

Heterogeneity of tumor immune microenvironment in malignant and metastatic change in LUAD is revealed by single-cell RNA sequencing

Haiqiang Wang¹, Guoliang Han¹, Jiakuan Chen¹

¹Department of Thoracic Surgery, Tangdu Hospital Air Force Medical University, Fourth Military Medical University, Xian, Shanxi, China

Correspondence to: Jiakuan Chen; email: tangduedu@afmmu.cn

Keywords: lung adenocarcinoma, single-cell, immune microenvironment, T regulatory cells, biomarkers

Received: February 10, 2023

Accepted: May 9, 2023

Published: June 16, 2023

Copyright: © 2023 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Lung adenocarcinoma (LUAD) is the most common type of non-small cell lung cancer and accounts for approximately 40% of all lung cancer cases. Multiple distant metastases are the major cause of mortality in lung cancer. In this study, single-cell sequencing datasets of LUAD were utilized to depict the transcriptome characteristic of LUAD based on the bioinformatic method. Firstly, the transcriptome landscape of heterogeneous cell types in LUAD was analyzed and memory T cells, NK cells, and helper T cells were revealed to be the common immune cells in tumor, normal, and metastasis tissue, respectively. Then, marker genes were calculated and 709 genes were identified to play a vital role in the microenvironment of LUAD. While macrophages were reported to act as one of the cells in LUAD, enrichment analysis of macrophage marker genes revealed the important role of macrophages in the activation of neutrophils. Next, the results of cell-cell communication analysis suggested that pericytes interact with broad immune cells via MDK-NCL pathways in metastasis samples, MIF-(CD74+CXCR4) and MIF-(CD74+CC44) interaction especially occurred between different cell types in tumor and normal samples. Finally, bulk RNA-seq was integrated to validate the prognosis effect of the marker gene and the maker gene of M2 macrophage, CCL20, showed the most related to LUAD prognosis. Besides, ZNF90 (Helper T cells), FKBP4 (memory T, helper T, Cytotoxic T, and B cells), CD79A (B cells), TPI1 (pericyte), and HOPX (epithelial cells, pericytes) were also pivotal in the pathology of LUAD, helping researchers understand the molecular insight of microenvironment in LUAD.

INTRODUCTION

Lung adenocarcinoma (LUAD) is the most common histological type of lung cancer and has the highest cancer mortality rate worldwide [1]. Immunotherapies had presented sustained clinical significance in LUAD [2]. However, due to intrinsic genomic and immunogenic heterogeneity, immunotherapy had limited efficacy and the effect for different patients may be inconsistent [3]. To unravel the role of the distinct genetic features of lung cancer subtypes among patients, new research opportunities and

technologies have arisen that might contribute to different therapeutic decisions [4, 5]. Plenty of research had reported that the tumor immune microenvironment (TIME) impacts cancer progression and metastases, as well as affects patient prognoses and outcomes [6, 7], which emphasizes the important role of immune cells, vital components of the tumor microenvironment, impacting on patient survival and tumor progression [8, 9]. Although researchers had made great efforts in the past decades, the related information about the role of TIME's characteristics in LUAD is limited.

Recently, next-generation sequencing, especially single-cell RNA sequencing is a possible option for the analysis of detailed cell population subtypes of LUAD from bulk tissue samples at single cell level [10, 11]. Previous researchers had conducted a comprehensive analysis of single-cell RNA-sequencing data of LUAD and identified prognostic signatures based on B cells [12] and natural killer cells [13]. Then, multivariate analysis was performed to demonstrate that the signature was an independent prognostic factor and highly correlated with immune-cell infiltrations and immune-suppressive states. Although several studies revealed the important role of immune cells and related marker genes impact on LUAD prognosis and process, a comprehensive understanding of TIME malignant and metastatic change in LUAD is still limited. In this study, we collected single-cell sequencing datasets of LUAD and applied an integrated bioinformatic method to depict the transcriptome characteristic of LUAD in normal, tumoral, and metastasis condition. Firstly, the transcriptome landscape of heterogeneous cell types in LUAD was analyzed and marker genes of each cell type were calculated, which were identified to play a vital role in the microenvironment of LUAD. GO and KEGG enrichment analysis, as well as GSEA, of marker genes were performed to explore the functional pathologies. Next, cell-cell communication analysis and trajectory analysis was performed and bulk RNA-seq were integrated to validate the prognosis effect of the marker genes. We aim to broaden the understanding of molecular insight of microenvironment in the malignant and metastatic change in LUAD and the results of the study provide novel insights into the TIME of LUAD patients in different pathological conditions, improve patient prognoses, and provide guidance for selecting clinical LUAD treatments.

MATERIALS AND METHODS

Data procession

Single-cell transcriptome files of GSE123902 [14] were downloaded from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) [15]. Bulk tumor transcriptome profiles and clinical information about LUAD were downloaded from The Cancer Genome Atlas (TCGA) database [16] via the UCSC Xena browser (<http://xenabrowser.net/>) [17]. R software (version: 4.1) was used for all the analyses in the manuscript.

Single-cell RNA seq analysis

Two single-cell transcriptome profiles from GSE123902 were collected and the “Seurat” package was applied to perform the single-cell RNA-seq

analysis. The batch effect from studies was removed with regularized negative binomial regression by the “Seurat” package. The non-linear dimensional reduction was performed with the t-SNE method. Cluster biomarkers were found by the “Seurat” package.

Cell-type specific markers identification

Analysis of differentially expressed genes (DEGs) between different samples in certain clusters was performed using the function “FindAllMarkers” in “Seurat” package [18] with standard comparison parameters. Invalid genes were removed and a threshold of absolute log₂ fold change > 1 and a significance threshold of adjusted *p*-value < 0.05 were chosen as criteria for valid DEGs. Then, cell-type-specifically expressed genes were also calculated with Seurat, and genes satisfying the above conditions were retained. Finally, take the intersection of DEGs between samples and cell-type-specifically expressed genes to be Cell-type specific marker genes.

Pathway enrichment analysis

The Gene Set Variation Analysis (GSVA) [19] R package was used for Gene Set Enrichment Analyze (GSEA) [20] of biological processes. The clusterProfiler [21] R package was applied to Gene Ontology (GO) enrichment analysis [22] and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis [23]. Genes related to immune-related pathways were collected from the Molecular Signatures Database (MSigDB) [24] and The GSEA algorithm was used to quantify the scores of immune-related pathways to explore the immune-relevant biological processes.

ssGSEA implementation

The marker genes of 28 immune cell subtypes were collected from public datasets. Infiltration levels of different immune cells, including innate and adaptive immunity cells in each LUAD sample were quantified using single-sample gene-set enrichment analysis (ssGSEA) [25]. We also used the estimation of stromal and immune cells in malignant tumors using expression data (ESTIMATE) algorithm [26] to calculate immune scores and stromal scores to determine levels of TIME immune.

Survival analysis

The mRNA profiles and related clinical information of LUAD samples were collected from TCGA database. Survival analysis based on the Cox model was used to estimate the survival risk of patients under different

conditions. For a specific marker gene, patients were divided into two groups according to median gene expression. Then, Kaplan-Meier curve analysis was performed with the R package “Survival” [27] and the p -value between the two groups was also calculated. Marker genes ($p < 0.05$) were identified to be biomarkers significantly related to LUAD prognosis.

Regulatory element analysis

Important regulatory transcription factors in tumor, normal, and metastasis samples were analyzed by single-cell regulatory network inference and clustering (SCENIC) algorithm [28]. To save calculation time, 1000 cells were randomly selected for analysis, and an important transcription factor was obtained according to the co-expression analysis of TF-target, which interacted with 2592 pairs of TF-targets, and regarded the regulatory network formed by TF and its target gene as a regulatory module, and calculated the area under the curve (AUC) value of the module to characterize its activity. The activity of different regulatory modules in different cell groups and sample types is different, indicating that the function of transcription factors is different in different tissue states.

Trajectory analysis

Monocle3 [29] was applied to explore differences in the developmental trajectories of immune cells between different samples. Trajectory analysis of different cell types in different samples shows that cell development trajectories are different between different samples, such as B cells have no obvious developmental relationship in the normal sample, while the cells in the tumor sample are divided into 3 clusters, and there are obvious transcript differences between each cluster, that is, differences in developmental relationships.

Cell-cell communication analysis

Signaling crosstalk is critical for informing diverse cellular decisions, including decisions to activate cell cycle or programmed cell death, undergo migration or differentiate along the lineage [30, 31]. Cell-cell interactions were analyzed using Cellchat [32], an open-source R package (<https://github.com/sqjin/CellChat>) to infer, visualize and analyze intercellular communications from scRNA-seq data, comparing differences in cell-to-cell interactions between sample classes. The result shows the intercellular interaction between normal samples and metastatic, tumor samples, it revealed that in normal samples, macrophage-related signaling pathways are more active, while in metastasis samples, macrophage-related signaling pathways are

missing, and epithelial cells in turn interact more with Regulatory T cells and monocyte.

