



Research article

Gene features of tumor-specific T cells relevant to immunotherapy, targeted therapy and chemotherapy in lung cancer

Ziwei Luo ^{a,b,1}, Xuefei Liu ^{c,f,1}, Ying Chen ^{b,1}, Lize Shen ^d, Hui Qin ^a, Qiongfang Zha ^a, Feng Hu ^e, Yali Wang ^{a,*}^a Department of Respiratory Medicine, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, 200120, China^b State Key Laboratory of Oncology in Southern China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, 510060, China^c School of Medicine, Southern University of Science and Technology, Shenzhen, Guangdong, 518055, China^d LC-Bio Technology Co.ltd, Hangzhou, 310018, China^e Department of Cardiology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, 200120, China^f Shenzhen Institute of Pediatrics, Shenzhen Children's Hospital, Shenzhen, Guangdong, 518038, China

ARTICLE INFO

Keywords:

Lung cancer
Tumor-specific T cell
Combination therapies
Immune infiltration
Immunotherapy

ABSTRACT

(1) *Background:* In lung cancer, the use of small-molecule inhibitors, chemotherapy and immunotherapy has led to unprecedented survival benefits in selected patients. Considering most patients will experience a relapse within a short period of time due to single drug resistance, combination therapy is also particularly important to improve patient prognosis. Therefore, more robust biomarkers to predict responses to immunotherapy, targeted therapy, chemotherapy and rationally drug combination therapies may be helpful in clinical treatment choices.

(2) *Methods:* We defined tumor-specific T cells (TSTs) and their features (TSTGs) by single-cell RNA sequencing. We applied LASSO regression to filter out the most survival-relevant TSTGs to form the Tumor-specific T cell score (TSTS). Immunological characteristics, enriched pathways, and mutation were evaluated in high- and low TSTS groups.

(3) *Results:* We identified six clusters of T cells as TSTs in lung cancer, and four most robust genes from 9 feature genes expressed only on tumor-specific T cells were screened to construct a tumor-specific T cells score (TSTS). TSTS was positively correlated with immune infiltration and angiogenesis and negatively correlated with malignant cell proliferation. Moreover, potential vascular-immune crosstalk in lung cancer provides the theoretical basis for combined anti-angiogenic and immunotherapy. Noticeable, patients in high TSTS had better response to ICB and targeted therapy and patients in the low TSTS group often benefit from chemotherapy.

(4) *Conclusion:* The proposed TSTS is a promising indicator to predict immunotherapy, targeted therapy and chemotherapy responses in lung cancer patients for helping clinical treatment choices.

* Corresponding author.

E-mail addresses: luozw@sysucc.org.cn (Z. Luo), 12231372@mail.sustech.edu.cn (X. Liu), cheny_sysucc@163.com (Y. Chen), lzshen@lc-bio.com (L. Shen), drqinhui@126.com (H. Qin), drzhaqiongfang@sina.com (Q. Zha), huf2007@163.com (F. Hu), wyl1586135@126.com (Y. Wang).¹ These authors contributed equally as first authors.<https://doi.org/10.1016/j.heliyon.2024.e28374>

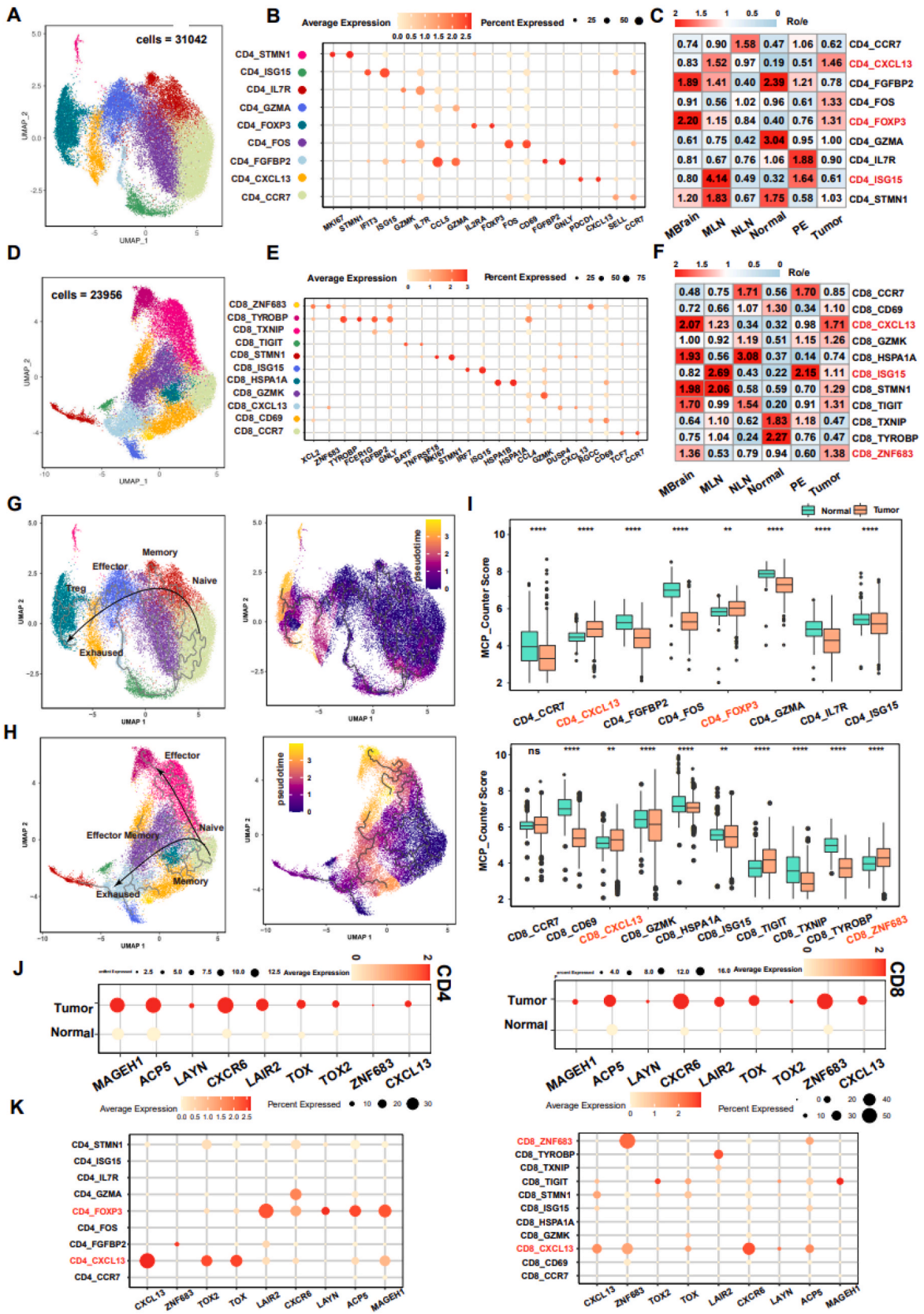
Received 11 December 2022; Received in revised form 14 March 2024; Accepted 18 March 2024

Available online 22 March 2024

2405-8440/© 2024 Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

<http://creativecommons.org/licenses/by-nc-nd/4.0/>.



(caption on next page)

Fig. 1. ScRNAseq identified tumor-specific T cell clusters and tumor-specific genes in lung cancer (A) UMAP projection of CD4⁺ T cells. Cells were color-coded for distinct clusters. (B) Dotplot exhibited the expression of marker genes (Top three) in nine subclusters of CD4⁺ T cells. (C) Tissue prevalence estimated by Ro/e score in CD4⁺ T cells. Cells derived from tumor, normal tissue, normal lymph node (NLN), metastases in lymph nodes (MLN), brain metastases (MBrain), and peritoneum (PE). (D) UMAP projection of CD8⁺ T cells. (E) Dotplot exhibited the expression of marker genes in eleven subclusters of CD8⁺ T cells. (F) Tissue prevalence estimated by Ro/e score in CD8⁺ T cells. (G) UMAP visualization in the trajectory of defined CD4 T cell types marked by colors, the dark line denoted cellular evolutionary directions (left). Visualization of single-lineage cells marked with inferred pseudotime by Monocle3 (using UMAP). The black curve denotes the inferred lineage (right). (H) UMAP visualization in the trajectory of defined CD8 T cell types marked by colors, the dark line denoted cellular evolutionary directions (left). Visualization of single-lineage cells marked with inferred pseudotime by Monocle3 (using UMAP). The black curve denotes the inferred lineage (right). (I) Boxplots representing the MCP_counter score of each subclusters of CD4 T cells (upper) and CD8 T cells (lower) in tumor versus normal. (J) Dotplot denotes the expression of each gene (row) in tumor versus normal. (K) Dotplot indicating the expression of tumor-specific T cell genes in each cluster of T cells.

1. Introduction

Lung cancer comprises the majority, around 25% of total cancer deaths worldwide and the 5-years overall survival (OS) has achieved very little progress in recent years [1]. Despite advances in our understanding of new pathogenesis and treatment options, unfortunately, lung cancer is still one of the highly aggressive and rapidly fatal tumor subtypes [2]. New insights into the causal roles in lung cancer spurred the latest belief that tumorigenesis is entirely driven not only by cancer-cell-intrinsic genetic aberrations, but also is affected by the tumor microenvironment (TME) [3]. Lung cancer develops in complex TMEs consisting of endothelial cells, tumor-infiltrating immune cells (TIICs) and fibroblast cells [4].

Immune checkpoint inhibitors (ICIs) which enhance immunization and relieve immune suppression and anti-angiogenesis therapy which inhibits tumor growth through inhibiting angiogenesis both had a significant effect on lung cancer. In ICI, a number of patients do not benefit from the initial therapy (primary resistance) and some responders recur after a period of response (acquired resistance) [5]. The suppressive tumor immune microenvironment (TIME) is the main barrier to ICI efficacy [6]. T cells with exhausted or regulatory phenotype are correlated with failure in anti-cancer immunity [7]. Tumor cell-extrinsic mechanisms that induce ICI resistance involve components within the TME, including regulatory T cells (Tregs), myeloid-derived suppressor cells (MDSCs), and other inhibitory immune cells [8]. In anti-angiogenesis therapy, the majority of agents have shown some evidence of activity, but increased toxicities have been observed, including an increased risk of death for some agents, limiting their development [9]. It is important to further understand the influence of immune cells in the TME on the proliferation of vascular endothelial cells. Pre-existing tumor-infiltrating lymphocyte (TIL) is the necessary precondition for potent tumor regression. Based on the status of pre-existing TIL, tumor microenvironments are classified into three types, in excluded infiltration type, where abnormal angiogenesis and immunosuppressive reactive stroma prevent the infiltration of T cells [10]. Neo-vasculatures upregulate some inhibitory signals for an anti-tumor immune response such as PD-L1, indoleamine 2, 3-dioxygenase (IDO), interleukin-6 (IL-6), and interleukin-10 (IL-10) [11]. The interaction between immunity and angiogenesis renders tumor immune escape and treatment resistance. Many clinical studies have been conducted to investigate the synergistic effect of ICI plus anti-angiogenesis therapy in patients and have achieved good efficacy [12].

