with myeloid cells. Both dot size and color indicate effect size (ES) reflecting the average scaled expression level and the proportion of cells expressing each gene within a cluster. (C) UMAP plots colored by the log-normalized expression levels of selected canonical marker genes for myeloid cell subtypes based on scRNA-seq data. Color intensity reflects expression in individual cells. (D) Kaplan–Meier plot showing worse clinical outcome in DLBCL patients with the higher proportion of *SPP1*$^+$ myeloid and *FCN1*$^+$ moonlike cells in external GEO RNA expression microarray datasets. Patients were stratified based on enrichment scores of specific cell-type signatures, and differences in overall survival were assessed.

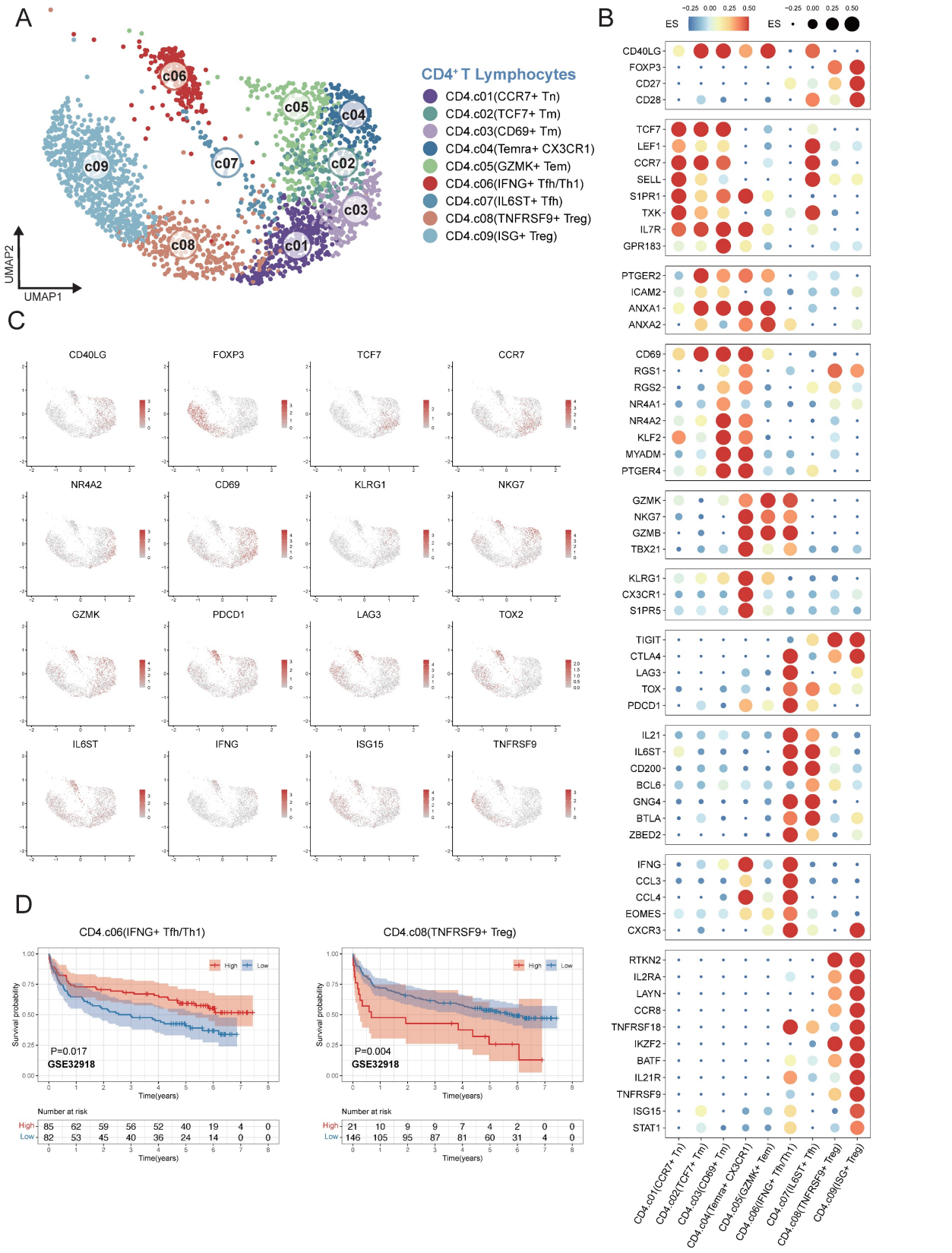

Figure 4. Fine-grained delineation of CD4⁺ T-cell subpopulations in DLBCL. (A) UMAP plot showing the major lineages of CD4⁺ T cells. Marker genes used for lineage definition were summarized in Supplementary Table S1. (B) Dot plot showing the expression of marker genes associated with CD4⁺ T cells. Both dot size and color indicate effect size (ES), reflecting the average scaled expression level and the proportion of cells expressing each gene within a cluster. (C) UMAP plots colored by the log-normalized expression levels of canonical CD4⁺ T-cell marker genes based on scRNA-seq data. Color intensity reflects expression in individual cells. (D) Kaplan–Meier survival curves demonstrating the prognostic relevance of selected signatures associated with CD4⁺ T-cell subsets in external GEO RNA expression microarray datasets. Patients were stratified based on enrichment scores of specific cell-type signatures, and differences in overall survival were assessed.