Immunocyte abundance calculation

T cells, one of the important components of the immune system, are divided into subpopulations in this project. Including Helper T cell, Memory T cell, Cytotoxic T cell, etc., but immune cells, including B cells and dendritic cells, have more subdivided subsets, it is difficult to distinguish them according to cell marker molecules, so the marker gene of immune cell subsets is collected, and the ssGSEA method is used to calculate the enrichment fraction of T cell subset genes in each cell, and the enrichment score is high. It indicates that the cell has a high functional similarity with the reference T cell.

RESULTS

Transcriptome landscape of heterogeneous cell types in lung adenocarcinomas (LUAD)

We collected single-cell sequencing dataset of 17 LUAD donors from GEO database (GSE123902). After strict quality control (see Method), we screened 37030 cells, consisting of 17908 cells from primary tumor samples, 8143 cells from metastasis samples, and 10988 cells from normal samples. To investigate the heterogeneity in different sample classes, all the scRNA-seq data were merged and normalized to create a global cell atlas [14]. Clustering with t-distributed stochastic neighbor embedding (TSNE) algorithm, all the cells were clustered into 28 clusters and then assigned to 17 cell types (Figure 1A; Supplementary Table 1) based on canonical markers reported previously [14, 33–40]. According to the clinical pathology, these samples were divided into primary tumor, metastasis, and normal groups (Figure 1B). Cytotoxic T cells and Memory T cells were the most common cells and immune cells, with 11 subpopulations of B cells, cytotoxic T cells, dendritic cells (DC), Helper T cells, M1 macrophage, M2 macrophage, Mast cells, memory T cells, monocyte, NK cells, and regulatory T cell, come up of 84.58% (31320/37030) of all the cells, thus provide us the opportunity to look insight of the immune micro-environment in different pathological samples.

Next, we elucidated the fraction of immune cell types in different sample classes (Figure 1C). In different pathological conditions, immune cells emerged in different abundance [41]. Based on comparisons of normal, metastasis, and tumor tissues of all immune cells, Memory T cell was predominant in tumor tissue, while NK cells were prevalent in normal tissue and

Helper T cells were the common immune cells in metastasis tissue. Besides, the proportion of pericytes and monocyte were much higher in metastasis tissue than in tumor and normal tissue and the proportion of B cell was higher in tumor tissue. In contrast, Cytotoxic T cell was the most common immune cells (Figure 1D)

while the distribution of cytotoxic T cell between different samples in more even. Taken together, our scRNA-seq analyses dissected the heterogeneity of the immune landscape between normal, tumor, and metastasis samples of LUAD, and different sample classes have different dominant cells.

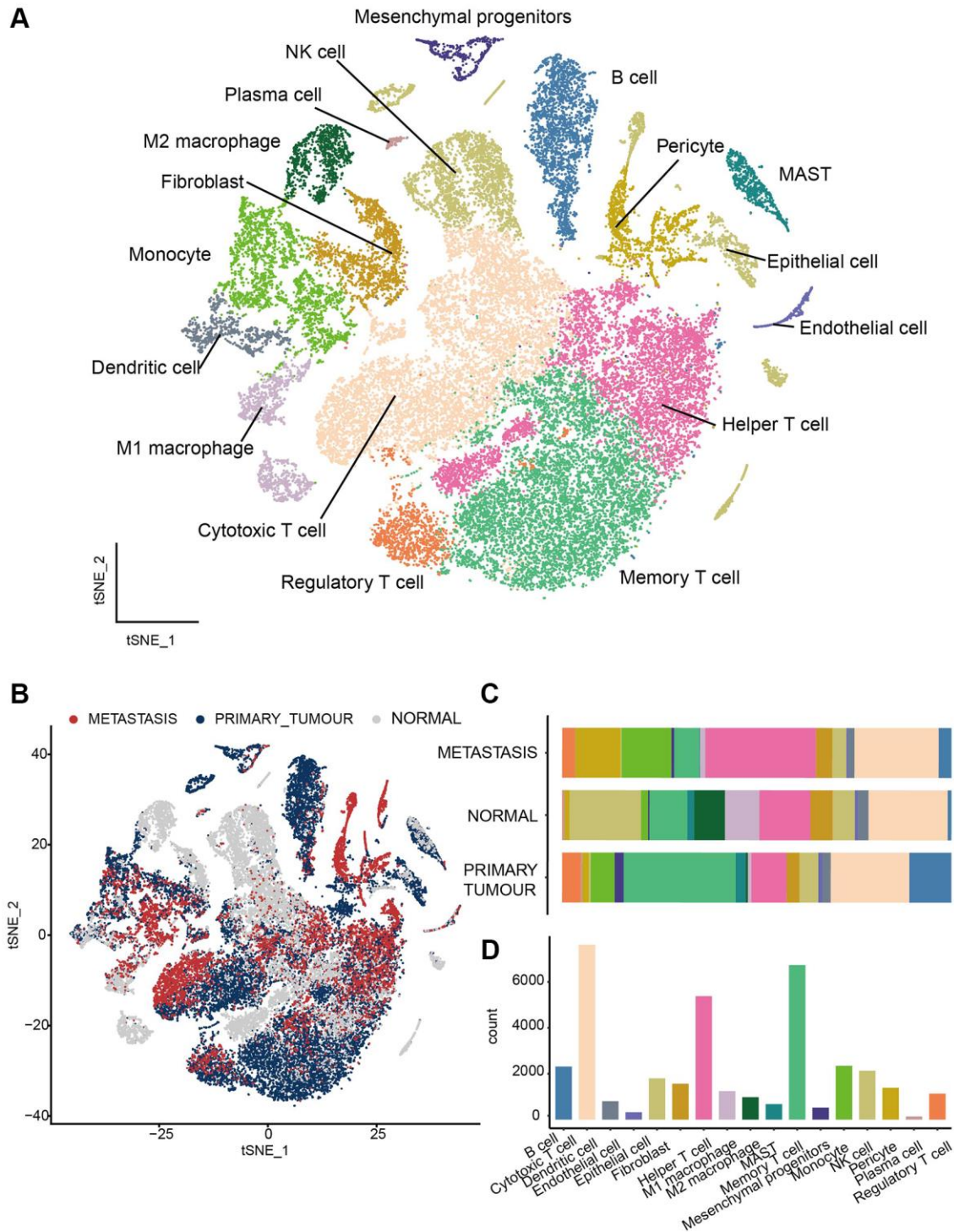


Figure 1. Transcriptome landscape of heterogeneous cell types in LUAD. (A) All the cells were assigned to 17 cell types. (B) The distribution of cell clusters among three groups, including primary tumors, metastasis lesions and normal tissues. (C) The fraction of immune cell types in different sample classes. (D) Histogram to visual Cell numbers in each cell type.

Identification of marker genes for types of immune cells in different sample classes

To further unveil the role of the immune cells in their microenvironment in different sample classes, we then investigated the differential expressed genes (DEGs) in each cell type in pairwise comparison between sample classes and specifically expressed in the cell type at the

same time. These genes were identified as marker genes for the cell groups (see Method). In total 709 genes were identified to play a vital role in the immune microenvironment of LUAD (Figure 2A). Interestingly, Epithelial cell has 629 marker genes, which reach the most, while the number of epithelial cells is only a small part of the total number of cells (5%). The cell with the second highest number of marker genes was