Recent advances in single-cell RNA sequencing (scRNA-seq) analysis have shed new light on the heterogeneity of T cell clusters in TME at single-cell resolution. Nevertheless, the tumor-specific T cells (TSTs) and how they affect the formation of the immune microenvironment, remains however poorly elucidated. How endothelial cells recruit TSTs to communicate is also unknown in lung cancer. The microbial abundance between cold and hot tumors has been even less discussed in lung cancer. Furthermore, combination strategies were identified in melanoma to overcome immunotherapy resistance [13], thus, further studies are required to identify predictive biomarkers for combination approaches in lung cancer.

In the present study, we aimed to uncover the characterization of the tumor-specific microenvironment by single-cell technology to guide the selection of therapeutic strategies. First, we identify TSTs with tumor-specific genes (TSTGs) for the tumor environment in lung cancer. Additionally, we applied Least absolute shrinkage and selection operator (LASSO) regression to filter out the most survival-relevant TSTGs to form the Tumor-Specific T Cell Score (TSTS). And we also defined these TSTS can receive the inhibitory signal from tip cells, a key phenotype in sprouting angiogenesis through specific crosstalk pairs. Finally, the TSTS was confirmed that can precisely predict the efficacy of combination strategies, immunotherapy, chemotherapy, and targeted therapy respectively in the 12 clinical treatment cohorts.

Thus, these findings suggest that selecting treatment strategies based on the characteristics of the tumor microenvironment is a more accurate and reliable approach and that TSTS can as an indicator to measure the choice of clinical treatment in lung cancer.

2. Results

2.1. ScRNA-seq identified tumor-specific T cell clusters and tumor-specific T cell features

To investigate the heterogeneity of tumor-infiltrated immune cells, we obtained scRNA-Seq data from patients diagnosed with lung cancer. After normalization and principal component analysis, we applied graph-based clustering (Fig. S1A, Table S1). These clusters could be assigned to eight known cell lineages through marker genes (Fig. S1B). The specific multicell group, characterized by high expression of MKI67, is a group of proliferating cells with active mitosis (Fig. S1C). We divided T cells into CD4⁺ and CD8⁺ T cells by CD4, CD8A and CD8B expression (Fig. S1D). Nine clusters of CD4⁺ T cells and eleven clusters of CD8⁺ T cells were defined (Fig. 1A–D). The expression of signature genes and known functional markers suggested the presence of conventional naïve (CD4_CCR7, CD4_IL7R,

CD8_CCR7), effector (CD4_GZMA, CD4_FGFBP2, CD8_GZMK and CD8_TYROBP), memory (CD4_FOS, CD8_TXNIP, CD8_ZNF683), exhausted clusters (CD4_CXCL13, CD8_CXCL13, CD8_TIGIT), CD4⁺ Tregs (CD4_FOXP3), proliferative clusters (CD4_STMN1, CD8_STMN1) (Fig. 1B–E Table S2). To determine tumor-specific T cells, we compared the observed and expected cell numbers of each cluster (Ro/e) (Fig. 1C–F) and the density of T cells between tumor and normal (Fig. S1E). Combining the above two conditions, the exhausted clusters (CD4_CXCL13 and CD8_CXCL13), ISG gene-related subcluster (CD4_ISG15, CD8_ISG15), regulatory clusters (CD4_FOXP3), and pre-exhausted (CD8_ZNF683) enriched in the tumor, which can be defined as tumor-specific T cells (TSTs) which highly expressed immune checkpoint inhibitors like PD1, CTLA4 and LAG3 (Fig. S1F). Through pseudotime analysis, we determined that CD4⁺ exhausted T cells were transformed from effector T cells and CD8 + exhausted cells were transformed from effector memory

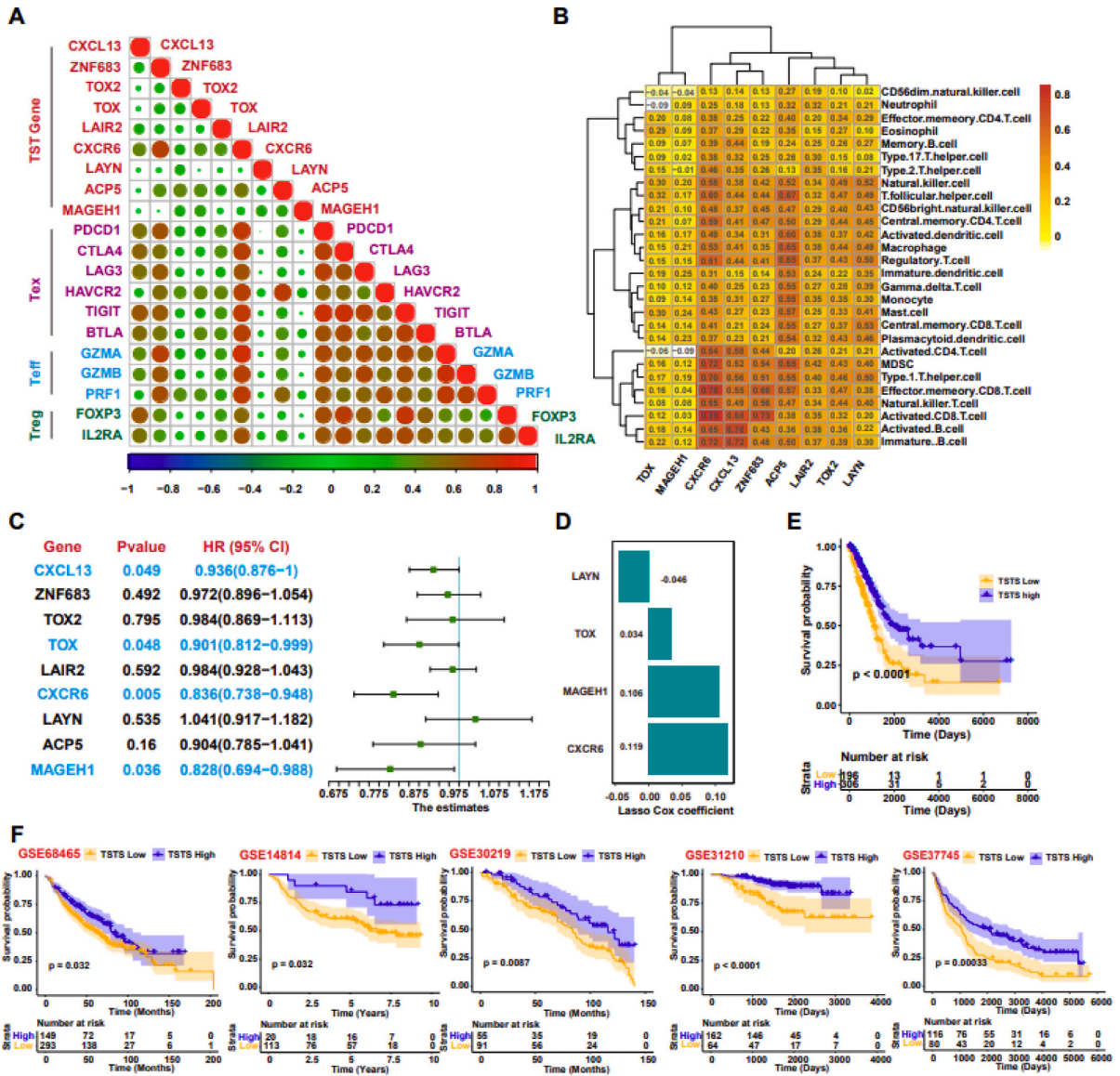
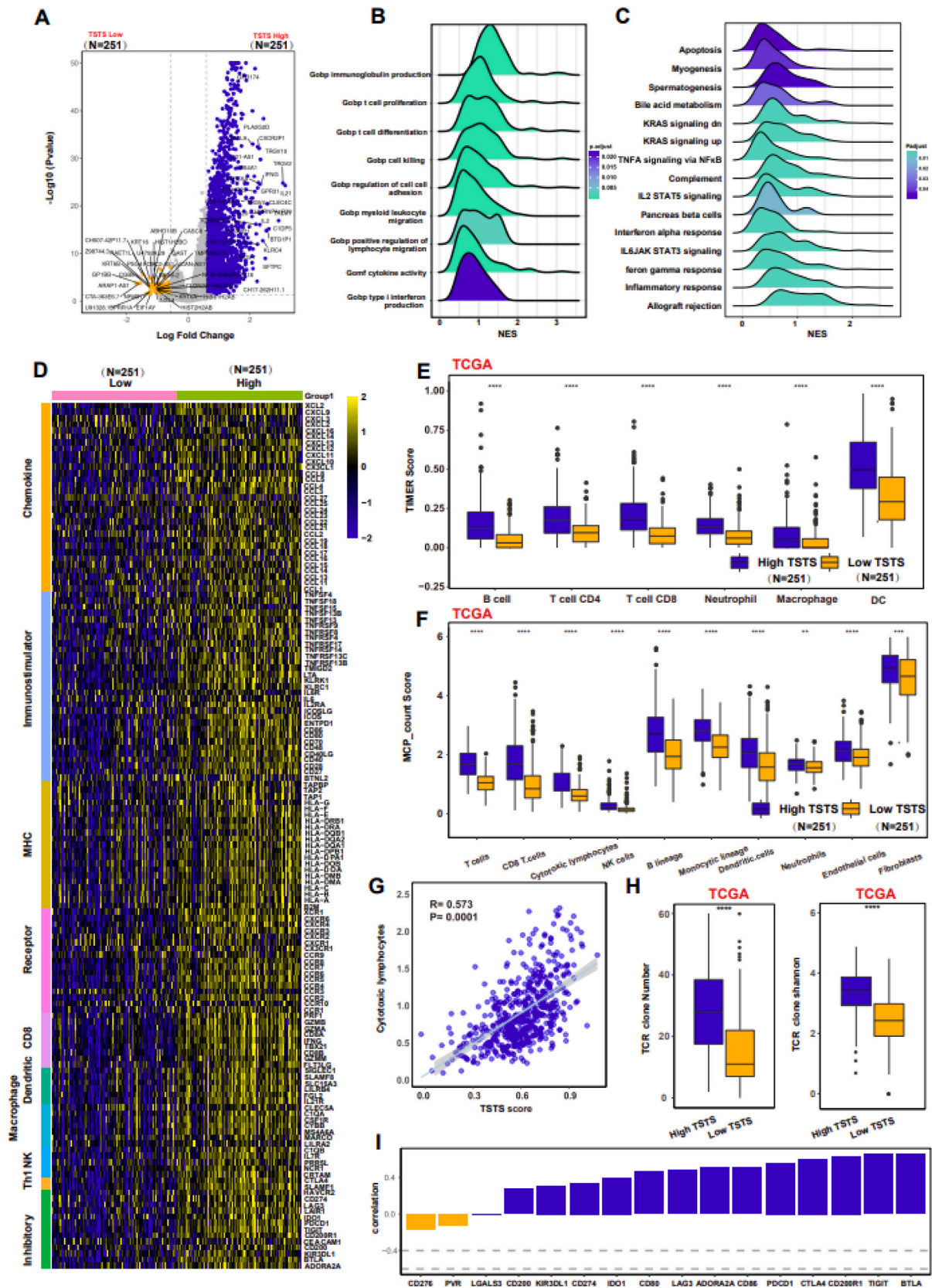


Fig. 2. Construction of tumor specific T cell score with survival outcomes (A) The correlation between the expression of TSTGs (CXCL13, ZNF683, TOX2, TOX, LAIR2, CXCR6, LAYN, ACP5 and MAGEH1) and the expression of immune checkpoint genes (PDCD1, CTLA4, LAG3, HAVCR2, TIGIT and BTLA), the effector T cells associated genes (GZMA, GZMB and PRF1) and the Tregs-related genes (FOXP3 and IL2RA). *P*-value < 0.001, Pearson correlation analysis for the datasets that meet the normal distribution, Spearman correlation analysis for those that are not normally distributed (detailed statistical information of these dataset is attached to Table S11). (B) The correlation between the expression of TSTGs and ssGSEA score which represents the infiltration of immune cells. (C) The *p*-value and HR of selected genes in univariable Cox regression analysis. (D) The bar plot shows the lasso cox coefficient of four optimal genes (LAYN, TOX, MAGEH1, and CXCR6) associated with prognosis. (E) Kaplan-Meier (K–M) curves demonstrated low TSTs group (orange) had noticeably poorer over survival than patients in the high TSTs group (blue). (F) Kaplan-Meier curve of validation cohorts (GSE68465, GSE14814, GSE30219, GSE31210, GSE37745) showing low TSTs group (orange) had noticeably poorer survival than patients in the high TSTs group. *P*-value was calculated by the log-rank test.