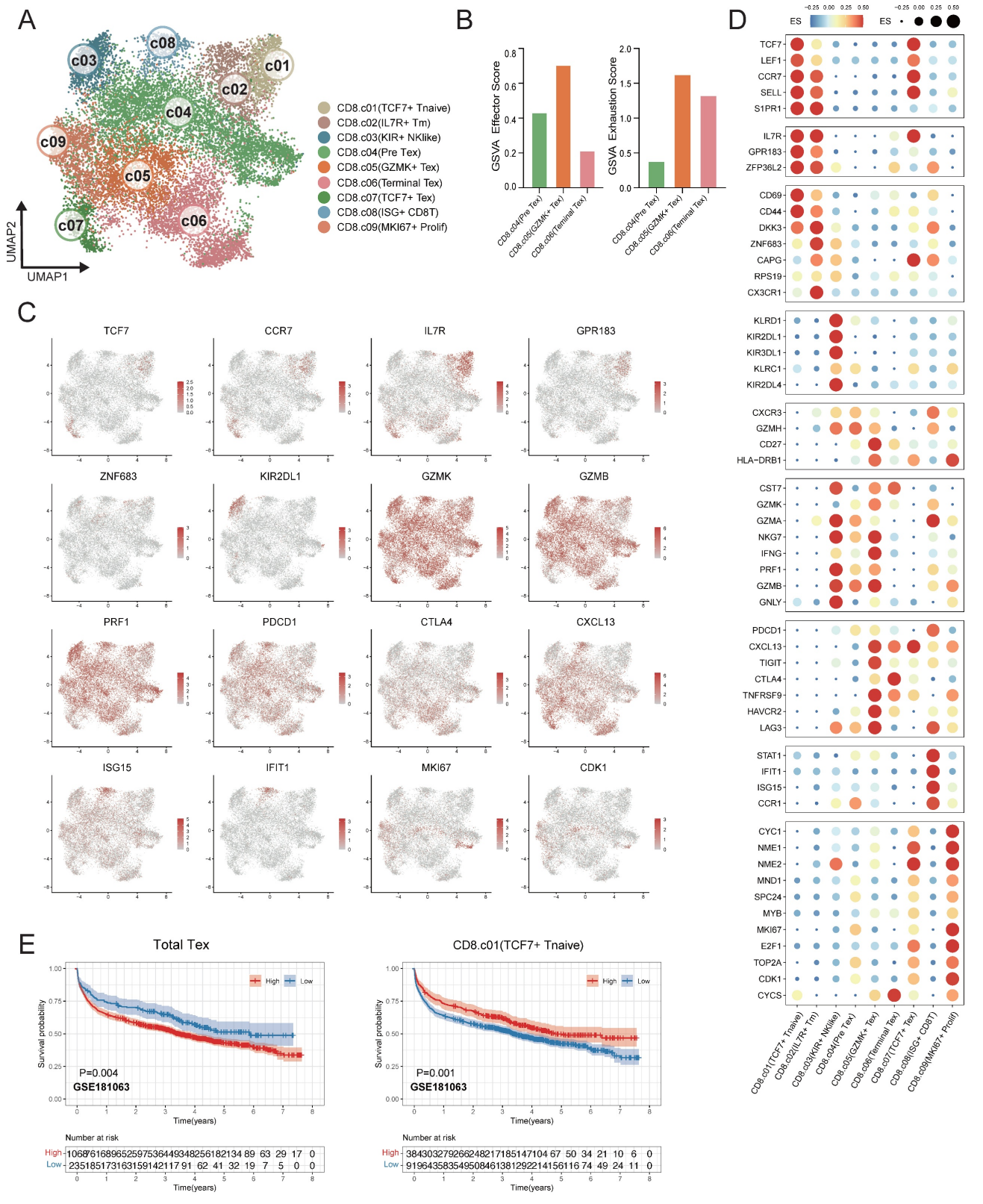

Figure 5. Fine-scale characterization of CD8⁺ T-cell subpopulations in DLBCL. (A) UMAP plot showing the major lineages of CD8⁺ T cells. Marker genes used for lineage definition are summarized in Supplementary Table S1. (B) Bar plots illustrating the GSVA scores for exhaustion and effector programs across Pre Tex, *GZMK*⁺ Tex and Terminal Tex cells. The exhaustion score was calculated based on the expression of *PDCD1*, *TOX*, *CXCL13*, *TIGIT*, *CTLA4*, *TNFRSF9*, *HAVCR2* and *LAG3*; the effector score was derived from *GNLY*, *GZMB*, *PRF1*, *IFNG*, *NKG7*, *GZMA*, *GZMK*, *CST7*, *TNF*, *FASL* and *TBX21*. (C) UMAP plots colored by log-normalized expression levels of selected canonical marker genes for CD8⁺ T-cell subsets based on scRNA-seq data. Color intensity reflects expression in individual cells. (D) Dot plot shows the expression of marker genes associated with CD8⁺ T cells. Both dot size and color indicate effect size (ES), reflecting the average scaled expression level and the proportion of cells expressing each gene within a cluster. (E) Kaplan–Meier survival curves evaluating the prognostic significance of selected CD8⁺ T-cell–related signatures in external GEO RNA expression microarray datasets. Patients were stratified based on enrichment scores of specific cell-type signatures, and differences in overall survival were assessed.

To validate the association between CD8⁺ T cell subsets in the tumor microenvironment and clinical outcomes, survival analyses were conducted in the independent DLBCL cohorts. Remarkably, the DLBCL patients had a better prognosis with a higher proportion of *TCF7*⁺ T naïve cells. In contrast, the elevated proportion total Tex cells (combination of Pre Tex + *GZMK*⁺ Tex + Terminal Tex + *TCF7*⁺ Tex) were related to diminished survival rate in DLBCL patients within GSE181063 (Figure 5E), GSE11318, GSE117556, GSE32918, GSE10846 and GSE87371 (Figure S4).

Interaction between SPP1⁺ macrophages and CD8⁺ cells

Subsequently, we employed CellChat to analyze intercellular communication within the DLBCL tumor microenvironment focusing on ligand-receptor interactions and associated signaling pathway. The heatmap plot showed that the major regulatory axis in the TME was the infiltrating CD8⁺ T lymphocytes and myeloid cells (Figure 6A). The network centrality analysis further clarified the functional roles of various cell types in mediating cell communications. Within the network of the TME, CD8⁺ T cells occupied a central