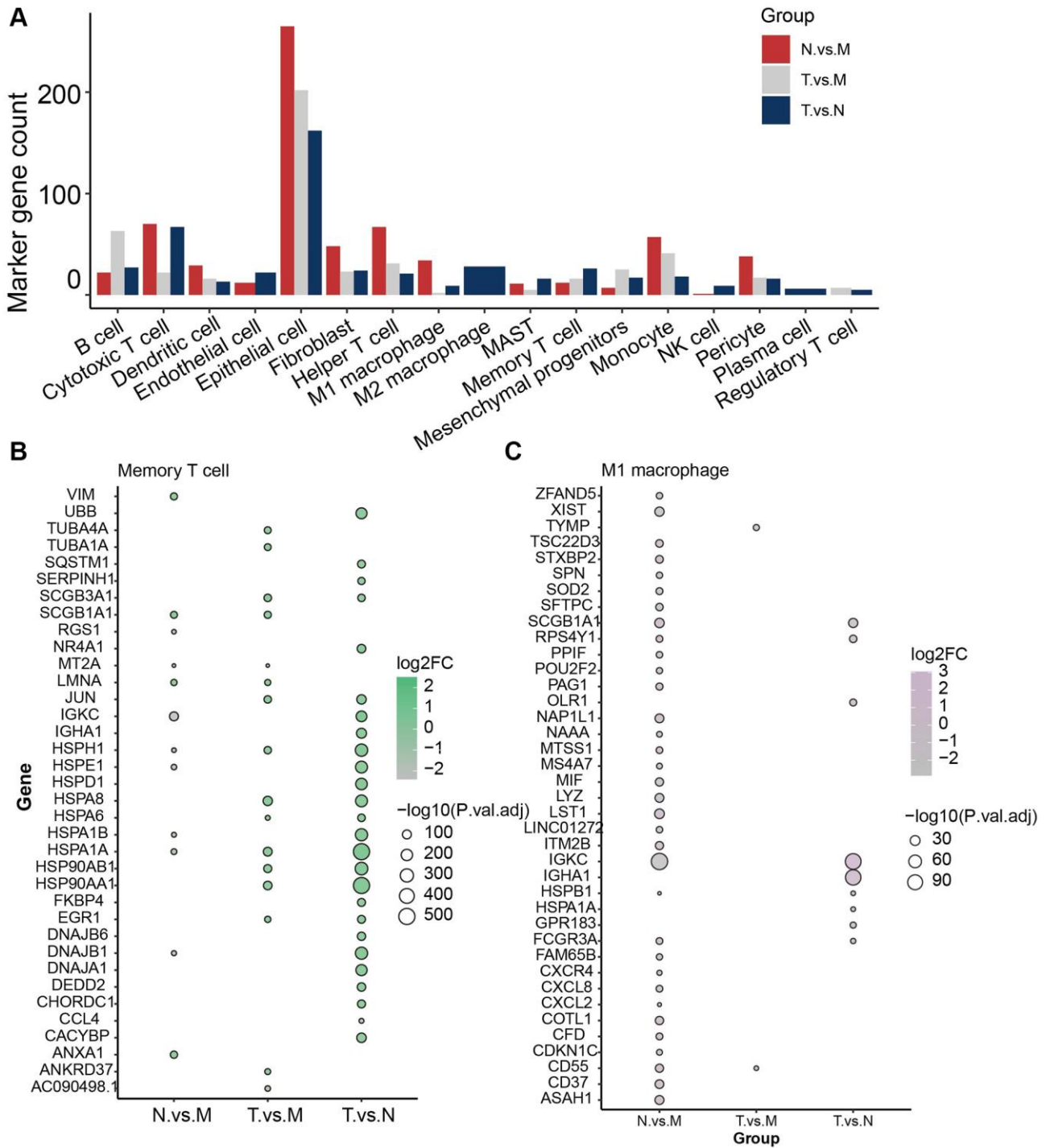


Figure 2. Identification of marker genes for type of immune cells in different sample classes. (A) Overview of total 709 marker genes. **(B)** Marker genes identified in memory T cells. **(C)** Marker genes identified in M1 macrophages.

cytotoxic T cell, with 159 marker genes and 7614 cells (20.6%) (Supplementary Table 2).

Investigation of key factors that induce the immune heterogeneity between different sample classes would provide a glimpse of the different malignancies and metastasis of LUAD patients in different pathological conditions. Memory T cell was the prevalent cell in tumor tissue and 26 genes were identified to be significantly differentially expressed between normal and tumor samples (Figure 2B), which were specifically expressed in memory T cells. HSPA1A and HSP90AA1 were the most significant ones. Whereas JUN, HSPH1, HSPA8, HSPA6, HSP90AB1, and EGR1 were found to be the DEGs in both Tumor vs. normal and tumor vs. metastasis group. In addition, M1/M2 macrophages were enriched in normal tissue (Figure 2C). IGKC, SCGB1A1 HSPB1, and FCGR3A were found to be differentially expressed in normal tissue compared with tumor and metastasis. Especially for metastasis tissue, Helper T cells were the most common cells and 119 markers were identified, such as VIM, ZFP36L2, TSC22D3, RPS29, and PRL41. These results reveal the molecular characteristics of different cell subtypes in different sample classes, thus facilitating our understanding of the microenvironment of LUAD.

Pathways analysis revealed immune cell heterogeneity

Specifically expressed genes contribute significantly to the heterogeneity between different sample classes and we further applied enrichment analysis to the marker genes to explore the biological mechanism affecting the microenvironment of LUAD. The results of GO enrichment analysis suggested that marker genes of M1 macrophage were enriched in immune-related pathways, such as “involved in immune response”, “neutrophil degranulation” and “regulation of mononuclear cell proliferation” (Figure 3A). Besides, markers of M2 macrophage enriched in not only immune-related pathways, such as “cellular response to chemokine”, “neutrophil chemotaxis” and “myeloid leukocyte migration” ion transportation pathways were also emphasized, such as “stress response to meta ion”, “response to zinc ion” and “cellular response to copper ion”(Figure 3B). For non-immune cells, the enrichment results of marker genes show that the cells are mainly related to protein targeting and localization (Figure 3C, 3D). Overall, the enrichment results of marker genes matched the functions of cell types in our cognition and deepened our understanding of the microenvironment of LUAD.

Functional analysis of marker genes showed that immune cells play an important role in affecting the

cancer microenvironment through immune-related pathways. Further GSEA suggested that cytotoxic T cells were significantly enriched in 2 cancer hallmarks and 9 immune signatures from MigDB and indicated to be active in response to tumor necrosis factor (TNF) with the regulation of NF κ B (Figure 4A). In addition, Helper T cells were founded to be enriched in genes positively correlated with memory B cell response (Figure 4B). Monocytes were active in the genes regulated by IL4, while suppressing the activity of polymorphonuclear leukocytes (Figure 4C). For non-immune cells. The marker genes of mesenchymal progenitors enriched in cancer hallmarks of epithelial-mesenchymal transition only. These results suggested broad communications between different subtypes of immune cells.

Cell-cell communication

The enrichment analysis results indicate that there is a wide range of cell-cell interactions between different types of cells and the tumor microenvironment is characterized by these complex interactions. To interrogate putative cell-cell interaction heterogeneity in different sample classes, we applied CellChat, which includes broad ligand-receptor pairs, to gain insights into interactions among all immune cell types. As a result, pericytes, which are predominant in metastasis tissues, were revealed to have broad interactions with B cells, cytotoxic T cells, dendritic cells, endothelial cells, epithelial cells, M1 macrophage, memory T cells, regulatory T cells, monocytes, and NK cells via the MDK-NCL pathways, which is an important hallmark in the change of tumor microenvironment, and serves as a sign of cancer-associated fibroblasts (CAF) activation to stimulate downstream pathways for facilitating tumor invasion [42]. What's more, Endothelial cells and epithelial cells were also found to be active in MDK-NCL pathways in communication with immune cells (Figure 5A, 5B). Besides, MIF-(CD74+CXCR4) and MIF-(CD74+CD44) pathways were widespread and are associated with multiple cell types in the tumor and normal tissues. Macrophage migration inhibitory factor (MIF) is an inflammatory cytokine that exhibits chemokine-like activities through non-cognate interactions with the chemokine receptors CXCR2 and CXCR4, in addition to activating the type II receptor CD74. Activation of the MIF-CXCR4 axes promotes leukocyte recruitment, mediating the exacerbating role of MIF in atherosclerosis [43]. MIF-(CD74+CXCR4) and MIF-(CD74+CC44) interactions happened in dendritic cells, M1 macrophage, M2 macrophage, monocyte, and plasma cells both in tumor and normal tissues (Figure 5C, 5D). Except for the same cell type in normal tissue, tumor tissues were also found B cell, cytotoxic T cell, MAST cell, memory T cell, and

regulatory T cell to have the interaction (Figure 5E, 5F), which uncovers the change of microenvironment between tumor and normal tissues.

Regulatory element analysis

To further dissect the immune heterogeneity of gene regulation, the single-cell regulatory network inference and clustering (SCENIC) analysis for DEGs was performed to assess the differences in the expression levels of transcription factors (TFs) between different sample classes. In general, metastasis, normal, and tumor tissues contained a similar number of regulons per cell (RPC), while in normal tissues, the number of

cells per regulon (CPR) was significantly higher (Figure 6A, 6C, 6E), which indicate that the microenvironment was more complex in metastasis and tumor tissue. The most five active regulators were identical in tumor and normal tissues, SPI1, CEBPB, JUNB, FOS, and KLF4 (Figure 6B, 6D, 6F). While in metastasis tissue, ATF3 was the third active regulator, which is a stress-induced transcription factor that plays a vital role in modulating metabolism, immunity, and oncogenesis [44].