(caption on next page)

Fig. 3. Enrichment of immune infiltration in high TSTS group (A) Volcano plot demonstrating the DE genes between high TSTS group and low TSTS group. (B) Enriched Gene Ontology (GO) pathways of high TSTS group shows immunological enhancement in high TSTS group. Statistical analysis was performed using GSEA analysis. The adjust P value labeled to legend. (C) Enriched hallmark pathways of the high TSTS group. Statistical analysis was performed using GSEA analysis. The adjust P value labeled to legend. (D) Heatmap shows immune-related genes (chemokine, immunostimulator, MHC, chemokine receptors and inhibitory immune checkpoint) and canonical genes of five TIICs (CD8⁺ T cells, DC, NK cells, macrophages, and Th1 cells) were highly expressed in high TSTS group. (E) By deconvolution method TIMER, boxplot demonstrated the six immune cells' score in the high TSTS versus the low TSTS group. (F) By deconvolution method MCP_counter analysis, boxplot demonstrated the ten types of cells' score in the high TSTS versus the low TSTS group. (G) The correlation between the TST scores and tertiary lymphoid structure (TLS) score. TLS score was calculated from the average expression of CCL19, CCL21, CXCL13, CCR7, SELL, LAMP3, CXCR4, CD86 and BCL6. Statistical significance was determined by Spearman's rank correlation analysis (p value < 0.001). (H) TCR clone number (left) and TCR clone diversity (right) was significantly elevated in high TSTS group. (I) Correlation between TSTS score and expression of known immune checkpoint genes. ns P > 0.05; *P < 0.05; ***P < 0.001; ****P < 0.0001. Statistical significance was determined by Spearman's rank correlation analysis (p value < 0.001).

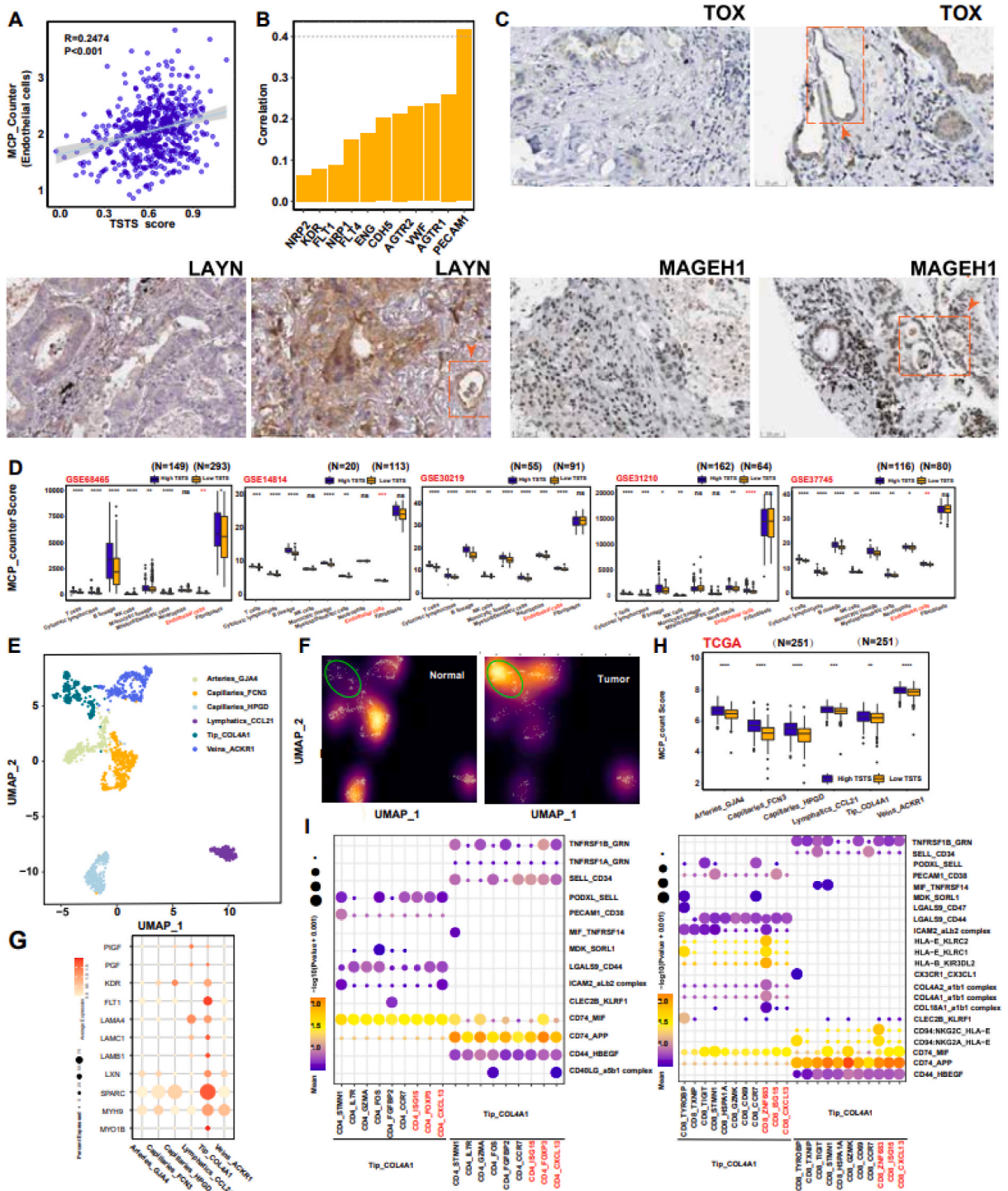
T cells (Fig. 1G and H). These T cells often lose their effector function and infiltrate in the microenvironment in tumor in the late stage of T cell transformation. We found the percentage of four representative TSTs was positively with pre-exhausted state cells which have effector function, which demonstrated that the more cells in the effector state, the more cells will transition to the exhausted state (Fig. S1G). Cells in the ISG gene related subcluster (CD4_ISG15, CD8_ISG15) had highly expressed ISG-related genes like ISG15, IFITM1 and IFITM3, but we found that these genes were expressed not only in T cells, but in myeloid cells can not represent TSTs (Fig. S1H). Using MCP_Counter deconvolution analysis with highly expressed genes in each cluster, we determined TSTs clusters-CD4_CXCL13, CD8_CXCL13, CD4_FOXP3 and CD8_ZNF683 had higher infiltration in tumor samples from TCGA compared with normal samples from GTEX (Fig. 1I; Table S3). Compared to normal tissues, nine representative genes -CXCL13, ZNF683, TOX2, TOX, LAIR2, CXCR6, LAYN, ACP5 and MAGEH1 were specially expressed in the tumor tissue in TSTs clusters (Fig. 1J, Fig. S1I) and these genes exclusively expressed in the TSTs clusters-CD4_CXCL13, CD4_FOXP3, CD8_CXCL13, CD8_ZNF683-were recognized as tumor-specific T cell genes (TSTGs) (Fig. 1K). These genes were expressed only on T cells and not on other cells in the TME (Fig. S1H). It is recognized that LAYN is closely related to tumorigenesis, and ZNF683 is chiefly implicated in the differentiation of T cells [14]. Both TOX2 and TOX are transcription factors, which correlate with T cell exhaustion and tumor immunosuppressive microenvironment [15].

2.2. Construction of tumor-specific T cell gene signatures

To avoid heterogeneity bias from single-cell data, we validated the correlation between TSTGs and immune-related genes using bulk RNA-seq in TCGA. The result demonstrated that the TSTGs (CXCL13, ZNF683, TOX2, TOX, LAIR2, CXCR6, LAYN, ACP5 and MAGEH1) and the exhausted T cells associated genes (PDCD1, CTLA4, LAG3, HAVCR2, TIGIT), the effector T cells associated genes (GZMA, GZMB), the Tregs-related genes (FOXP3, IL2RA) were positively correlated (Fig. 2A). Subsequently, single-sample Gene Set Enrichment Analysis (ssGSEA) was applied in TCGA to infer from bulk-RNA seq to reveal the relative level of immune cell infiltration. Positive correlation between gene expression of TSTGs and immune cell infiltration score was proved that TSTGs can reflect the state of immune infiltration in bulk RNA-seq (Fig. 2B–Tables S4–5). The immune microenvironment is often related to over survival. The expression of these genes was further assessed using univariate Cox regression analysis. The results indicated that 4 of 9 candidate genes were significantly correlated with OS of lung cancer patients (Fig. 2C, p < 0.05). In order to identify the optimal prognostic biomarker, the least absolute shrinkage and selection operator (LASSO) logistic regression was applied to screen the tumor-specific features and recruited to construct a multi-gene prognostic signature (Figs. S2A and B). The four genes including LAYN, TOX, MAGEH1 and CXCR6 are considered to be the optimal biomarkers with the best features that build tumor-specific T cell score (TSTS) (Fig. 2D). In the TCGA training set, patients were divided into two groups according to the TSTS score. The patients with lower TSTS score had a worse overall prognosis compared to the patients with higher TSTS score (Fig. 2E), which was validated in five independent validation cohorts (Fig. 2F). Further survival analysis demonstrated that the high TSTS score group showed a decreased disease-free survival than the low TSTS score group (Fig. S2C). To test the independence of our scoring system, a multivariate Cox regression analysis was then performed. After adjusting by common clinical parameters such as age, gender, stage and smoking status, the TSTS score still displayed robust assessment ability (Fig. S2D). In clinical staging, we also noted that the TSTS score represents early-clinical stages in TCGA dataset (Fig. S2E). Early-stage lung cancer patients had higher TSTS score, suggesting that there may be higher tumor-specific T cells infiltrating in the TME at this time. To further investigate the clinical potentiality of the TSTS score of lung cancer, stratified analysis based on these clinical characteristics was implemented. The results indicated that the TSTS score seemed to predict OS of lung cancer patients in subgroup of T1, T2, T3, T4, N0, N1, N2, N3, M0 and M1 (Fig. S2F). Patients in low groups had significantly poorer clinical outcomes than those in high groups. The prognostic value of the TSTS score for the survival outcome of patients was verified in pan-cancer types using TCGA datasets, TSTS may be used as a biomarker of pan-cancer to predict patients' survival (Fig. S3A).