hub position and exhibited significant signal senders' and receivers' potential. Myeloid cells mainly acted as robust signal initiators with limited ability to receive signals, suggesting they might function as important upstream regulators in the immune signaling cascade (Figure 6B). Notably, the *SPP1*⁺ macrophages may play an important regulatory role in the DLBCL TME due to their remarkably strong signal sending capacity.

The study also identified that *SPP1*⁺ macrophages were widespread communicators to all clusters of CD8⁺ T cells. Early CD8⁺ T cells, such as CD8⁺ naïve and Pre Tex cells, showed limited communication efficacy with *SPP1*⁺ macrophages. On the other hand, the activated CD8⁺ subsets, including *KIR*⁺ NK like CD8+ T cells and *GZMK*⁺ Tex cells exhibited increased incoming signals from *SPP1*⁺ macrophages (Figure 6C). As indicated by these findings, *SPP1*⁺ macrophages may interact preferentially with activated CD8⁺ T cells which may alter their functional states thus remodeling DLBCL tumor microenvironment.

We carried out non-negative matrix factorization analysis (NMF) to identify key outgoing (sending) and incoming (receiving) signaling and key ligand–receptor interactions patterns. The heatmap visualization showed the contributions of various cell types to different signaling. Myeloid cells especially the *SPP1*⁺ macrophages exhibited strong signal-sending capacities within the outgoing signaling heatmap. Conversely, the incoming signaling map indicated that CD8⁺ T cells, especially Pre-Tex, *GZMK*⁺ Tex and Terminal Tex subsets, exhibited significant signal-receiving activity. Meanwhile, pathway enrichment analysis showed that the SPP1 ligand-receptor axis was largely responsible for these interactions (Figure 6D-E). The data implied that myeloid cell clusters, specifically *SPP1*⁺ macrophages, may facilitate the shift of CD8⁺ T cells from the effector state toward terminal exhaustion while contributing to the remodeling of the tumor microenvironment and immune suppression.

Discussion

In our study, we analyzed single-cell and RNA expression microarray data to reveal the possible relationship between the immune landscape transcriptome and clinical outcomes in DLBCL patients. We created a comprehensive cell atlas encompassing both malignant and nonmalignant cell populations, providing evidence that the diversity of myeloid and T lymphocyte subpopulations, along with specific transcriptomic gene signatures, significantly predicted the prognosis of DLBCL patients.