Immune cell abundance prediction

To further unveil the role of the immune cells in their microenvironment in different sample classes, we then

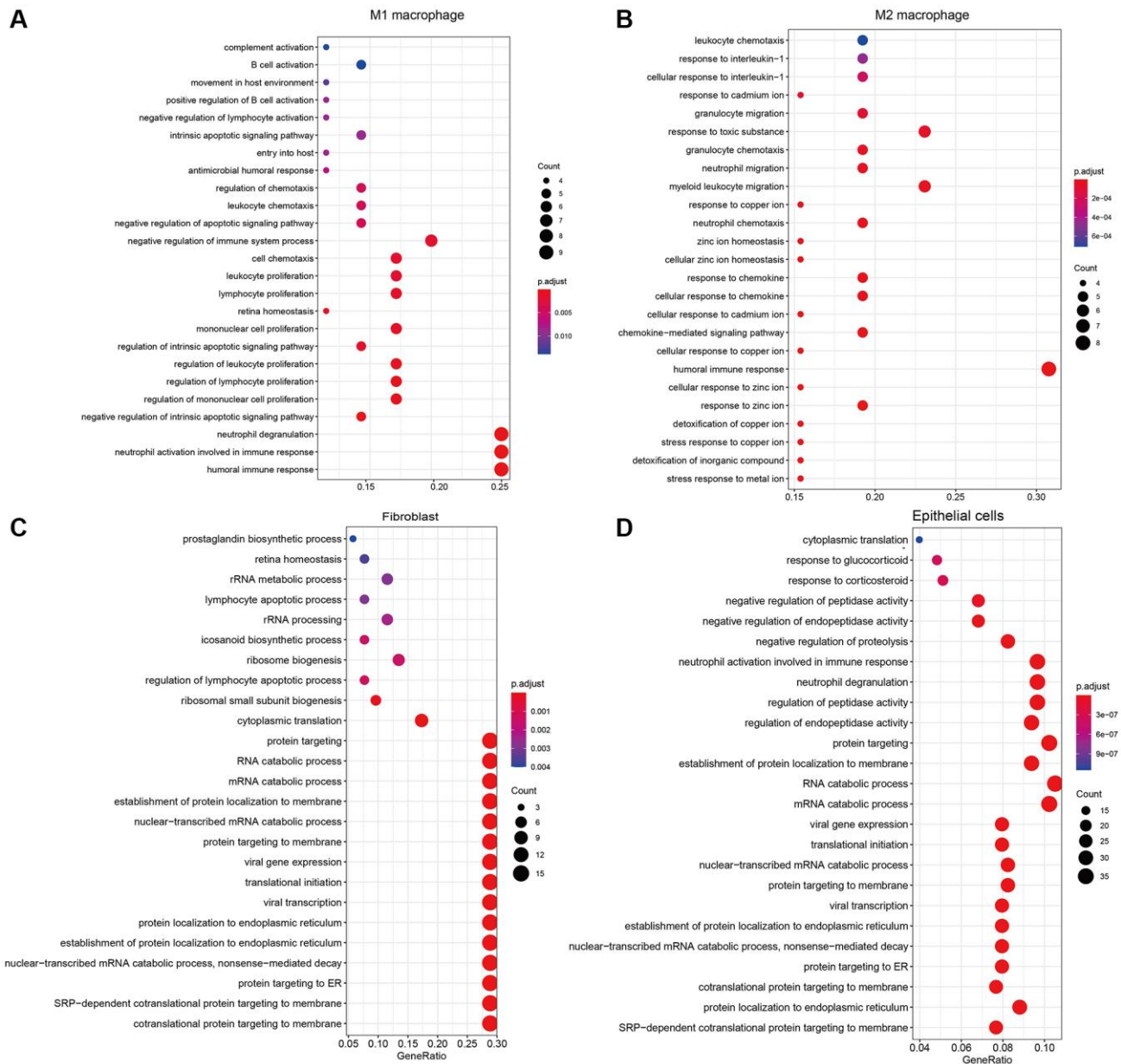


Figure 3. The results of GO enrichment analysis for marker genes of (A) M1 macrophages, (B) M2 macrophages, (C) Fibroblasts, (D) Epithelial cells.

investigated the immune cell type composition through ssGSEA algorithm (see Method). As a result, the marker genes of Naïve CD4 T cell, Naïve CD8 T cell, exhausted T cell, T central memory (Tcm), T effector memory (Tem), B cell, and monocyte had relatively high enrichment fraction in all cells (Figure 7A–7C). The proportion of cytotoxic T cells in each sample was similar, but the analysis of immune cell abundance showed that the enrichment fraction of CD4+ T cells with cytotoxic activity was significantly higher in cytotoxic T cells than in other immune cells, like regulatory T cells and memory T cells

(Figure 7D). Besides, the enrichment scores of all the immune cells in Cytotoxic T cells were significantly higher in tumor tissues than in metastasis and normal tissue (Figure 7E). Unsurprisingly, monocytes had higher enrichment scores for marker genes in monocyte, M1 macrophage, and M2 macrophage (Figure 7F). Besides, monocytes also had significantly high enrichment scores in type 2 conventional dendritic cells, while the scores of memory dendritic cells in monocyte were significantly low (Figure 7G–7I). These results indicated that monocyte may exhibit vital biological functions in LUAD.

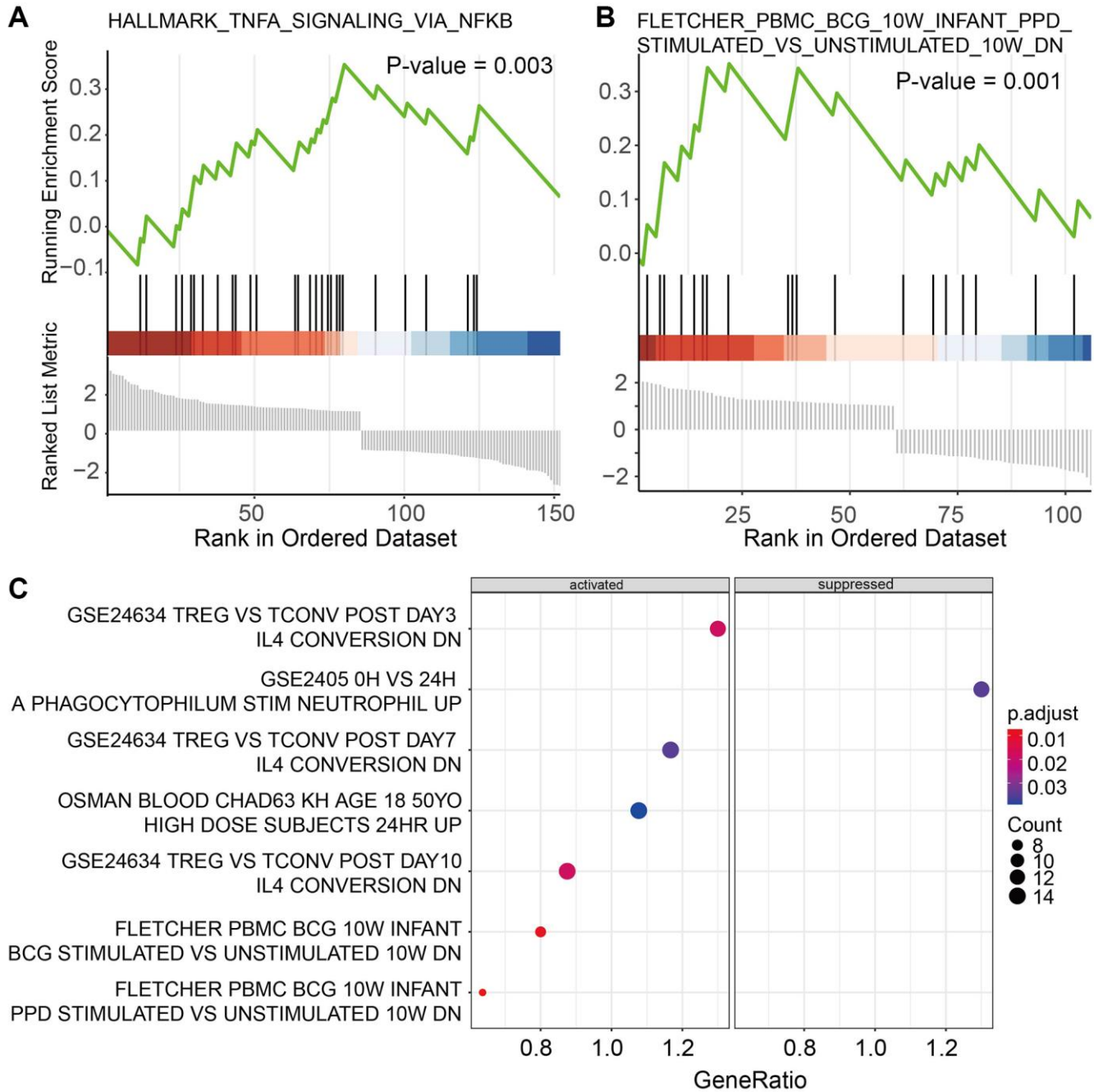


Figure 4. Functional analysis of marker genes for (A) cytotoxic T cells, (B) Helper T cells, and (C) Monocytes.