2.3. Immune infiltration was enhanced in the high TSTS group

Furthermore, in the lung cancer samples from TCGA, firstly the high and low groups are divided at the median, then it is intriguing to investigate that the high TSTS group was featured by high immune infiltration, which enriched genes were immune-related genes such as CXCL9, IFNG, IL2, MS4A1, CLEC6A, and KLRC4 (Fig. 3A, Table S6). The elevated gene of the high TSTS group was enriched in immunological enhancement pathways, including immunoglobulin production, positive regulation of lymphocyte activation, T cell



(caption on next page)

Fig. 4. Angiogenesis and tumor-specific T cell infiltration have synergistic effects (A) Positive correlation between the TSTS and the endothelial cells' score (calculated by MCP-counter). Statistical significance was determined by Spearman's rank correlation analysis (p value < 0.001). (B) Positive correlation between TSTS and the expression of vascular related genes. P -value < 0.001, Pearson correlation analysis for the datasets that meet the normal distribution, Spearman correlation analysis for those that are not normally distributed (detailed statistical information of these dataset is attached to Table S11). (C) Immunostaining of representative genes (TOX, LAYN, and MAGEH1). Orange arrows and dashed boxes indicate blood vessels. (D) Utilization of MCP_counter de-convolution algorithm to compare the infiltration of endothelial cells between TSTS low and high groups in five independent cohorts. (E) UMAP analysis identified six clusters that named arteries, veins, capillaries and tips of the blood vasculature, as well as the lymphatic vessels. (F) Density plot of UMAP representation comparing normal and tumor samples. Solid line highlights the different infiltration level between normal and tumor. (G) Dotplot shows the expression of angiogenesis and endothelial cell growth genes in Tip_COL4A1 cluster. (H) By MCP_counter analysis, infiltration score of six clusters endothelial cells presents different state in the high and low TSTS group. (I) Dotplot showing the statistically significant ligand-receptor interactions between Tip_COL4A1 cells and subclusters of T cells, inferred by CellPhoneDB in which cluster of tumor specific T cells highlighted by red. P values are indicated by circle size. ns $P > 0.05$; * $P < 0.05$; *** $P < 0.001$; **** $P < 0.0001$.

proliferation, cell killing, and Type I interferon production (Fig. 3B). Similarly, the Hallmarks pathway in the high TSTS group also revealed highly significant enrichment in immunological activity including interferon-alpha response, interferon-and gamma response, inflammatory response (Fig. 3C–Table S7). Five immunomodulators (chemokine, immunostimulator, MHC, and chemokine receptors, inhibitory immune checkpoint) and the infiltration levels of five types of TIICs ($CD8^+$ T cells, DC, NK cells, macrophages, and Th1 cells) were enriched in high TSTS (Fig. 3D). MCP counter, XCELL, EPIC, quanTiseq and TIMER algorithms were used to calculate an immune score and estimate the abundance of various types of immune cells. We found higher proportion of $CD4^+$ T cells,

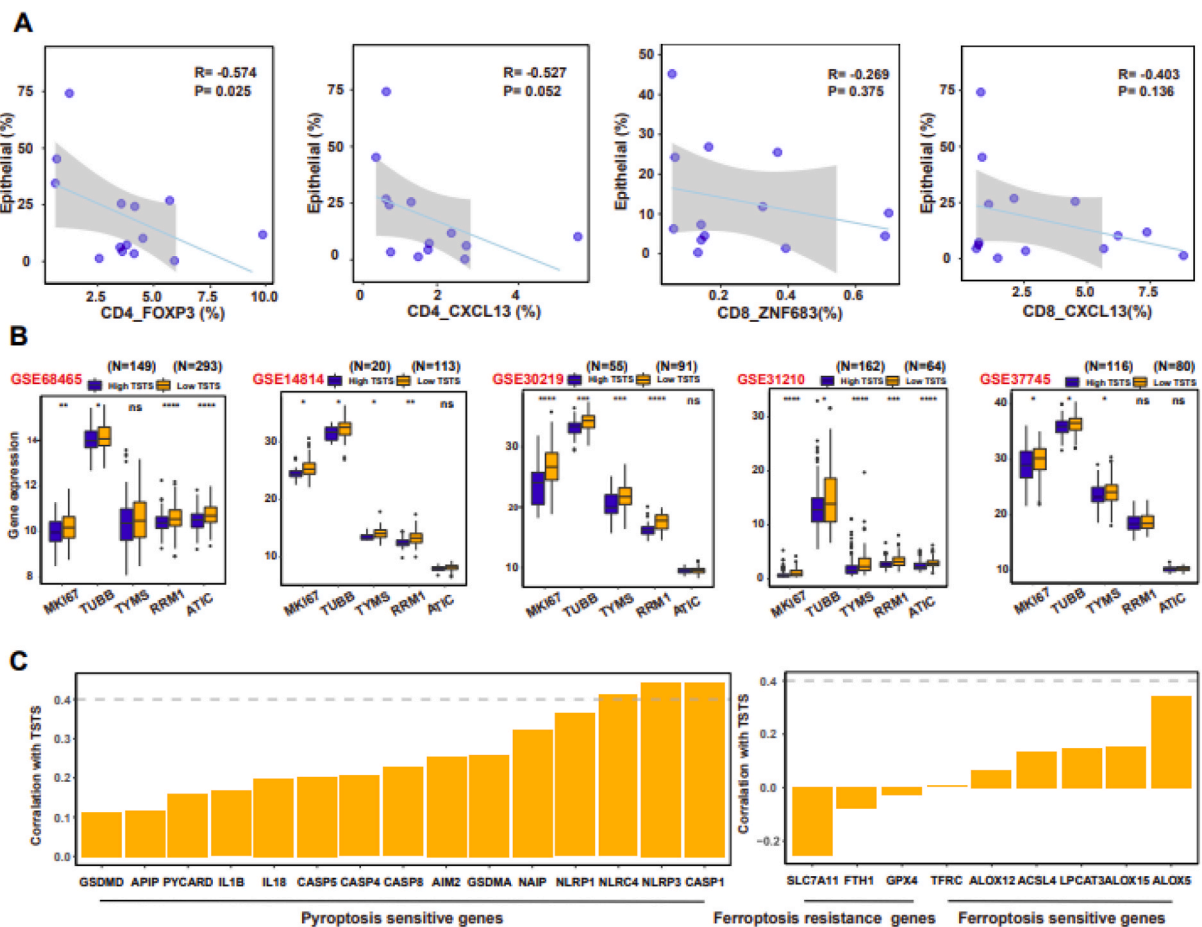
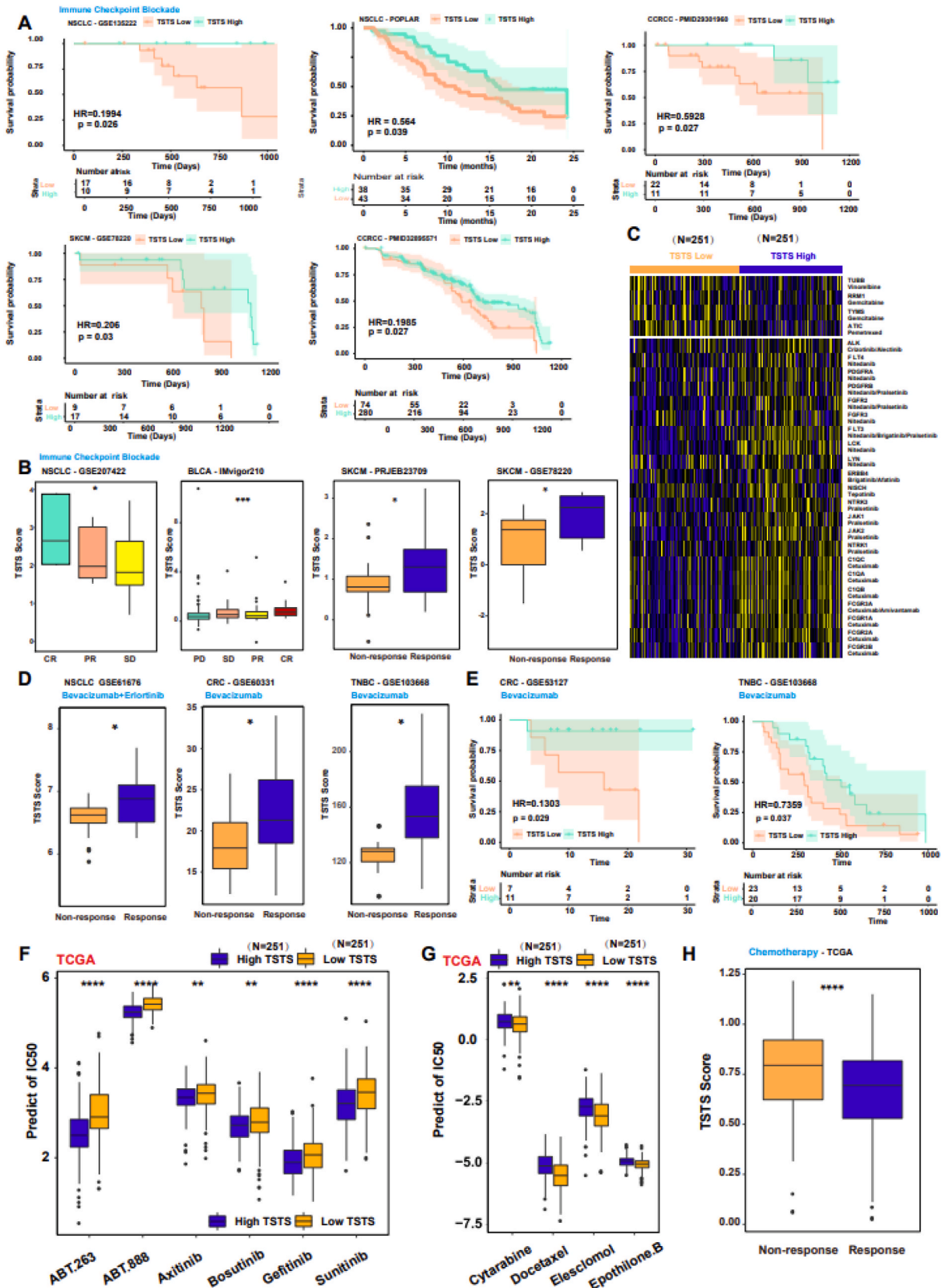


Fig. 5. The low TSTS group showed a state of rapid proliferation of malignant cells (A) The correlation analysis between the proportion of Tregs cells (CD4_FOXP3), exhausted T cell cells (CD4_CXCL13 and CD8_CXCL13), pre-exhausted cells (CD8_ZNF683) and the proportion of epithelial cells. Each dot represents a patient. In one patient, the ratio was calculated as the proportion of the number of cells in the appointed group to the total number of cells. Statistical significance was determined by Spearman's rank correlation analysis (p value < 0.001). (B) Boxplot shows the different expression level of proliferation-related genes in high TSTS and TSTS group in five independent cohorts. (C) In the TCGA dataset, the correlation between the TSTS and the expression of pyroptosis sensitive genes (left) and ferroptosis resistance or sensitive genes (right). Statistical significance was determined by Spearman's rank correlation analysis (p value < 0.001).