Secreted Phosphoprotein-1 (SPP1), also known as Osteopontin (OPN) was extensively expressed across a variety of cell types, including T cells, B cells and myeloid cells[12]. Additionally, SPP1 was found in body fluids such as serum, bovine milk, and human urine, where it played a role in intercellular communication and the extracellular matrix[13]. Elevated levels of circulating SPP1 in serum, as well as increased SPP1 expression in tumor cells, had been associated with poor prognosis in multiple cancer types by promoting tumor cell proliferation, migration, and invasion[14,15]. Regarding circulating SPP1, the expression level of SPP1 in tumor tissue did not correlate with plasma SPP1 levels and patient outcomes, suggesting that non-malignant cells may contribute to plasma SPP1 concentrations[16]. Recently, the expression of SPP1 by macrophages had garnered significant attention in the scientific community. Eri Matsubara and colleagues have reported that elevated SPP1 expression in tumor-associated macrophages (TAMs) was indicative of a poor prognosis in patients with lung adenocarcinoma[17]. More recently, Ruben Bill et al. demonstrated that SPP1 played a role in the progression of macrophage polarization, with IFN-γ-induced CXCL9 and hypoxia-induced SPP1 emerging as critical features of the tumor microenvironment[18]. Furthermore, the knockdown of SPP1 in macrophages had been shown to mitigate tumor cell migration and induce the Th1 response by downregulating PD-L1[19]. Our findings revealed that

SPP1⁺ macrophages fostered an immunosuppressive microenvironment in DLBCL by suppressing T-cell activation and effector function, consistent with the finding that SPP1 promotes metastatic tumor growth in non-small-cell lung cancer[20]. Importantly, *in vivo* administration of the SPP1-blocking antibody suppressed liver and lung metastases[21]. Thus, the abundance of *SPP1*⁺ TAMs may serve as a predictive biomarker for patient stratification and as a therapeutical target to overcome immunotherapy resistance in DLBCL. Future studies integrating SPP1 pathway may constitute a rational and promising strategy to restore anti-tumor immunity and improve outcomes in patients with refractory disease.