Marker gene validation in bulk RNA-seq

To further explore the impact of cell-type specifically expressed marker genes on LUAD prognosis, mRNA expression matrix and clinical information of TCGA-LUAD were collected and applied to survival analysis. The results of survival analysis suggested that CCL20, a marker of M2 macrophage, is the most significant gene related to LUAD prognosis (p -value = 0.00048) (Figure 8A) and patients with low expression of CCL20 have a better prognosis. Besides, TPI1 is the marker gene of Pericyte and patients with low expression of TPI1 have a better prognosis (p -value = 0.00053) (Figure 8B). Epithelial cells have 16 markers genes significantly

correlated with LUAD prognosis, including CHRNA5 (p -value = 0.0015) (Figure 8C), GAPDH (p -value = 0.0018) (Figure 8D), ERO1B (p -value = 0.0019) (Figure 8E), and CLIC6 (p -value = 0.0067) (Figure 8F). In general, 25 marker genes of 10 cell types were identified to be significantly related to LUAD prognosis and these genes provide molecular insight into the function of specific cell types in LUAD pathology (Supplementary Table 3).

DISCUSSION

In this study, we collected normal, tumor, and metastasis LUAD datasets in single-cell sequencing and

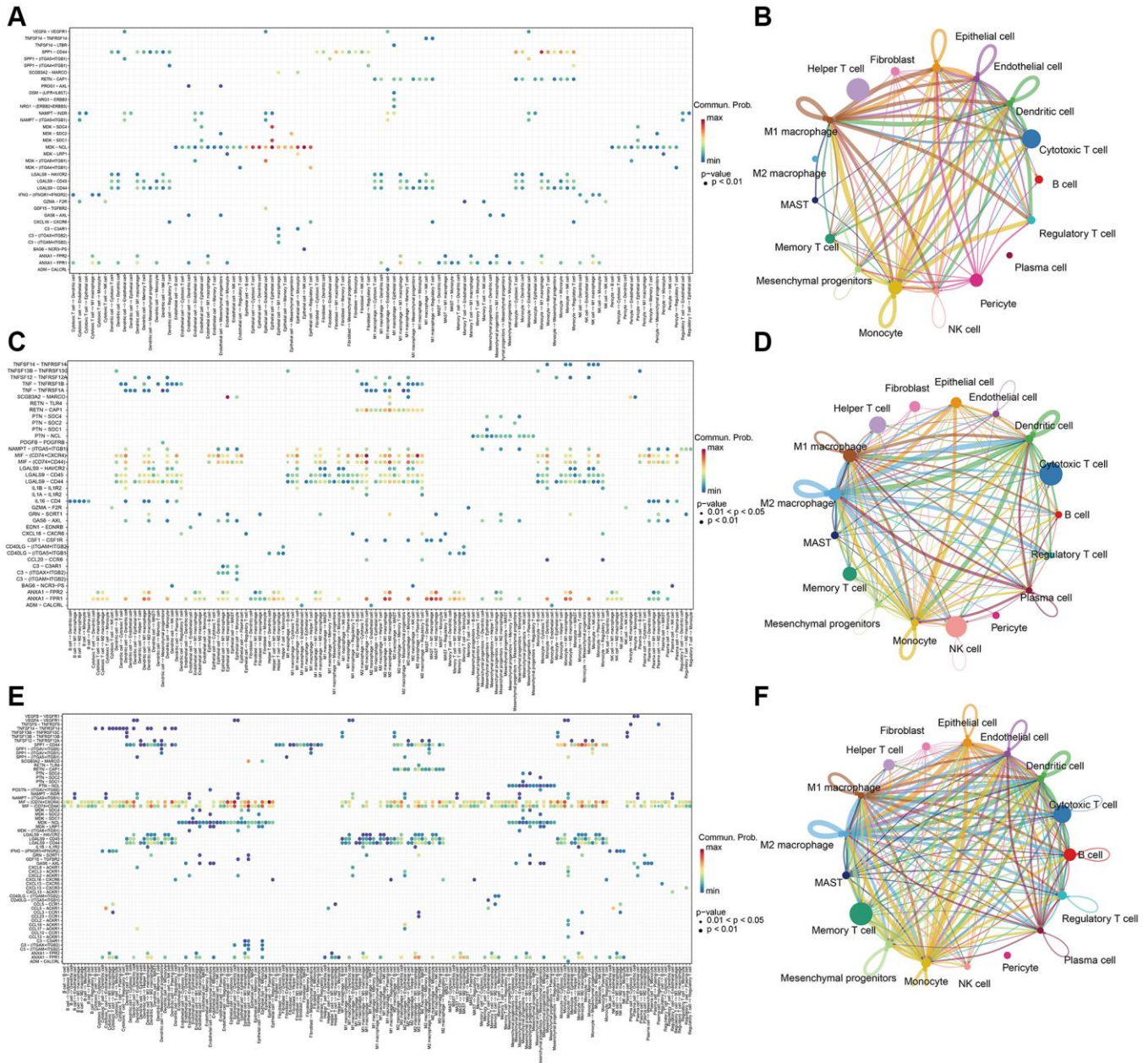


Figure 5. Cell-cell communication analysis with CellChat in (A, B) metastasis samples, (C, D) Normal samples, and (E, F) tumor samples.

applied bioinformatic methods to explore the distinct molecular features of different sample types. As a result, the cellular landscapes of different subclasses of LUAD were depicted. With the bioinformatic method (see Method), marker genes of each cell type in LUAD were identified and enrichment analysis was used to explore the biological function of these marker genes. Combined with trajectory analysis and cell-cell communication analysis, the different micro-environments between different sample classes were characterized. For example, CCL20 encodes protein displaying chemotactic activity for lymphocytes and can repress the proliferation of myeloid progenitors. Previous research reported that high CCL20 expression is a poor prognostic marker for LUAD patients, and is associated with enhanced epithelial-mesenchymal transition (EMT), inflammatory response, and activated TNF pathway in LUAD. While low CCL20 expression blocked the detrimental effects of high TGF- β on survival and effectively improved patients' response to anti-PD-L1 therapy [45]. In our study, CCL20 is specifically expressed in M2 macrophage, and the expression in normal samples is much higher than in other samples, which indicates that M2 macrophage plays an important role in LUAD through the function of CCL20. Further enrichment

analysis of marker genes of LUAD revealed that the significantly enriched GO terms were chemokine-mediated signaling pathway, myeloid leukocyte migration, positive regulation of leukocyte migration, calcium-mediated signaling, and positive regulation of ERK1/ERK2 cascade. Besides, IL-17 signaling pathway and TNF signaling pathway were the most significant KEGG pathways. Marker genes shared in these pathways were TREM2, CXCL3, CCL20, IL1B, CXCL8, and CCL4. Then, the result of cell-cell communication analysis illustrated different signal pathways in different sample classes. In normal samples, CCL20-CCR6 interaction occurs in M2 macrophage->DC, M2->Treg. The ligand-receptor pair CCL20-CCR6 is responsible for the chemotaxis of dendritic cells (DC), effector/memory T-cells, and B-cells, which plays an important role at skin and mucosal surfaces under homeostatic and inflammatory conditions, as well as in pathology, including cancer and various autoimmune diseases. While in meta samples, macrophage function is weaker. In conclusion, this study has revealed the characteristic of normal, tumor, and metastasis LUAD tissues and associated immune microenvironment, and further illuminated changes in cellular and molecular dynamics in different sample classes.