(caption on next page)

Fig. 6. TSTS identified as a predictor of immunotherapy, chemotherapy and targeted/anti-angiogenic therapy (A) Kaplan-Meier (K–M) curves demonstrated low TSTS group (red) had noticeably poorer survival than patients in high TSTS group (green) in immunotherapy treatment cohorts (B) The distribution of TSTS between responders (stable disease, partial response and complete response) and non-responders (progressive disease) in immunotherapy treatment cohorts. (C) Heatmap shows the different expression level of NSCLC-related drug-target genes between low TSTS and high TSTS ($P < 0.05$) (D) The distribution of TSTS between responders (stable disease, partial response and complete response) and non-responders (progressive disease) in anti-angiogenic treatment cohorts. (E) Kaplan-Meier (K–M) curves demonstrated low TSTS group (red) had noticeably poorer survival than patients in high TSTS group (green) in anti-angiogenic treatment cohorts. (F) Targeted-therapeutic sensitivity of six common drugs were estimated and compared in TCGA cohort. (G) Chemotherapeutic sensitivity of four common drugs were estimated and compared in TCGA cohort. (H) The distribution of TSTS between responders (stable disease, partial response and complete response) and non-responders (progressive disease) in TCGA chemotherapy treatment cohorts. ns $P > 0.05$; * $P < 0.05$; *** $P < 0.001$; **** $P < 0.0001$.

CD8⁺T cells, NK cells and B cells in lung cancer with higher TSTS (Fig. 3E–F, Tables S8–9). Consistent with the above studies, TSTS closely related to immunopotentiality validated by multiple deconvolution algorithms (Fig. S4A). Meanwhile, TSTS highly positively correlated with cytotoxic lymphocytes (Fig. 3G). To our surprise, there were a large number of endothelial cells in the high TSTS group (Fig. 3F, Fig. S4A). In addition, the higher number and diversity of T-cell receptors (TCRs) in the high TSTS group than that in the low TSTS group ($P < 0.01$) suggested higher tumor antigen stimulation and TCR expansion occurs in the high TSTS group (Fig. 3H). Combined cell compositions and pathway of the high TSTS group, more favorable immune environment consisting of immune cells and immune components exist in higher groups. Meanwhile, the most ICI-related genes used as therapy targets were the highly positively correlated with TSTS, a direct clue that TSTS may be suitable for ICI therapy (Fig. 3I).

2.4. Angiogenesis and tumor-specific T cell infiltration have synergistic effects

In consideration of the infiltration of endothelial cells in the high TSTS group, we evaluated how endothelial cells are involved in the formation of tumor microenvironment in the high TSTS group. We found TSTS score was positively correlated with MCP_counter score of endothelial cells (Fig. 4A). We also found TSTS score had positive correlation with migration (NRP1), and vessel formation (FLT1, KDR, NRP1, FLT4, NRP2 and ENG), adhesion (CDH5 and PECAM1) related gene expression level (Fig. 4B). By immunohistochemical analysis, we found high expression of TOX, LAYN, and MAGEH1 proteins that make up the TSTS fraction in the perivascular area. In contrast, immunohistochemical samples in which no obvious blood vessels could be found often showed negative protein staining (Fig. 4C). With the utilization of the MCP_counter de-convolution algorithm, T cells and endothelial cells had higher infiltration level in the high TSTS group in five independent validation bulk datasets (Fig. 4D). We used spatial transcriptome from a patient with lung cancer to demonstrate that the emergence of TSTS was often accompanied by vascular infiltration (Fig. S5A). In order to better understand the relationship between immune infiltration and angiogenesis, we divided all endothelial cells into six clusters (Fig. 4E; Table S2). By integrating the function-related gene signatures in previous study [16], we determined the arteries, capillaries, tips, and veins of the blood vasculature, as well as the lymphatic vessels (Fig. S5B). Remarkably, cluster Tip_COL4A1, which enriched in tumor tissue the key phenotype in sprouting angiogenesis, expressed a signature of genes involved in angiogenesis and endothelial cell growth (FLT1, KDR, PGF, PIGF and SPARC), cell adhesion, migration (LAMC1, LAMA4, LAMB1 and SPARC), facilitating their navigating role in the process of vessel sprouting (Fig. 4F and G). With the MCP_counter de-convolution algorithm, we found that six endothelial cell clusters had high infiltration level in the high TSTS group in TCGA (Fig. 4H). These observations suggested that angiogenesis serves as an important characteristic in the high TSTS group, and the tip EC phenotype is likely to be a promising target of anti-angiogenic therapy (AAT), especially anti-vascular endothelial growth factor receptor (VEGFR) antibodies in lung cancer. To dissect the complex activities of tip cells in TME, we subsequently used CellPhoneDB to investigate the molecular communication between T cells and tip cells. In this analysis, we observed widespread interactions between Tip_COL4A1 and TSTs cells, which expressed the CD94: NKG2A receptors on surface and received the inhibitory signal HLA-E from Tip_COL4A1 (Fig. 4I–Table S10). As a result, CD8⁺ T cell activation is suppressed, it was proved a mechanism of immune escape in various tumor types such as glioma, melanoma, or colorectal cancer [17], it may hint the presence of tip cells may contributed to immune escape in lung cancer. Taken together, these observations reflected the therapeutic potential for AAT alone or in combination with ICB in lung cancer.

2.5. Elevated malignant cell proliferation status in the low TSTS group

Of note, In single cell analysis, the proportion of epithelial cells was negatively correlated with tumor-specific T cells (TSTs) which also found in immune cell deconvolution of TCGA dataset (Fig. S4A), suggesting that more epithelial cells were clustered in the low TSTS group (Fig. 5A). Moreover, genes involved in proliferation (MK167 and TUBB) and synthesis of important nucleotides necessary for cell growth (TYMS, RRM1 and ATIC) had high expression in low TSTS groups in five validation cohorts (Fig. 5B). And we found that the TSTS score was positively correlated with the expression of pyroptosis and ferroptosis sensitive genes and negatively correlated with the expression of ferroptosis resistance genes which suggests that there is an anti-apoptotic mechanism in the low TSTS group (Fig. 5C).

2.6. Associated oncogenic event reveals different TSTS states

Subsequently, we probed into the deep-seated reasons at the genome level underlying the distinct features between these two groups. The comparison of the proportions of tumor mutation burden (TMB) indicated no discernible difference (Fig. S6A, $p > 0.05$).

Nevertheless, the mutant landscape of the top 10 most frequently mutated genes was not significantly different between the two groups (Fig. S6B). Strikingly, KEAP1, DGKB, PSG8, and ITPR2, were identified as statistically different in odds ratios for low versus high groups (Fig. S6C). The Low TSTS group displayed enriched gene mutation of KEAP1, which accelerate proliferation and oncogenic transformation [18]. We found that there were the same or different mutational signatures in the high and low TSTS groups. Tobacco signature, APOBEC signature and defective DNA mismatch repair signature can explain most of the mutation signatures. The aging-associated signature SBS5 only appeared in the high TSTS group, and SBS40 only appeared in the low TSTS group. This may be an important reason for two kinds of microenvironment infiltration in lung cancer (Figs. S6D–E). What's more, we do further analysis to investigate whether mutation and TSTS can provide a more effective prediction of tumor progression. The results indicate that among the common mutations (TP53, KRAS, EGFR and ALK) in lung adenocarcinoma, the survival outcomes predicted by environmental factors represented by TSTS and mutations are fairly consistent, with no statistically significant differences (Fig. S6F). We found that when patients with STK11 mutation and lower TSTS score, the patient had a worse prognosis, indicating that STK11 mutations could make this indicator more clinically meaningful.

2.7. TSTS was identified as predictor of immunotherapy and targeted therapy

Taken together, high TSTS represents high immune environment with more exhausted T cells infiltration combining abundant tertiary lymphoid structure and angiogenesis while low TSTS exhibited state of high malignant cell proliferation, (Fig. S5B). Further in the clinical treatment cohort, in the NSCLC immunotherapy cohort, the high TSTS group was significantly associated with better survival outcome, suggesting that TSTS can predict immunotherapy efficacy, and this ability was also validated in the immunotherapy cohort of the CCRCC, BLCA and SKCM (Fig. 6A). The prognosis of the high TSTS group was better due to higher ICB response rates (Fig. 6B). Furthermore, we examined a single-cell dataset (GSE207422) of individuals undergoing *anti*-PD-1 therapy and assessed the levels of TSTS gene expression across groups with different treatment outcomes. The results revealed that TSTS genes were notably enriched in the CD4_CXCL13, CD4_FOXP3, and CD8_CXCL13 cell subclusters (Figs. S7A–B). Importantly, the degree of enrichment within these three cell subgroups was significantly higher in the partial response (PR) group compared to the stable disease (SD) group. This provides compelling evidence that TSTS is indeed a relevant indicator associated with the efficacy of immunotherapy (Fig. S7C).

Notably, comparing TSTS score to other well-recognized immunotherapy predictive signature [19,20] in the NSCLC immunotherapy cohort, Cytotoxic T cells (CTL), IFNG, T cell inflamed GEP score were statistically significant in predicting the efficacy of immunotherapy, furthermore, the Hazard Rates (HR) of TSTS was the smallest of all indicators (TSTS HR = 0.20, CTL HR = 0.45, IFNG HR = 0.36, T cell inflamed GEP = 0.37, CD8 HR = 0.40), implying that TSTS was the strongest protective factor for immunotherapy (Fig. S8A).

Patients with DGKB, PSG8, and ITPR2 mutations correlated with better survival than those with no mutation (Fig. S8B). Nevertheless, the patients carrying KEAP1 mutation had a poor survival outcome (Fig. S8C). Thus, the gene mutations mentioned above may serve as potential biomarkers and predictors of therapeutic efficacy in lung cancer patients.

Additionally, we searched the Drugbank database for therapeutic drugs and drug targets that have currently been approved in non-small-cell lung cancer (NSCLC). Genes that are the targets of chemotherapy drugs (vinorelbine, Gemcitabine and Pemetrexed) were highly expressed in the low TSTS group. Genes that are the targets of targeted therapy drugs (ALK inhibitor, EGFR inhibitor and antiangiogenic inhibitor) were highly expressed in the high TSTS group (Fig. 6C). Strikingly, the result of the NSCLC cohort provided unique insights into the better synergistic effect of Erlotinib (EGFR-TKI) combined with Bevacizumab (an antiangiogenic drug) in high TSTS group ($p < 0.05$). By analysis Two antiangiogenic therapy (Bevacizumab) cohorts, we also found patients with better responses had higher TSTS (Fig. 6D). Obviously, in antiangiogenic therapy cohorts, patients with high TSTS may have a better prognosis outcome than those with low TSTS (Fig. 6E). The estimated IC50 values of targeted therapy drugs were significantly elevated in TSTS-low samples of TCGA cohort (Fig. 6F); however, the estimated IC50 values of chemotherapy drugs were significantly elevated in TSTS-high samples (Fig. 6G). And in the chemotherapy cohort from the TCGA database, the low TSTS group had a better response efficacy (Fig. 6H).