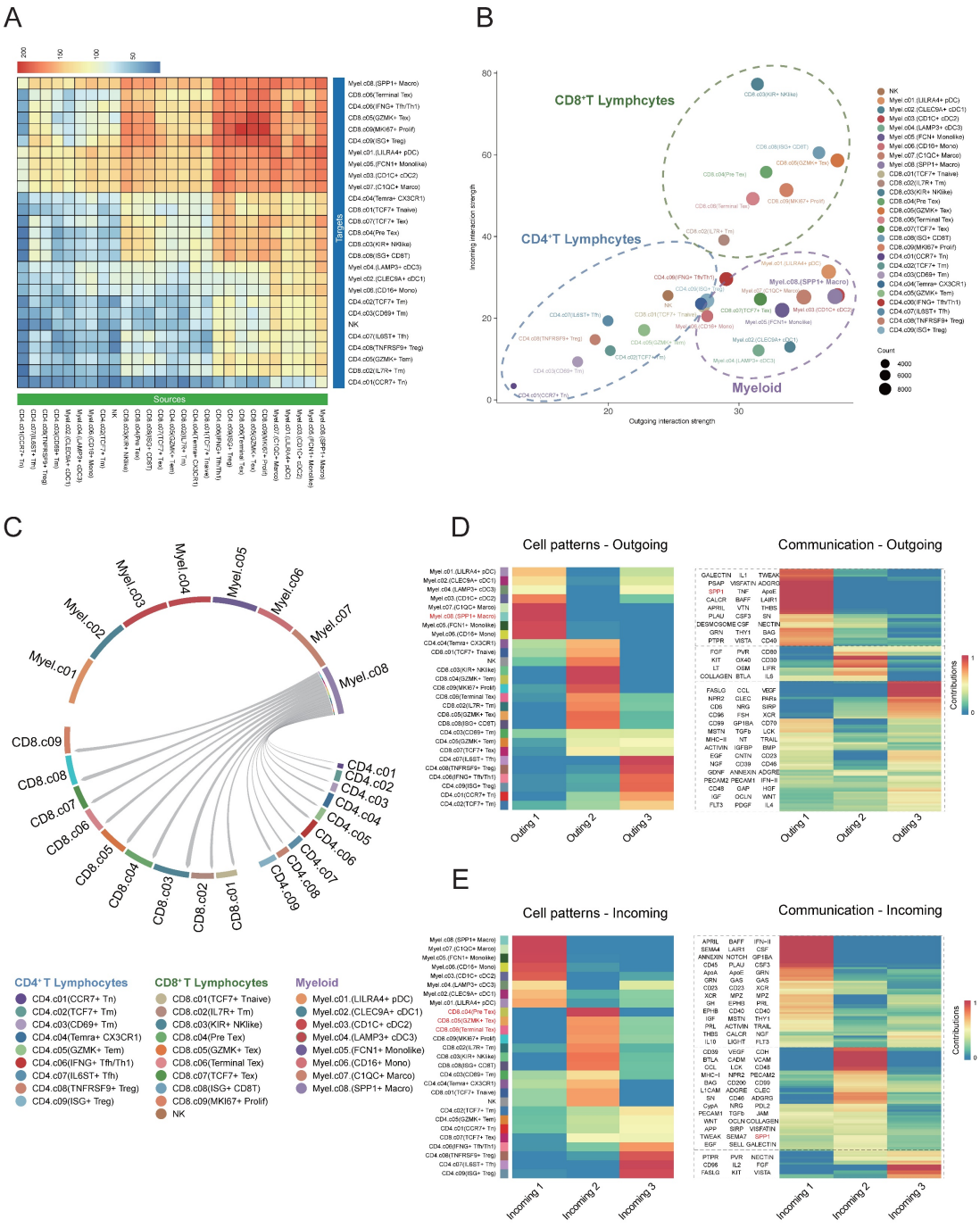

Figure 6. Cell-cell communication between SPP1⁺ cells and CD8⁺ T cells in the DLBCL tumor microenvironment. (A) Heatmap showing the inferred cell-cell communication events between different subclusters. Columns represent signal-sending cells, and rows represent signal-receiving cells. The color represented the strength of interaction. (B) Scatter plot depicting cell-cell communication network centrality. The X-axis represents the contribution of each cell type in transmitting signals, while the Y-axis indicates its role in receiving signals. (C) Circular chord diagram illustrating interaction strength between *SPP1*⁺ Macro cells and T cells. Each ribbon represents a communication event, with *SPP1*⁺ Macro cells as signal sources and T cells as recipients. (D) Left: Heatmap of NMF-based clustering of cells according to outgoing signaling patterns. Color intensity reflects the relative contribution of each cell type to the corresponding signaling module. Right: Heatmap showing the composition of signaling pathways within each NMF-defined module. Color intensity denotes the strength of intercellular communication for each pathway. (E) Heatmap as in (D) but based on incoming signaling patterns.

The complement system, a critical component of the innate immune system, was instrumental in the recognition and elimination of pathogens[22]. Numerous studies had demonstrated that the expression of C1q was correlated with markers indicative of M2-like macrophage phenotypes. Zhang et al. identified that C1q-positive TAMs exhibited high expression levels of several macrophage markers, including *CD163* and *APOE*, as well as inhibitory molecules such as *Tim-3* and *PD-1*, suggesting an immunosuppressive role within the tumor microenvironment[23]. Furthermore, research by Lubka Roumenina et al. indicated that TAMs were the predominant cell type responsible for C1q production, and the density of C1q-positive cells were associated with poor prognosis in advanced clear cell renal cell carcinoma (ccRCC)[24]. Additionally, C1q-expressing TAMs and complement activation products were found to promote inflammation, T-cell exhaustion, and tumor progression. Notably, the ablation of C1q, C4, and C3 in mice resulted in a reduction in tumor growth[24].

T cell exhaustion referred to a range of dysfunctional states observed in antigen-specific $CD8^+$ T cells, a concept initially characterized in the context of chronic viral infections[25]. Due to mechanisms of immunotolerance and immunosuppression, $CD8^+$ T cells infiltrating tumors often failed to achieve full activation, subsequently transitioning into an exhausted and dysfunctional state characterized by diminished proliferative capacity, reduced cytokine production, and impaired tumor cell lysis[26]. Gene expression profiling had demonstrated that exhausted T cells exhibited upregulation of immune checkpoint receptors, including PD-1, CTLA-4, Tim-3, LAG-3[27–29]. The therapeutic targeting of these inhibitory receptors, particularly CTLA-4 and PD-1, with specific antibodies had led to significant improvements in clinical outcomes for patients with advanced melanoma, non-small-cell lung cancer, renal cell carcinoma, and B cell malignancies[30–33]. Numerous studies have indicated that the extent of exhausted T-cell infiltration correlates with the response rates to immunotherapy. Similarly, in our study, DLBCL patients with a higher percentage of total exhausted $CD8^+$ T cells within tumor microenvironment tended to have worse survival outcomes. More importantly, *TCF7* expression regulated differentiation of T cells in tumor tissue[34]. The loss of T-cell exhaustion regulatory genes, such as *NR4a1* and *NR4a2*, triggered *TCF7* expression promoted anti-tumor immunity in tumor infiltration $CD8^+$ T cells [35]. Similarly, the infiltration of stem-like exhausted T precursor/ progenitor cells, such as $TCF7^+$ Tex cells, were correlated with functionality of memory T cells and favorable prognosis of DLBCL patients.