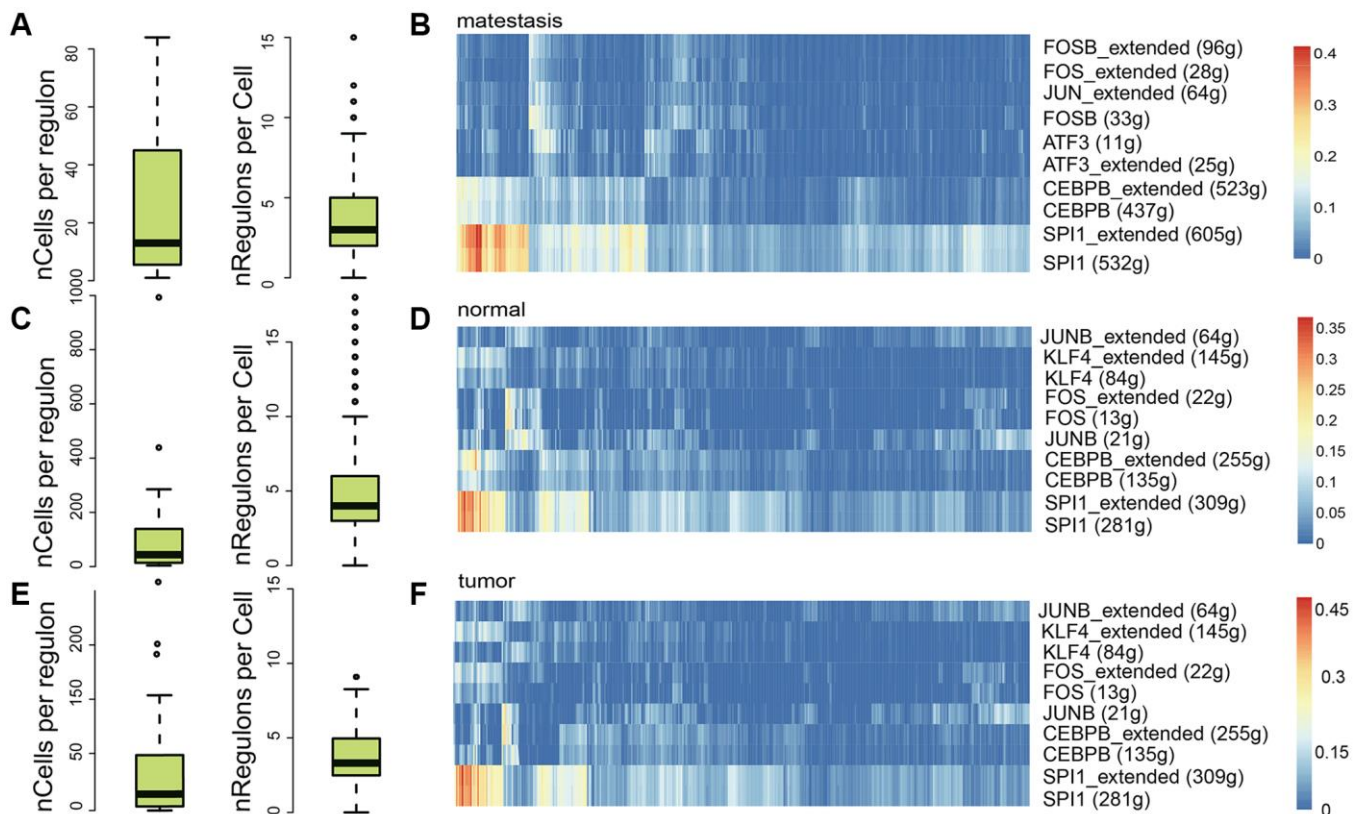


Figure 6. Regulatory element analysis with SCENIC in different sample classes. The number of cells per regulon and the regulon activity in metastasis (A, B), normal (C, D), and tumor (E, F) samples, respectively.

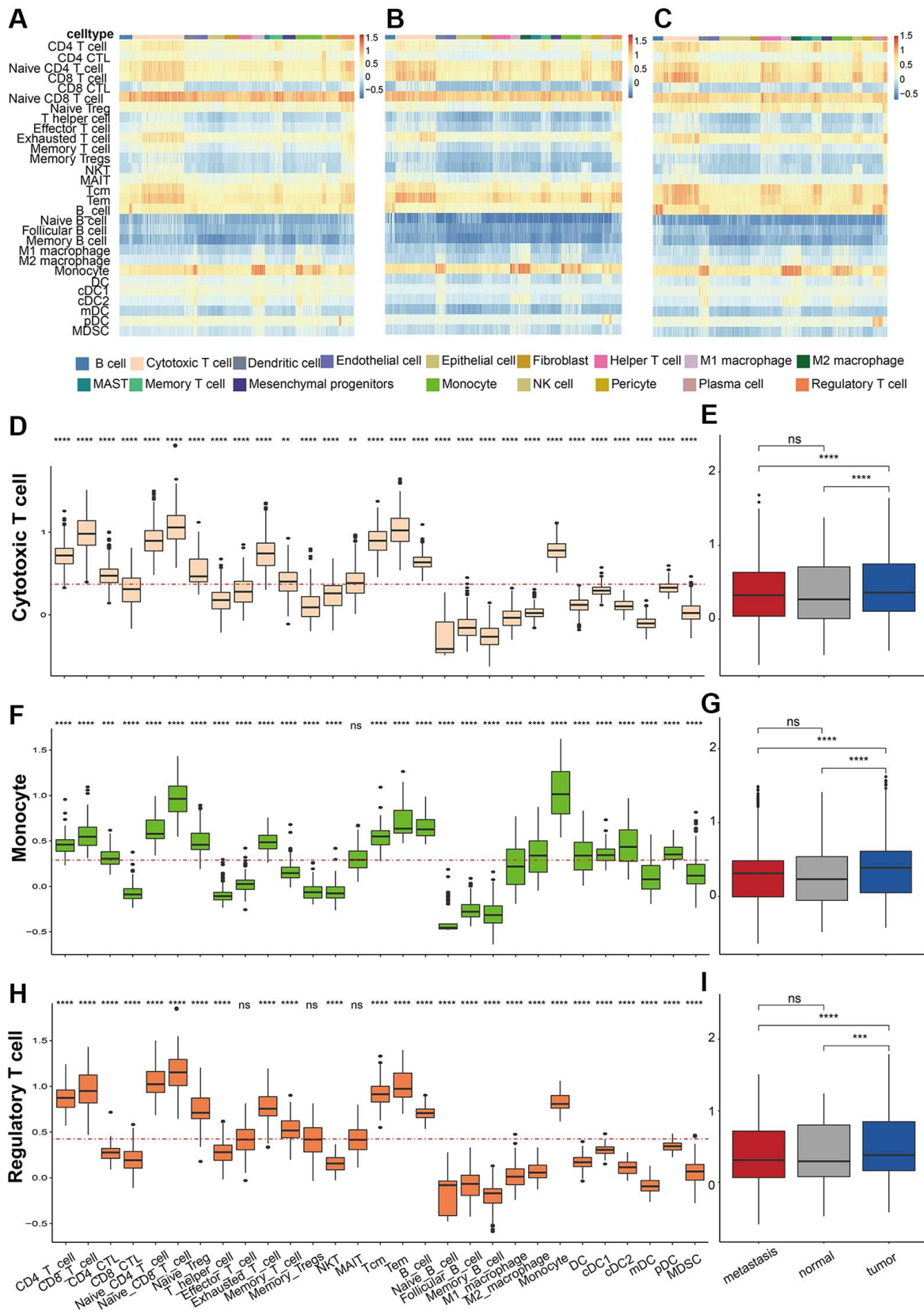


Figure 7. Immune cell abundance prediction. Heatmap of Estimate score in (A) metastasis, (B) normal samples, and (C) tumor samples in different cells. The predicted estimate score of cytotoxic T cell markers in (D) different cells and (E) samples. The predicted estimate score of monocyte markers in (F) different cells and (G) samples. The predicted estimate score of regulatory T cell markers in (H) different cells and (I) samples. *P* values were shown as: **p* < 0.05, ***p* < 0.01, and ****p* < 0.001.