Taken together, patients of advanced lung cancer with low TSTS were preferable for chemotherapy, whereas patients of early lung cancer with high TSTS fit targeted therapy (antiangiogenic inhibitor and small molecule inhibitors) and/or in combination with immunotherapy.

3. Discussion

TILs especially T lymphocytes, which are the main anti-tumor cells. There is ample evidence to demonstrate that the presence of TILs is associated with enhanced host response to the tumor and therefore is associated with improved patient outcomes [21].

Previous studies revealed that TILs consist of not only those T cells specific for TSAs, but also the ones that recognize epitopes unrelated to the tumor, such as virus-specific antigens, known as bystander T cells [22]. In contrast to bystander T cells, although exhibiting a dysfunctional phenotype, tumor-specific T cells possess significant capabilities of tumor-specific killing [23]. Therefore, we tried to find and use tumor-specific T cells related index to measure the level of immune infiltration and to precisely predict the response to ICB. Meanwhile, the strategy of combining ICB with ramucirumab demonstrated favorable effects in clinical trials [24–26]. Indicators urgently needed to guide combination therapy offer hope for the treatment of advanced lung cancer.

Through the single-cell sequencing technology, we first identified unique TSTs. Intriguingly, our findings with the gene expression profiles in TSTs further identified the tumor-specific T cell features, consisting of CXCL13, ZNF683, TOX2, TOX, LAIR2, CXCR6, LAYN, ACP5 and MAGEH1. LASSO Cox regression model was successively used to filter out the most robust markers and tumor-specific T cell-related prognostic score (TSTS) was established. Therefore, a TSTS model consisting of LAYN, TOX, MAGEH1 and CXCR6, classifies NSCLC patients into high TSTS group with better prognosis and low TSTS group with poor prognosis. Strikingly, TOX was reported as a critical regulator of tumor-specific T cell differentiation, and highly expressed in dysfunctional TST cells from tumors [27]. Remarkably, LAYN has been corroborated to suppress CD8⁺ T cells, and high expression of LAYN, MAGEH1 in Treg cells correlated with poor prognosis of NSCLC and colorectal cancer (CRC) [28]. Also, we found that chemokine receptor CXCR6 was enriched in CD4⁺ Treg and CD8⁺ exhausted T cell clusters. Knockdown of CXCR6 inhibited cell proliferation migration and invasion of hepatocellular carcinoma (HCC) cells, followed by the deregulated expression of vascular endothelial growth factor (VEGF) [29]. CXCR6 is vital for effector-like activate cytotoxic lymphocytes (CTLs) in the TME to maximize their antitumor efficacy before progressing to irreversible dysfunction [30].

Higher tumor antigen stimulation and TCR expansion occurs in the high TSTS group than that in the low TSTS group suggested higher ability to recognize tumor-specific antigen that ultimately kill tumor cells. Meanwhile the exhaustion of effector cells indicates that the more effector T cells become toxic before moved towards terminal differentiation. By using bulk RNA sequencing, we identified that immune-related pathways and cellular adhesion pathways were significantly elevated in the high TSTS group, immune-related genes, such as PD1, TIGIT and CTLA4 were highly expressed in the high TSTS group. The Low TSTS group exhibits low immune infiltration and high proliferative properties. Therefore, converse to the immunologically hot (T cells inflamed) TME in high TSTS group, cold (non-T cells inflamed) TME have been related with low TSTS, indicating that TSTS defines hot or cold TME. Ultimately, this study validated the ability of TSTS as a predictor of immunotherapy efficacy in pan-cancer in the SKCM, BTCC, CCRCC and NSCLC immunotherapy datasets and found strong predictive power in NSCLC, SKCM, CCRCC, and BTCC. We believe that this can be explained by that T cell subgroups are relatively stable. and do not tend to be heterogeneous across cancer types, resulting in the presence of tumor-specific T cells that are likely to be shared across cancer types. This phenomenon is also consistent with that reported by Zheng, L. et al. [31].

Strikingly, a strong positive correlation between TSTS and angiogenesis was identified in our present work. Sustained angiogenesis is considered to be an important hallmark of NSCLC [32]. Angiogenesis contributes to NSCLC development by providing oxygen and nutrients and facilitating tumor growth and metastasis [33]. We found that the tumor angiogenesis was predominantly featured by tip cell infiltration—the main cells at the tips of vascular sprouts integrating multiple activities during angiogenesis [34,35]. The sprouting angiogenesis promotes the creation of emerging blood vessels from pre-existing vessels, along with the tip endothelial cells, which may significantly affect immunity. Our research confirmed angiogenesis high enriched in high TSTS group. Subsequently, we further identified the comprehensive intercellular communication between tip cells and tumor-specific T cells. Notably, HLA-E/ β 2m-CD94/NKG2A, a new druggable inhibitory immune checkpoint, was found to be a highly communicative ligand pair between tip cells and CD8⁺ exhausted T cells, which contributes to the functional exhaustion of cytotoxic NK cells and CD8 T cells [36, 37]. In addition to inhibiting the induction of local hypoxia, promoting the aggregation of MDSCs, Treg and leading to a suppressive tumor immune microenvironment, inhibition of neovascularization can reduce the expression of immunosuppressive check pairs and provide new evidence for synergistic effects of immunotherapy and anti-angiogenesis therapy. High TSTS group, enriched in tip cells, were proved better response in Bevacizumab (AAT)-therapy cohort. Thus, it is plausible to make recommendations on the combination of anti-angiogenic and immunotherapy for patients with high TSTS.

Overall, low TSTS patients had cold TME, characterized by decreased immune infiltration, decreased cells apoptosis and increased cells proliferation. These patients represent a phenotype of “immune desert”, which has been associated with the poor outcome to ICB [38,39]. Surprisingly, not only were epithelial cells enriched in the low TSTS group but also the TSTS score is negatively correlated with the proliferative genes and Ferroptosis resistance genes and is positively correlated with pyroptosis sensitive genes ferroptosis sensitive genes, implying that proliferation and growth of cells in the low TSTS group would be more sensitive, meaning that patients in the low TSTS group tended to have a better response to chemotherapy. This is one of the unexpected findings of this study.

Meanwhile, also, we found that TMB as an immunotherapy-guided biomarker did not seem to differ between high and low TSTS groups. This may explain why previous research has reported that TMB has a limited predictive effect on non-small cell lung cancer [40]. It should be noted that TSTs, as an indicator of immune inflammation, is higher in patients with early-stage lung cancer, while patients with advanced lung cancer tend to be low TSTS with “immune desert”. This is in line with one previous clinical trial that earlier use of immunotherapy will bring more significant benefits to NSCLC patients [41].

The treatment of advanced NSCLC is mostly based on a combination of chemotherapy or targeted therapy. In recent years, immunotherapy targeting programmed cell death protein 1 (PD-1) and its ligand (PD-L1) or cytotoxic T-lymphocyte-associated antigen 4 (CTLA-4) has made significant progress in the treatment of NSCLC. *Anti*-CTLA-4 inhibitor has achieved breakthroughs in the treatment of NSCLC, demonstrating significant clinical efficacy, prolonging patient survival and dramatically changing the treatment landscape of NSCLC. Thus, such a biomarker, which can evaluate the efficacy of chemotherapy, targeted therapy and immunotherapy, brings a new breakthrough in the stratification of patients for comprehensive drug use.

The limitation of our approach should also be taken into consideration. Although RNA sequencing, whole exon sequencing, and single-cell RNA sequencing were used and validated in as many independent cohorts as possible, all conclusions were based on bioinformatics analysis. Our results warrant further validation in a large number of real-world cohorts.

4. Methods and materials

4.1. Data acquisition

Single-cell RNA-sequencing (ScRNA-seq) data (accession number: GSE131907, GSE207422) of lung adenocarcinoma samples from the initial publication was downloaded and reanalyzed for this manuscript. Spatial transcriptomic data for one patient with NSCLC were downloaded from 10X (<https://www.10xgenomics.com/resources/datasets/human-lung-cancer-11-mm-capture-area-ffpe-2-standard>). The Cancer Genome Atlas (TCGA) data: For comparison within a single TCGA-LUAD cohort, lung adenocarcinoma (LUAD) RNA Sequencing data (TPM), somatic mutation data and survival information were downloaded from the UCSC Xena data portal (<https://xenabrowser.net/datapages/>). Raw RNA-seq was annotated by Gencode V23 and quantified by RSEM (v1.3.3), tpm values are extracted, $\log_2(\text{tpm}+0.001)$ transformed, and combined all sample. Gene Expression Omnibus (GEO): RNA sequencing data (number: GSE68465, GSE14814, GSE30219, GSE31210 and GSE37745) of lung cancer with detailed survival data were downloaded as validation queues. RNA sequencing of immunotherapy cohort of non-small-cell lung cancer (GSE135222, GSE207422) and melanoma (GSE78220 and PRJEB23709) were downloaded from GEO. RNA sequencing of the immunotherapy cohort of clear cell renal cell carcinoma (PMID32895571 and PMID29301960) was downloaded from supplementary materials. POPLAR immunotherapy cohort of non-small-cell lung cancer was applied from European Genome-Phenome Archive (<https://ega-archive.org/studies/EGAS00001004343>). RNA sequencing of the immunotherapy cohort of bladder cancer was downloaded from R package Mvigor210CoreBiologies [42]. MSK-IMPACT assay of immunotherapy cohort was downloaded from cbiportal (<http://www.cbiportal.org/>). RNA sequencing of patients with antiangiogenic treatment in non-small-cell lung cancer (GSE61676), –triple-negative breast cancer (GSE103668), colorectal cancer (GSE60331 and GSE53127) was downloaded from GEO. All public datasets are downloads of the processed datasets, so background correction, normalization and expression value calculation follow the original authors' processing.

4.2. Tumor mutational burden (TMB) estimates

TMB was defined as the number of somatic non-synonymous single nucleotide variants. For the TCGA-LUAD dataset, somatic mutation data were retrieved from the UCSC Xena multi-omics database platform (https://xenabrowser.net/datapages/?dataset=TCGA.LUAD.sampleMap%2Fmutation_broad&host=https%3A%2F%2Ftcga.xenahubs.net&removeHub=https%3A%2F%2Fxcena.treehouse.gi.ucsc.edu%3A443) and preprocessed at the Broad Institute Genome Sequencing Center [43]. WES data were generated on the Illumina platform. Mutation calls were calculated using the MuTect method [44], and only calls with variant allele frequency (VAF) > 4.0% were included. The R package “maftools” [45] was then used to calculate the total number of somatic non-synonymous point mutations within each sample. The high and low TMB groups are cut off by the median. The copy number variation data were calculated by GISTIC2 [46].

4.3. Immunohistochemical staining

Six samples of IHC was available from the Human Protein Atlas (HPA, <https://www.proteinatlas.org/>). This database contains the immunohistochemical (IHC) staining profiles of numerous proteins in a variety of cancerous and non-cancerous tissues based on more than 8800 antibodies. All samples detailed information was attached in [Table S12](#).