The sequence of processes facilitating DNA replication and cell division was referred to as the cell cycle. The advancement of cancer was characterized by aberrant activity within the cell cycle[36]. DNA mutations contributed to the dysregulation of cell cycle kinases; a phenomenon frequently associated with abnormal cell division and the uncontrolled proliferation characteristic of cancer cells. Notably, aberrant activation of cyclin-dependent kinases (CDKs), a common occurrence in human malignancies, underscored the rationale for developing synthetic CDK inhibitors as anticancer therapeutics[37]. Concerning the molecular signatures within DLBCL tumor cells, our results illustrated that cell cycle, DNA replication, and p53 signal pathway related genes were significantly upregulated in malignant B cells. Consistent with our results, the upregulated cyclins, cell cycle-specific antigens PCNA and mini-chromosome maintenance (MCM) proteins promoted the progression of cell cycle regulation, DNA replication in various kinds of tumor development [38–40].

In summary, the intra-tumoral heterogeneity and microenvironmental interactions characteristic of DLBCL present promising opportunities for prognostic stratification and the advancement of novel immunotherapeutic strategies.

Material and Methods

Data process and single cell sequence analysis

Two DLBCL scRNA-seq datasets were sourced from GSA-Human (HRA002297)[41] and GEO (GSE182436)[10]. The HRA002297 dataset comprises single-cell data derived from nine DLBCL tumor tissue samples, while the GSE182436 dataset included data from three patients. After merging the datasets, we utilized Seurat (v5.2.1) to process the Unique Molecular Identifier (UMI) count matrices which represent the number of unique transcripts captured per cell and conducted quality control, keeping 54,199 cells and 41,166 genes for subsequent analysis[42,43]. The scRNA-seq data was processed through log transformation, normalization, and dimensionality reduction, followed by visualization. The Variance Stabilizing Transformation (VST) method pinpointed the top 2,000 highly variable genes, and Principal Component Analysis (PCA) was used for dimensionality reduction, with UMAP applied for clustering and visualization[44]. Batch effects were corrected using Harmony Integration in Seurat. The Leiden algorithm was employed for multiscale clustering to identify cell meta clusters, which were

annotated based on gene expression profiles[45]. Myeloid cells, $CD4^+$ T cells, and $CD8^+$ T cells were then analyzed using the same preprocessing and clustering approach. The related marker genes were chosen according to canonical immune cell signatures reported in prior single-cell studies of the tumor microenvironment and hematopoietic lineages (Supplementary Table S1) [46–48]. Concerning the microarray data, we selected seven independent DLBCL RNA expression microarray data cohorts from the Gene Expression Omnibus (GEO) database (GSE10846, GSE11318, GSE31312, GSE32918, GSE87371, GSE117556 and GSE181063), which totally contained clinical information and expression data from over 3,000 DLBCL patients.

Malignancy Identification

First, we analyzed BCR sequencing data and identified dominant monoclonal and polyclonal B cell subsets. Subsequently, we performed single-cell Copy Number Variation (CNV) analysis for these subsets using the InferCNV (v1.22.0) algorithm. The malignant B cells were identified based on BCR clonality analysis, κ/λ chain ratio and InferCNV results[49]. The reference population for InferCNV was selected from a cluster containing cells derived from multiple patient samples that showed no detectable copy number alterations, supporting its suitability as a non-malignant baseline.

Cellular composition deconvolution and survival analysis

Seven cohort datasets were obtained from the GEO database, which contained over 3,000 samples and survival data of patients. We utilized scRNA-seq clustering to determine a transcriptional signature for specific cell subpopulations. Next, we used the BayesPrism algorithm to establish a prognosis model using risk scores generated from these datasets to determine its relationship with patients' survival outcome[49]. The Overall Survival (OS) was defined as the time from initiation of disease until death from any cause. Survival R package (Version 3.8-3) was employed to study the correlation of risk factors with survival probability[50]. Univariate and multivariate Cox regression analyses were conducted, presented with 95% confidence intervals, to assess the effect of the mentioned factors on survival endpoints. Results shown statistically significant if p value < 0.05. We only considered findings that showed consistent survival trends as well as statistical significance in at least three studies as clinically relevant.

Cell-cell communication analysis

CellChat (version 2.1.2) was used to analyze intercellular communication based on a curated ligand-receptor interactions[51]. The analysis was performed on the basis of gene expression data, quantified as transcriptome-wide read counts mapped to protein-coding genes. The average expression level for each identified cell cluster was computed and served as input to infer communication probabilities and reconstruct signaling networks.

KEGG pathway enrichment analysis

Pseudobulk count matrices for each cell type were generated from the Seurat object and normalized using DESeq2. Principal Component Analysis (PCA) was performed to evaluate sample reproducibility and the stability of inter-sample variation. Differentially Expressed Genes (DEGs) were identified using a threshold of $|\log_2 \text{fold change}| > 1.5$ and $p < 0.05$, and visualized with volcano plots generated by ggplot2. KEGG pathway enrichment analysis of DEGs was conducted using cluster Profiler (v4.14.4), and the top pathways were ranked by p-values. Gene Set Variation Analysis (GSVA) (v2.0.5) was performed on selected gene sets or pathways to calculate enrichment scores for further analysis (Supplementary Table S2).