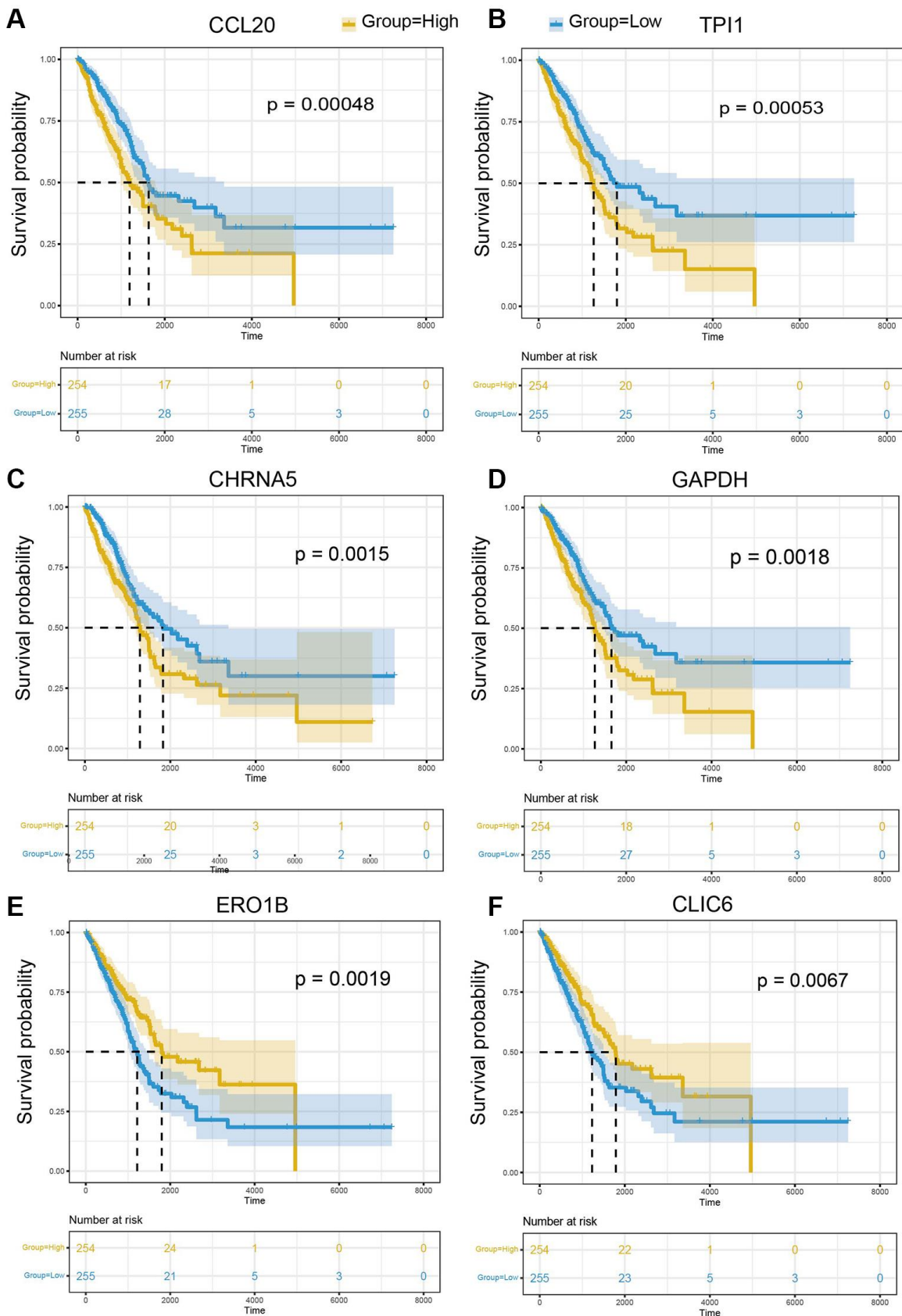


Figure 8. The prognostic significance of marker genes in LUAD. Kaplan-Meier survival curves of (A) CCL20, (B) TPI1, (C) CHRNA5, (D) GAPDH, (E) ERO1B, (F) CLIC6 in LUAD. The median of gene expression is utilized as the cut-off. Yellow: higher expression of indicated gene; Blue: lower expression of indicated gene.

AUTHOR CONTRIBUTIONS

Jiakuan Chen formulated the study concept and design; Haiqiang Wang wrote the original manuscript draft; Guohan Liang do the bioinformatical analyses; Jiakuan Chen and Haiqiang Wang critically revised the manuscript. All authors contributed to the article and approved the submitted version.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest related to this study.

FUNDING

No funding was provided for this study.

REFERENCES

1. Ren C, Li J, Zhou Y, Zhang S, Wang Q. Typical tumor immune microenvironment status determine prognosis in lung adenocarcinoma. *Transl Oncol.* 2022; 18:101367. <https://doi.org/10.1016/j.tranon.2022.101367> PMID:35176624
2. Zhao G, Song D, Wu J, Yang S, Shi S, Cui X, Ren H, Zhang B. Identification of OTUD6B as a new biomarker for prognosis and immunotherapy by pan-cancer analysis. *Front Immunol.* 2022; 13:955091. <https://doi.org/10.3389/fimmu.2022.955091> PMID:36052059
3. Gao F, Liu J, Gan H. The expression of POSTN and immune cell infiltration are prognostic factors of lung adenocarcinoma. *Medicine (Baltimore).* 2022; 101:e30187. <https://doi.org/10.1097/MD.00000000000030187> PMID:36042584
4. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature.* 2012; 489:519–25. <https://doi.org/10.1038/nature11404> PMID:22960745
5. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature.* 2014; 511:543–50. <https://doi.org/10.1038/nature13385> PMID:25079552
6. Meurette O, Mehlen P. Notch Signaling in the Tumor Microenvironment. *Cancer Cell.* 2018; 34:536–48. <https://doi.org/10.1016/j.ccell.2018.07.009> PMID:30146333
7. Remark R, Becker C, Gomez JE, Damotte D, Dieu-Nosjean MC, Sautès-Fridman C, Fridman WH, Powell CA, Altorki NK, Merad M, Gnjatic S. The non-small cell lung cancer immune contexture. A major determinant of tumor characteristics and patient outcome. *Am J Respir Crit Care Med.* 2015; 191:377–90. <https://doi.org/10.1164/rccm.201409-1671PP> PMID:25369536
8. Gordon SR, Maute RL, Dulken BW, Hutter G, George BM, McCracken MN, Gupta R, Tsai JM, Sinha R, Corey D, Ring AM, Connolly AJ, Weissman IL. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. *Nature.* 2017; 545:495–9. <https://doi.org/10.1038/nature22396> PMID:28514441
9. Huntington ND, Cursons J, Rautela J. The cancer-natural killer cell immunity cycle. *Nat Rev Cancer.* 2020; 20:437–54. <https://doi.org/10.1038/s41568-020-0272-z> PMID:32581320
10. Yang L, He YT, Dong S, Wei XW, Chen ZH, Zhang B, Chen WD, Yang XR, Wang F, Shang XM, Zhong WZ, Wu YL, Zhou Q. Single-cell transcriptome analysis revealed a suppressive tumor immune microenvironment in EGFR mutant lung adenocarcinoma. *J Immunother Cancer.* 2022; 10:e003534. <https://doi.org/10.1136/jitc-2021-003534> PMID:35140113
11. He D, Wang D, Lu P, Yang N, Xue Z, Zhu X, Zhang P, Fan G. Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene.* 2021; 40:355–68. <https://doi.org/10.1038/s41388-020-01528-0> PMID:33144684
12. Song P, Li W, Wu X, Qian Z, Ying J, Gao S, He J. Integrated analysis of single-cell and bulk RNA-sequencing identifies a signature based on B cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma. *Cancer Immunol Immunother.* 2022; 71:2341–54. <https://doi.org/10.1007/s00262-022-03143-2> PMID:35152302
13. Song P, Li W, Guo L, Ying J, Gao S, He J. Identification and Validation of a Novel Signature Based on NK Cell Marker Genes to Predict Prognosis and Immunotherapy Response in Lung Adenocarcinoma by Integrated Analysis of Single-Cell and Bulk RNA-Sequencing. *Front Immunol.* 2022; 13:850745. <https://doi.org/10.3389/fimmu.2022.850745> PMID:35757748
14. Laughney AM, Hu J, Campbell NR, Bakhoun SF, Setty M, Lavallée VP, Xie Y, Masilionis I, Carr AJ, Kottapalli S, Allaj V, Mattar M, Rekhtman N, et al. Regenerative

- lineages and immune-mediated pruning in lung cancer metastasis. *Nat Med.* 2020; 26:259–69.
<https://doi.org/10.1038/s41591-019-0750-6>
PMID:[32042191](https://pubmed.ncbi.nlm.nih.gov/32042191/)
15. Clough E, Barrett T. The Gene Expression Omnibus Database. *Methods Mol Biol.* 2016; 1418:93–110.
https://doi.org/10.1007/978-1-4939-3578-9_5
PMID:[27008011](https://pubmed.ncbi.nlm.nih.gov/27008011/)
 16. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM, and Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; 45:1113–20.
<https://doi.org/10.1038/ng.2764>
PMID:[24071849](https://pubmed.ncbi.nlm.nih.gov/24071849/)
 17. Navarro Gonzalez J, Zweig AS, Speir ML, Schmelter D, Rosenbloom KR, Raney BJ, Powell CC, Nassar LR, Maulding ND, Lee CM, Lee BT, Hinrichs AS, Fyfe AC, et al. The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.* 2021; 49:D1046–57.
<https://doi.org/10.1093/nar/gkaa1070>
PMID:[33221922](https://pubmed.ncbi.nlm.nih.gov/33221922/)
 18. Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. *J Vis Exp.* 2021.
<https://doi.org/10.3791/62528>
PMID:[34605806](https://pubmed.ncbi.nlm.nih.gov/34605806/)
 19. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics.* 2013; 14:7.
<https://doi.org/10.1186/1471-2105-14-7>
PMID:[23323831](https://pubmed.ncbi.nlm.nih.gov/23323831/)
 20. Powers RK, Goodspeed A, Pielke-Lombardo H, Tan AC, Costello JC. GSEA-InContext: identifying novel and common patterns in expression experiments. *Bioinformatics.* 2018; 34:i555–64.
<https://doi.org/10.1093/bioinformatics/bty271>
PMID:[29950010](https://pubmed.ncbi.nlm.nih.gov/29950010/)
 21. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb).* 2021; 2:100141.
<https://doi.org/10.1016/j.xinn.2021.100141>
PMID:[34557778](https://pubmed.ncbi.nlm.nih.gov/34557778/)
 22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet.* 2000; 25:25–9.
<https://doi.org/10.1038/75556>
PMID:[10802651](https://pubmed.ncbi.nlm.nih.gov/10802651/)
 23. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000; 28:27–30.
<https://doi.org/10.1093/nar/28.1.27>
PMID:[10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)
 24. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015; 1:417–25.
<https://doi.org/10.1016/j.cels.2015.12.004>
PMID:[26771021](https://pubmed.ncbi.nlm.nih.gov/26771021/)
 25. Xiao B, Liu L, Li A, Xiang C, Wang P, Li H, Xiao T. Identification and Verification of Immune-Related Gene Prognostic Signature Based on ssGSEA for Osteosarcoma. *Front Oncol.* 2020; 10:607622.
<https://doi.org/10.3389/fonc.2020.607622>
PMID:[33384961](https://pubmed.ncbi.nlm.nih.gov/33384961/)
 26. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun.* 2013; 4:2612.
<https://doi.org/10.1038/ncomms3612>
PMID:[24113773](https://pubmed.ncbi.nlm.nih.gov/24113773/)
 27. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol.* 2014; 21:686–94.
<https://doi.org/10.1007/s12350-014-9908-2>
PMID:[24810431](https://pubmed.ncbi.nlm.nih.gov/24810431/)
 28. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, van den Oord J, Atak ZK, Wouters J, Aerts S. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods.* 2017; 14:1083–6.
<https://doi.org/10.1038/nmeth.4463>
PMID:[28991892](https://pubmed.ncbi.nlm.nih.gov/28991892/)
 29. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol.* 2014; 32:381–6.
<https://doi.org/10.1038/nbt.2859>
PMID:[24658644](https://pubmed.ncbi.nlm.nih.gov/24658644/)
 30. Stubbington MJT, Rozenblatt-Rosen O, Regev A, Teichmann SA. Single-cell transcriptomics to explore the immune system in health and disease. *Science.* 2017; 358:58–63.
<https://doi.org/10.1126/science.aan6828>
PMID:[28983043](https://pubmed.ncbi.nlm.nih.gov/28983043/)
 31. Giladi A, Cohen M, Medaglia C, Baran Y, Li B, Zada M, Bost P, Blecher-Gonen R, Salame TM, Mayer JU, David E, Ronchese F, Tanay A, Amit I. Dissecting cellular