4.4. Single-cell transcriptome sequencing and data preprocessing

Seurat (v3.1.3) R toolkit was used to analyze the single-cell transcriptome sequencing data [47]. Firstly, cells with low quality were filtered out [48]. Briefly, the dead or dying cells with more than 10% mitochondrial RNA content were removed, and the cells with a too low number (less than 200) were also removed. Cell doublets were predicted using DoubletFinder with default parameters. Then, the filtered gene expression matrix for each sample was normalized using the “NormalizeData” function in Seurat with default parameters which employs a global-scaling normalization method “LogNormalize” then multiplies by a scale factor (10,000 by default) and log-transforms. Only highly variable genes remained using the “FindVariableFeatures” function in Seurat. Next, “Runharmony” functions in harmony were used to integrate the gene expression matrices of all samples, where batch effects between different samples have been adjusted [49]. Next, the “RunPCA” function was used to perform the principal component analysis (PCA) and the “FindNeighbors” function was used to construct a K-nearest-neighbor graph. Next, the most representative principal components (PCs) selected based on PCA were used for clustering analysis with the “FindCluster” function to determine different cell types. Lastly, UMAP was used to visualize the different cell types.

We annotated the cell types using the following rules: Based on the most ten differentially expressed genes that were derived using the “wilcoxau” function in presto, genes such as CD3D, CD3G were used as T cell markers, SDC1, TNFRSF17 were used as plasma cell markers, CD68, CD14 were used as myeloid cell markers, MS4A2, CPA3 were used as mast cell markers, ACTA2, COL6A1 were used as fibroblast/smc cell markers, KRT18, EPCAM were used as epithelial cell markers, VWF, PECAM1 were used as endothelial cell markers, CD19, MS4A1 were used as B cell markers. CD4 and CD8 gene expression were used to differentiate CD4⁺ and CD8⁺ T cells. Sub-cluster of CD4⁺ T cells, CD8⁺T cells and endothelial were named by the first marker gene.

4.5. Pathway analysis of single-cell

The pathway analysis was performed using the Correlation Adjusted MEan RANk gene set test (CAMERA) that has been implemented in the SingleSeqGset (version 0.1.2) R package. Lists of metabolic genes and pathways were obtained from the KEGG database (<http://www.kegg.jp>) were used for the CAMERA analysis [48].

4.6. Tissue distribution of clusters

We used to observe and expected cells in each cluster $R_{o/e}$ = observed/expected, and expected cell numbers for each combination of cell clusters and tissues are obtained from the chi-square test. We defined $R_{o/e} > 1$ as specific tissue cluster [50].

4.7. Trajectory analysis

Monocle 3 was used to infer the cell differentiation trajectories [51]. These algorithms place the cells along with a trajectory corresponding to a biological process by taking advantage of an individual cell's asynchronous progression under an unsupervised framework.

4.8. Cell-cell communication analysis

To investigate the potential cell-cell communications between T cells and endothelial cells from the tumor, we performed ligand-receptor analyses using CellPhoneDB [52] (version 2.0.6). CellPhoneDB applies an algorithm that considers only receptors and ligands with broad expression among the tested cell types, followed by calculating the likelihood of cell-type specificity of a given receptor-ligand complex with a sufficient number of permutations.

4.9. Selection of candidate genes and establishment of tumor-specific T cell signatures

Tumor-specific T cell clusters can be defined by R_{oe} . Genes expressed in these clusters and did not express in any other clusters were involved in the establishment of the TSTS score. We chose these genes for more functional investigation and developed a possible risk score using the lasso cox regression algorithm [53]. Finally, minimum criteria defined six genes and their constants, choosing the perfect penalty parameter λ related to the min 20fold-cross validation inside the training set. The EIC score was calculated:

$$\sum_{i=1}^n (Coef_i * x_i)$$

Coef_i is the coefficient and x_i is the value of chosen genes.

4.10. Functional and pathway enrichment analysis of bulk RNA-Seq

Limma R package was used to identify the signaling pathways that were differentially 473 activated between the low TSTS score and high TSTS score groups [54]. Signal pathways specific to the tumor microenvironment (TME) phenotype between the high TSTS and low TSTS groups were explored by GSEA, Gene Set Enrichment Analysis (GSEA) was used with adjusted $p < 0.05$ using the clusterfiler R package [55].

4.11. Immunological characteristics of the TME in lung cancer

Evaluation of Immune Cell Infiltration The proportions of immune cell types (i.e., TICs) in each sample (with immune infiltration scores) were calculated using the Timer [56] and MCP_counter algorithm [57], EPIC, Xcell, quanTIseq. Dividing between the high and low groups for comparing the proportions of immune cell types in the text are thresholded at the median.

4.12. Deciphering mutational signature operative in the genome

We used R packages Maftools to extract mutational signatures from the aggregated WES genomic data of NSCLC cohorts [58]. The ExtractSignatures function from Maftools factorized the mutation portrait matrix into two nonnegative matrices, "signatures" and "contributions," where "signatures" represented mutational processes and "contributions" represented the corresponding mutational activities. Mutational signatures were annotated by calculating cosine similarity against 30 validated mutational signatures in COSMIC, Version 2 [59].

4.13. Spatial transcriptome analysis

To infer the spatial organization of certain cell subtypes, we used Seurat R package to integrate spatial and single-cell data. Raw UMI counts were normalized by “SCTransform” function. Dimensionality reduction and clustering were performed as before. Cell subtypes distributions were visualized in spatial context over H&E images.

4.14. Statistical analysis

Dividing between the high and low groups for all survival curves in the text are thresholded at the median and the sample sizes between high and low groups are indicated in the pictures. The Kaplan-Meier method was used to estimate OS or FPS, the log-rank test was used to compare the Kaplan-Meier curves. A two-sided P value of less than 0.05 was considered significant. All sample sizes were large enough to ensure proper statistical analysis. Pearson correlation analysis for the datasets that meet the normal distribution, Spearman correlation analysis for those that are not normally distributed (detailed statistical information of these dataset is attached to [Table S11](#)). Statistical analyses were performed using GraphPad Prism (GraphPad Software, Inc.). P values < 0.05 were considered statistically significant. All *t*-test analyses were one-tailed *t*-tests (paired or unpaired depending on the experiments).

5. Conclusions

In summary, we applied single-cell RNA sequencing to lung cancer patients and identified tumor-specific T cells and their features. The model constructed by LAYN, TOX, MAGEH1 and CXCR6 exclusively expressed in the tumor-specific T cells, which score represented how extend of the immune-inflammatory state and angiogenesis state, dividing lung cancer patients into two types. Patients with high TSTS are sensitive to immunotherapy and targeted therapy, while patients with low TSTS are sensitive to chemotherapy. Collectively, this article provides a basic explanation of the environmental context in which TSTS is used as a treatment decision-making strategy. And the limitation of our approach should also be taken into consideration. Although RNA sequencing, whole exon sequencing, and single-cell RNA sequencing were used and validated in as many independent cohorts as possible, all conclusions were based on bioinformatics analysis. Our results warrant further validation in a large number of real-world cohorts. The underlying mechanism should be further investigated.

Source(s) of support

The study was supported by National Natural Science Foundation of China (81800050), and Natural science fund of Yangzhou City (YZ2017119).

Ethics approval

Not applicable.

Consent to publish

Not applicable.

Consent to participate

Not applicable.

Data availability statement

The processed Single-cell RNA-sequencing (ScRNA-seq) data of lung adenocarcinoma samples are available in GEO (accession number: GSE131907, GSE207422). The processed RNA sequencing data of lung cancer (number: GSE68465, GSE14814, GSE30219, GSE31210 and GSE37745), non-small-cell lung cancer (GSE135222, GSE207422) and melanoma (GSE78220 and PRJEB23709), and cohorts with antiangiogenic treatment in non-small-cell lung cancer (GSE61676), -triple-negative breast cancer (GSE103668), colorectal cancer (GSE60331 and GSE53127) are available in GEO. The processed RNA sequencing of the immunotherapy cohort of bladder cancer are available in R package Mvigor210CoreBiologies [42]. POPLAR immunotherapy cohort of non -small-cell lung cancer are available in European Genome-Phenome Archive.

(<https://ega-archive.org/studies/EGAS00001004343>). MSK-IMPACT assay of immunotherapy cohort are available in cbiportal (<http://www.cbiportal.org/>). The TCGA-LUAD RNA Sequencing data (TPM), somatic mutation data and survival information are available in the UCSC Xena data portal (<https://xenabrowser.net/datapages/>). The processed RNA sequencing of the immunotherapy cohort of clear cell renal cell carcinoma (PMID32895571 and PMID29301960) are available in the original authors' supplementary materials. Spatial transcriptomic data for one patient with NSCLC are available in 10X (<https://www.10xgenomics.com/resources/datasets/human-lung-cancer-11-mm-capture-area-fpfe-2-standard>).

The source code for TSTS is available at <https://github.com/liuxuefei99/LungTSTS>.

CRediT authorship contribution statement

Ziwei Luo: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xuefei Liu:** Writing – review & editing, Visualization, Software, Project administration, Methodology, Data curation, Conceptualization. **Ying Chen:** Visualization, Software, Methodology, Data curation. **Lize Shen:** Visualization, Software, Data curation. **Hui Qin:** Visualization, Software, Methodology, Data curation. **Qiongfang Zha:** Validation, Software, Data curation. **Feng Hu:** Visualization, Methodology, Data curation. **Yali Wang:** Writing – review & editing, Writing – original draft, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yali Wang reports financial support was provided by National Natural Science Foundation of China. Yali Wang reports financial support was provided by Natural science fund of Yangzhou City. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the professor of Zhixiang Zuo at Sun Yat-sen University Cancer Center, Guangzhou. And we also acknowledge Yueyu Huang's critical guidance in the statistical aspects during the second revision process.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e28374>.