Abbreviations

DLBCL: diffuse large B-cell lymphoma; BCR: B-cell receptor; CNV: copy number variation; NHL: non-Hodgkin lymphomas; GCB: B-cell-like; ABC: activated B-cell-like; GEP: mRNA expression profiling; ECM: extracellular matrix; TME: tumor microenvironment; scRNA-seq: single-cell RNA sequencing; TCR: T-cell receptor; UMAP: uniform manifold approximation and projection; ISGs: interferon-stimulated genes; Pre Tex: pre-exhaustion T cells; Terminal Tex: terminally exhausted T cells; SPP1: secreted phosphoprotein-1; OPN: osteopontin; TAMs: tumor-associated macrophages; ccRCC: clear cell renal cell carcinoma; CDKs: cyclin-dependent kinases; MCM: mini-chromosome maintenance; VST: variance stabilizing transformation; PCA: principal component analysis; GEO: gene expression omnibus; OS: overall survival; DEGs: differentially expressed genes; GSVA: gene set variation analysis; UMI: unique molecular identifier.

Supplementary Material

Supplementary figures and tables.
https://www.jcancer.org/v17p0469s1.pdf

Acknowledgements

This study has been supported by grants from National Natural Science Foundation of China (82470178, 81970179), Beijing Natural Science Foundation (Z220023, L254088), Science Foundation of Peking University Cancer Hospital (JC202401) and Chinese Society of Clinical Oncology Foundation (Grant No. Y-2024AZ(BTK)MS—0024).

Competing Interests

The authors have declared that no competing interest exists.