- crosstalk by sequencing physically interacting cells. *Nat Biotechnol.* 2020; 38:629–37.
<https://doi.org/10.1038/s41587-020-0442-2>
PMID:[32152598](https://pubmed.ncbi.nlm.nih.gov/32152598/)
32. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q. Inference and analysis of cell-cell communication using CellChat. *Nat Commun.* 2021; 12:1088.
<https://doi.org/10.1038/s41467-021-21246-9>
PMID:[33597522](https://pubmed.ncbi.nlm.nih.gov/33597522/)
33. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, Ping Y, Li F, Shi A, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* 2019; 47:D721–8.
<https://doi.org/10.1093/nar/gky900>
PMID:[30289549](https://pubmed.ncbi.nlm.nih.gov/30289549/)
34. Blom B, Spits H. Development of human lymphoid cells. *Annu Rev Immunol.* 2006; 24:287–320.
<https://doi.org/10.1146/annurev.immunol.24.021605.090612>
PMID:[16551251](https://pubmed.ncbi.nlm.nih.gov/16551251/)
35. Welner RS, Pelayo R, Kincade PW. Evolving views on the genealogy of B cells. *Nat Rev Immunol.* 2008; 8:95–106.
<https://doi.org/10.1038/nri2234>
PMID:[18204470](https://pubmed.ncbi.nlm.nih.gov/18204470/)
36. Farag SS, Caligiuri MA. Human natural killer cell development and biology. *Blood Rev.* 2006; 20:123–37.
<https://doi.org/10.1016/j.blre.2005.10.001>
PMID:[16364519](https://pubmed.ncbi.nlm.nih.gov/16364519/)
37. Dahlin JS, Hallgren J. Mast cell progenitors: origin, development and migration to tissues. *Mol Immunol.* 2015; 63:9–17.
<https://doi.org/10.1016/j.molimm.2014.01.018>
PMID:[24598075](https://pubmed.ncbi.nlm.nih.gov/24598075/)
38. Terry RL, Miller SD. Molecular control of monocyte development. *Cell Immunol.* 2014; 291:16–21.
<https://doi.org/10.1016/j.cellimm.2014.02.008>
PMID:[24709055](https://pubmed.ncbi.nlm.nih.gov/24709055/)
39. Murray PJ, Wynn TA. Protective and pathogenic functions of macrophage subsets. *Nat Rev Immunol.* 2011; 11:723–37.
<https://doi.org/10.1038/nri3073>
PMID:[21997792](https://pubmed.ncbi.nlm.nih.gov/21997792/)
40. Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG, Treutlein B, Tan K, Tan V, Liu FC, Looney MR, Matthay MA, Rock JR, Chapman HA. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature.* 2015; 517:621–5.
<https://doi.org/10.1038/nature14112>
PMID:[25533958](https://pubmed.ncbi.nlm.nih.gov/25533958/)
41. Wang C, Yu Q, Song T, Wang Z, Song L, Yang Y, Shao J, Li J, Ni Y, Chao N, Zhang L, Li W. The heterogeneous immune landscape between lung adenocarcinoma and squamous carcinoma revealed by single-cell RNA sequencing. *Signal Transduct Target Ther.* 2022; 7:289.
<https://doi.org/10.1038/s41392-022-01130-8>
PMID:[36008393](https://pubmed.ncbi.nlm.nih.gov/36008393/)
42. Liu XY, Liu YB, Xu JC, Zhang YF, Ruan YY, Zhao Y, Wu LF, Hu JW, Zhang Z, He MJ, Chen TY, Xu XY, Zhang JW, et al. Single-cell transcriptomic analysis deciphers key transitional signatures associated with oncogenic evolution in human intramucosal oesophageal squamous cell carcinoma. *Clin Transl Med.* 2023; 13:e1203.
<https://doi.org/10.1002/ctm2.1203>
PMID:[36855810](https://pubmed.ncbi.nlm.nih.gov/36855810/)
43. Rajasekaran D, Gröning S, Schmitz C, Zierow S, Drucker N, Bakou M, Kohl K, Mertens A, Lue H, Weber C, Xiao A, Luker G, Kapurniotu A, et al. Macrophage Migration Inhibitory Factor-CXCR4 Receptor Interactions: EVIDENCE FOR PARTIAL ALLOSTERIC AGONISM IN COMPARISON WITH CXCL12 CHEMOKINE. *J Biol Chem.* 2016; 291:15881–95.
<https://doi.org/10.1074/jbc.M116.717751>
PMID:[27226569](https://pubmed.ncbi.nlm.nih.gov/27226569/)
44. Ku HC, Cheng CF. Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Front Endocrinol (Lausanne).* 2020; 11:556.
<https://doi.org/10.3389/fendo.2020.00556>
PMID:[32922364](https://pubmed.ncbi.nlm.nih.gov/32922364/)
45. Fan T, Li S, Xiao C, Tian H, Zheng Y, Liu Y, Li C, He J. CCL20 promotes lung adenocarcinoma progression by driving epithelial-mesenchymal transition. *Int J Biol Sci.* 2022; 18:4275–88.
<https://doi.org/10.7150/ijbs.73275>
PMID:[35864953](https://pubmed.ncbi.nlm.nih.gov/35864953/)

SUPPLEMENTARY MATERIALS

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 2 and 3.

Supplementary Table 1. Marker gene list.

Cell type	Marker genes
B cell	CD19, MS4A1, CD79A, CD79B, SPIB
Cytotoxic T cell	CD3D, CD3G, CD8A, CD8B
Dendritic cell	XCR1, CLEC9A, THBD, CD1C, CD1E, FCER1A
Endothelial cell	CDH5, PECAM1, VWF
Epithelial cell	EPCAM, CDH1
Fibroblast	LUM, DCN, COL1A2, COL1A1
Helper T cell	CD3D, CD3G, CD4, TBX21, IFNG, IL4, CXCR5
M1 macrophage	CD68, IL1B, IL6, TNF
M2 macrophage	MRC1, CD163, TGFB1, IL10, FN1
MAST	TPSAB1, TPSB2, CPA3
Memory T cell	PTPRC, CCR7, SELL
Mesenchymal progenitors	FDGFRA, LGR5
Monocyte	CD14, FCN1, FCGR3A, VCAN
NK cell	KLRD1, NKG7, KLRC1, FCGR3A, CD3D-, CD3G-
Pericyte	RGS5, ACTA2, MCAM, PDGFRB, NOTCH3
Plasma cell	CD79A, JCHAIN, MZB1, IGHG1
Regulatory T cell	FOXP3, IL2RA, CTLA4, IKZF2, CD25

Supplementary Table 2. Marker genes in each cell type.

Supplementary Table 3. Survival analysis of all the marker genes.