References

- [1] R. Alföldi, et al., Single cell mass cytometry of non-small cell lung cancer cells reveals complexity of in vivo and three-dimensional models over the petri-dish, *Cells* 8 (2019), <https://doi.org/10.3390/cells8091093>.
- [2] Y. Li, et al., Identification of a novel prognosis-associated ceRNA network in lung adenocarcinoma via bioinformatics analysis, *Biomed. Eng. Online* 20 (2021) 117, <https://doi.org/10.1186/s12938-021-00952-x>.
- [3] Z. Li, D. Jiang, S. Yang, MiR-490-3p inhibits the malignant progression of lung adenocarcinoma, *Cancer Manag. Res.* 12 (2020) 10975–10984, <https://doi.org/10.2147/cmar.S258182>.
- [4] M. Xiao, et al., AC138128.1 an intronic lncRNA originating from ERCC1 implies a potential application in lung cancer treatment, *J. Cancer* 10 (2019) 3608–3617, <https://doi.org/10.7150/jca.31832>.
- [5] X. Lv, et al., Polymorphism in lncRNA AC008392.1 and its interaction with smoking on the risk of lung cancer in a Chinese population, *Cancer Manag. Res.* 10 (2018) 1377–1387, <https://doi.org/10.2147/cmar.S160818>.
- [6] C. Zeltz, et al., LOXL1 is regulated by integrin $\alpha 11$ and promotes non-small cell lung cancer tumorigenicity, *Cancers (Basel)* 11 (2019), <https://doi.org/10.3390/cancers11050705>.
- [7] P. Sharma, et al., Primary, adaptive, and acquired resistance to cancer immunotherapy, *Cell* 168 (2017) 707–723, <https://doi.org/10.1016/j.cell.2017.01.017>.
- [8] S. Devalaraja, et al., Tumor-derived retinoic acid regulates intratumoral monocyte differentiation to promote immune suppression, *Cell* 180 (2020) 1098–1114. e16, <https://doi.org/10.1016/j.cell.2020.02.042>.
- [9] W. Chung, et al., Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer, *Nat. Commun.* 8 (2017) 15081, <https://doi.org/10.1038/ncomms15081>.
- [10] P.M. Ellis, Anti-angiogenesis in personalized therapy of lung cancer, *Adv. Exp. Med. Biol.* 893 (2016) 91–126, https://doi.org/10.1007/978-3-319-24223-1_5.
- [11] P.S. Hegde, V. Karanikas, S. Evers, The where, the when, and the how of immune monitoring for cancer immunotherapies in the era of checkpoint inhibition, *Clin. Cancer Res.* 22 (2016) 1865–1874, <https://doi.org/10.1158/1078-0432.Ccr-15-1507>.
- [12] M. Liu, et al., Targeting the Ido1 pathway in cancer: from bench to bedside, *J. Hematol. Oncol.* 11 (2018) 100, <https://doi.org/10.1186/s13045-018-0644-y>.
- [13] E.M. Van Allen, et al., Genomic correlates of response to CTLA-4 blockade in metastatic melanoma, *Science* 350 (2015) 207–211, <https://doi.org/10.1126/science.aad0095>.
- [14] P. Perdiguer, et al., Insights into the evolution of the prdm1/blimp1 gene family in teleost fish, *Front. Immunol.* 11 (2020) 596975, <https://doi.org/10.3389/fimmu.2020.596975>.
- [15] M.L.H. Katayama, et al., Stromal cell signature associated with response to neoadjuvant chemotherapy in locally advanced breast cancer, *Cells* 8 (2019), <https://doi.org/10.3390/cells8121566>.
- [16] J. Goveia, et al., An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates, *Cancer Cell* 37 (2020) 421, <https://doi.org/10.1016/j.ccell.2020.03.002>.
- [17] J. Eugène, et al., The inhibitory receptor CD94/NKG2A on CD8(+) tumor-infiltrating lymphocytes in colorectal cancer: a promising new druggable immune checkpoint in the context of HLA-E/ $\beta 2m$ overexpression, *Mod. Pathol.* 33 (2020) 468–482, <https://doi.org/10.1038/s41379-019-0322-9>.
- [18] Z. Fan, et al., Nrf2-Keap1 pathway promotes cell proliferation and diminishes ferroptosis, *Oncogenesis* 6 (2017) e371, <https://doi.org/10.1038/oncsis.2017.65>.
- [19] M. Ayers, et al., IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade, *J. Clin. Invest.* 127 (2017) 2930–2940, <https://doi.org/10.1172/jci91190>.
- [20] N.S. Patil, et al., Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer, *Cancer Cell* 40 (2022) 289–300.e4, <https://doi.org/10.1016/j.ccell.2022.02.002>.
- [21] Q. Fu, et al., Prognostic value of tumor-infiltrating lymphocytes in melanoma: a systematic review and meta-analysis, *OncoImmunology* 8 (2019) 1593806, <https://doi.org/10.1080/2162402x.2019.1593806>.
- [22] W. Scheper, et al., Low and variable tumor reactivity of the intratumoral TCR repertoire in human cancers, *Nat. Med.* 25 (2019) 89–94, <https://doi.org/10.1038/s41591-018-0266-5>.

- [23] Z. Liu, et al., Detecting tumor antigen-specific T cells via interaction-dependent fucosyl-biotinylation, *Cell* 183 (2020) 1117–1133.e19, <https://doi.org/10.1016/j.cell.2020.09.048>.
- [24] M.A. Socinski, et al., Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC, *N. Engl. J. Med.* 378 (2018) 2288–2301, <https://doi.org/10.1056/NEJMoa1716948>.
- [25] F. Zhou, et al., Immune-checkpoint inhibitors plus chemotherapy versus chemotherapy as first-line treatment for patients with extensive-stage small cell lung cancer, *J. Immunother. Cancer* 8 (2020), <https://doi.org/10.1136/jitc-2020-001300>.
- [26] R.S. Herbst, et al., Ramucicromab plus pembrolizumab in patients with previously treated advanced non-small-cell lung cancer, gastro-oesophageal cancer, or urothelial carcinomas (JVDf): a multicohort, non-randomised, open-label, phase 1a/b trial, *Lancet Oncol.* 20 (2019) 1109–1123, [https://doi.org/10.1016/s1470-2045\(19\)30458-9](https://doi.org/10.1016/s1470-2045(19)30458-9).
- [27] A.C. Scott, et al., TOX is a critical regulator of tumour-specific T cell differentiation, *Nature* 571 (2019) 270–274, <https://doi.org/10.1038/s41586-019-1324-y>.
- [28] M. De Simone, et al., Transcriptional landscape of human tissue lymphocytes unveils uniqueness of tumor-infiltrating T regulatory cells, *Immunity* 45 (2016) 1135–1147, <https://doi.org/10.1016/j.immuni.2016.10.021>.
- [29] J.M. Xu, et al., Blockade of the CXCR6 signaling inhibits growth and invasion of hepatocellular carcinoma cells through inhibition of the VEGF expression, *Int. J. Immunopathol. Pharmacol.* 27 (2014) 553–561, <https://doi.org/10.1177/039463201402700411>.
- [30] M. Di Pilato, et al., CXCR6 positions cytotoxic T cells to receive critical survival signals in the tumor microenvironment, *Cell* 184 (2021) 4512–4530.e22, <https://doi.org/10.1016/j.cell.2021.07.015>.
- [31] L. Zheng, et al., Pan-cancer single-cell landscape of tumor-infiltrating T cells, *Science* 374 (2021) abe6474, <https://doi.org/10.1126/science.abe6474>.
- [32] X. Zhang, et al., Blocking CD47 efficiently potentiated therapeutic effects of anti-angiogenic therapy in non-small cell lung cancer, *J. Immunother. Cancer* 7 (2019) 346, <https://doi.org/10.1186/s40425-019-0812-9>.
- [33] H. Ding, et al., Long distance from microvessel to cancer cell predicts poor prognosis in non-small cell lung cancer patients, *Front. Oncol.* 11 (2021) 632352, <https://doi.org/10.3389/fonc.2021.632352>.
- [34] M.J. Siemerink, et al., Endothelial tip cells in ocular angiogenesis: potential target for anti-angiogenesis therapy, *J. Histochem. Cytochem.* 61 (2013) 101–115, <https://doi.org/10.1136/0022155412467635>.
- [35] P. Carmeliet, R.K. Jain, Molecular mechanisms and clinical applications of angiogenesis, *Nature* 473 (2011) 298–307, <https://doi.org/10.1038/nature10144>.
- [36] N. Lee, et al., HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A, *Proc. Natl. Acad. Sci. U.S.A.* 95 (1998) 5199–5204, <https://doi.org/10.1073/pnas.95.9.5199>.
- [37] M. Zheng, et al., Functional exhaustion of antiviral lymphocytes in COVID-19 patients, *Cell. Mol. Immunol.* 17 (2020) 533–535, <https://doi.org/10.1038/s41423-020-0402-2>.
- [38] R.S. Herbst, et al., Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients, *Nature* 515 (2014) 563–567, <https://doi.org/10.1038/nature14011>.
- [39] M.W. Teng, et al., Classifying cancers based on T-cell infiltration and PD-L1, *Cancer Res.* 75 (2015) 2139–2145, <https://doi.org/10.1158/0008-5472.Can-15-0255>.
- [40] R.M. Samstein, et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types, *Nat. Genet.* 51 (2019) 202–206, <https://doi.org/10.1038/s41588-018-0312-8>.
- [41] G.M. Blumenthal, et al., Milestone analyses of immune checkpoint inhibitors, targeted therapy, and conventional therapy in metastatic non-small cell lung cancer trials: a meta-analysis, *JAMA Oncol.* 3 (2017) e171029, <https://doi.org/10.1001/jamaoncol.2017.1029>.
- [42] S. Mariathasan, et al., TGF β attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells, *Nature* 554 (2018) 544–548, <https://doi.org/10.1038/nature25501>.
- [43] M. Goldman, et al., The UCSC Xena platform for public and private cancer genomics data visualization and interpretation, *bioRxiv* (2018) 326470.
- [44] K. Cibulskis, et al., Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples, *Nat. Biotechnol.* 31 (2013) 213–219.
- [45] A. Mayakonda, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, *Genome Res.* 28 (2018) 1747–1756.
- [46] C.H. Mermel, et al., GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers, *Genome Biol.* 12 (2011) R41, <https://doi.org/10.1186/gb-2011-12-4-r41>.
- [47] A. Butler, et al., Integrating single-cell transcriptomic data across different conditions, technologies, and species, *Nat. Biotechnol.* 36 (2018) 411–420, <https://doi.org/10.1038/nbt.4096>.
- [48] D. Wu, G.K. Smyth, Camera: a competitive gene set test accounting for inter-gene correlation, *Nucleic Acids Res.* 40 (2012) e133, <https://doi.org/10.1093/nar/gks461>.
- [49] I. Korsunsky, et al., Fast, sensitive and accurate integration of single-cell data with Harmony, *Nat. Methods* 16 (2019) 1289–1296, <https://doi.org/10.1038/s41592-019-0619-0>.
- [50] C. Zheng, et al., Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing, *Cell* 169 (2017) 1342–1356.e16, <https://doi.org/10.1016/j.cell.2017.05.035>.
- [51] C. Trapnell, et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells, *Nat. Biotechnol.* 32 (2014) 381–386, <https://doi.org/10.1038/nbt.2859>.
- [52] M. Efremova, et al., CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes, *Nat. Protoc.* 15 (2020) 1484–1506, <https://doi.org/10.1038/s41596-020-0292-x>.
- [53] W. Cheng, et al., Bioinformatic profiling identifies an immune-related risk signature for glioblastoma, *Neurology* 86 (2016) 2226–2234, <https://doi.org/10.1212/wnl.0000000000002770>.
- [54] Z.S. Bao, et al., Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling, *CNS Neurosci. Ther.* 20 (2014) 112–118, <https://doi.org/10.1111/cns.12171>.
- [55] G. Yu, et al., clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters, vol. 16, *OMICS*, 2012, pp. 284–287, <https://doi.org/10.1089/omi.2011.0118>.
- [56] B. Li, et al., Comprehensive analyses of tumor immunity: implications for cancer immunotherapy, *Genome Biol.* 17 (2016) 174, <https://doi.org/10.1186/s13059-016-1028-7>.
- [57] E. Becht, et al., Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression, *Genome Biol.* 17 (2016) 218, <https://doi.org/10.1186/s13059-016-1070-5>.
- [58] D.C. Gulhan, et al., Detecting the mutational signature of homologous recombination deficiency in clinical samples, *Nat. Genet.* 51 (2019) 912–919, <https://doi.org/10.1038/s41588-019-0390-2>.
- [59] L.B. Alexandrov, et al., Deciphering signatures of mutational processes operative in human cancer, *Cell Rep.* 3 (2013) 246–259, <https://doi.org/10.1016/j.celrep.2012.12.008>.