References

1. A Clinical Evaluation of the International Lymphoma Study Group Classification of Non-Hodgkin's Lymphoma. Blood. 1997;89(11):3909-3918.
2. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127(20):2375-2390.
3. Salles G, Barrett M, Foà R, et al. Rituximab in B-Cell Hematologic Malignancies: A Review of 20 Years of Clinical Experience. Adv Ther. 2017;34(10):2232-2273.
4. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. 2000;403.
5. Miao Y, Medeiros LJ, Li Y, Li J, Young KH. Genetic alterations and their clinical implications in DLBCL. Nat Rev Clin Oncol. 2019;16(10):634-652.
6. Scott DW, Mottok A, Ennishi D, et al. Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. J Clin Oncol. 2015;33(26):2848-2856.
7. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med. 2018;378(15):1396-1407.
8. Lenz G, Wright G, Dave SS, et al. Stromal Gene Signatures in Large-B-Cell Lymphomas. N Engl J Med. 2008;359(22):2313-2323.
9. Kotlov N, Bagaev A, Revuelta MV, et al. Clinical and biological subtypes of B-cell lymphoma revealed by microenvironmental signatures. Cancer Discov. 2021;11(6):1468-1489.
10. Steen CB, Luca BA, Esfahani MS, et al. The Landscape of Tumor Cell States and Ecosystems in Diffuse Large B Cell Lymphoma. Cancer Cell. 2021;39(10):1422-1437.e10.
11. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, et al. The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. Cell. 2020;181(2):236-249.
12. Zhao H, Chen Q, Alam A, et al. The role of osteopontin in the progression of solid organ tumour. Cell Death Dis. 2018;9(3):1-15.
13. Haylock DN, Nilsson SK. Osteopontin: a bridge between bone and blood. Br J Haematol. 2006;134(5):467-474.
14. Kothari AN, Arffa ML, Chang V, et al. Osteopontin—A Master Regulator of Epithelial-Mesenchymal Transition. J Clin Med. 2016;5(4):39.
15. Shevde LA, Samant RS. Role of osteopontin in the pathophysiology of cancer. Matrix Biol J Int Soc Matrix Biol. 2014;37:131-141.
16. Mack PC, Redman MW, Chansky K, et al. Lower Osteopontin Plasma Levels Are Associated With Superior Outcomes in Advanced Non–Small-Cell Lung Cancer Patients Receiving Platinum-Based Chemotherapy: SWOG Study S0003. J Clin Oncol. 2008;26(29):4771-4776.
17. Matsubara E, Yano H, Pan C, et al. The Significance of SPP1 in Lung Cancers and Its Impact as a Marker for Protumor Tumor-Associated Macrophages. Cancers. 2023;15(8):2250.
18. Bill R, Wirapati P, Messemaker M, et al. CXCL9:SPP1 macrophage polarity identifies a network of cellular programs that control human cancers. Science. 2023;381(6657):515-524.
19. Zhang Y, Du W, Chen Z, Xiang C. Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. Exp Cell Res. 2017;359(2):449-457.
20. Shojaei F, Scott N, Kang X, et al. Osteopontin induces growth of metastatic tumors in a preclinical model of non-small lung cancer. J Exp Clin Cancer Res CR. 2012;31(1):26.
21. Li H, Lan L, Chen H, et al. SPP1 is required for maintaining mesenchymal cell fate in pancreatic cancer. Nature. Published online September 24, 2025:1-7.
22. Dunkelberger JR, Song WC. Complement and its role in innate and adaptive immune responses. Cell Res. 2010;20(1):34-50.
23. Zhang S, Peng W, Wang H, et al. C1q+ tumor-associated macrophages contribute to immunosuppression through fatty acid metabolic reprogramming in malignant pleural effusion. J Immunother Cancer. 2023;11(8):e007441.
24. Roumenina LT, Daugan MV, Noé R, et al. Tumor Cells Hijack Macrophage-Produced Complement C1q to Promote Tumor Growth. Cancer Immunol Res. 2019;7(7):1091-1105.
25. Moskophidis D, Lechner F, Pircher H, Zinkernagel RM. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. Nature. 1993;362(6422):758-761.
26. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol. 2015;15(8):486-499.
27. Thommen DS, Schumacher TN. T Cell Dysfunction in Cancer. Cancer Cell. 2018;33(4):547-562.
28. Ahmadzadeh M, Johnson LA, Heemskerk B, et al. Tumor antigen–specific CD8 T cells infiltrating the tumor express high levels of PD-1 and are functionally impaired. Blood. 2009;114(8):1537-1544.
29. Blackburn SD, Shin H, Haining WN, et al. Coregulation of CD8+ T cell exhaustion during chronic viral infection by multiple inhibitory receptors. Nat Immunol. 2009;10(1):29-37.
30. Greenbaum AM, Fromm JR, Gopal AK, Houghton AM. Diffuse large B-cell lymphoma (DLBCL) is infiltrated with activated CD8+ T-cells despite immune checkpoint signaling. Blood Res. 2022;57(2):117-128.
31. Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus Everolimus in Advanced Renal Cell Carcinoma. N Engl J Med. 2015;373(19):1803-1813.
32. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Non-squamous Non-small Cell Lung Cancer. N Engl J Med. 2015;373(17):1627-1639.
33. Robert C, Ribas A, Wolchok JD, et al. Anti-programmed-death-receptor-1 treatment with pembrolizumab in ipilimumab-refractory advanced melanoma: a randomised dose-comparison cohort of a phase 1 trial. Lancet Lond Engl. 2014;384(9948):1109-1117.
34. Siddiqui I, Schaeuble K, Chennupati V, et al. Intratumoral Tcf1+PD-1+CD8+ T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. Immunity. 2019;50(1):195-211.e10.
35. Srirat T, Hayakawa T, Mise-Omata S, et al. NR4a1/2 deletion promotes accumulation of TCF1+ stem-like precursors of exhausted CD8+ T cells in the tumor microenvironment. Cell Rep. 2024;43(3):113898.
36. Matthews HK, Bertoli C, de Bruin RAM. Cell cycle control in cancer. Nat Rev Mol Cell Biol. 2022;23(1):74-88.
37. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer. 2009;9(3):153-166.
38. Bertoli C, Skotheim JM, de Bruin RAM. Control of cell cycle transcription during G1 and S phases. Nat Rev Mol Cell Biol. 2013;14(8):518-528.
39. Strzalka W, Ziemienowicz A. Proliferating cell nuclear antigen (PCNA): a key factor in DNA replication and cell cycle regulation. Ann Bot. 2011;107(7):1127-1140.
40. Juríková M, Danihel Ľ, Polák Š, Varga I. Ki67, PCNA, and MCM proteins: Markers of proliferation in the diagnosis of breast cancer. Acta Histochem. 2016;118(5):544-552.
41. Liu N, Jiang C, Yao X, et al. Single-cell landscape of primary central nervous system diffuse large B-cell lymphoma. Cell Discov. 2023;9(1):55.
42. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression. Nat Biotechnol. 2015;33(5):495-502.
43. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019;20:296.
44. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. J Open Source Softw. 2018;3(29):861.
45. Traag VA, Waltman L, van Eck NJ. From louvain to leiden: guaranteeing well-connected communities. Sci Rep. 2019;9:5233.
46. Cheng S, Li Z, Gao R, et al. A pan-cancer single-cell transcriptional atlas of tumor infiltrating myeloid cells. Cell. 2021;184(3):792-809.e23.
47. Zheng L, Qin S, Si W, et al. Pan-cancer single-cell landscape of tumor-infiltrating T cells. Science. 2021;374(6574):abe6474.
48. Yang Y, Chen X, Pan J, et al. Pan-cancer single-cell dissection reveals phenotypically distinct B cell subtypes. Cell. 2024;187(17):4790-4811.e22.
49. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. Nat Cancer. 2022;3(4):505-517.

50. Allignol A, Latouche A. CRAN task view: survival analysis. February 9, 2025. Accessed June 20, 2025. https://CRAN.R-project.org/view=Survival
51. Jin S, Guerrero-Juarez CF, Zhang L, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021;12:1088.