



Research article

Comprehensive analysis of single-cell RNA and bulk RNA sequencing based on M2 tumor-associated macrophage and angiogenesis-related genes to assess prognosis and therapeutic response in lung adenocarcinoma

Anbang Liu^{a,1}, Gengqiu Liu^{b,1}, Xiaohuai Wang^b, Dongqing Yan^b, Junhang Zhang^{b,*}, Li Wei^{b,**}

^a Department of Thoracic Surgery, Qingdao Municipal Hospital, 266000, Qingdao, Shandong, China

^b Department of Thoracic Surgery, The Seventh Affiliated Hospital, Sun Yat-sen University, No. 628, Zhenyuan Road, Guangming Dist., Shenzhen, 518107, China

ARTICLE INFO

Keywords:
LUAD
M2 TAM
Angiogenesis
Prognosis
Therapeutic response

ABSTRACT

M2 tumor-associated macrophage (M2 TAM), a crucial component of the tumor microenvironment, has a significant impact on tumor invasion and metastasis in the form of angiogenesis for lung adenocarcinoma (LUAD). In this study, both single-cell RNA and bulk RNA sequencing data were analyzed to identify 12 M2 TAM and angiogenesis-related genes (OLR1, CTSL, HLA-DPB1, NUPR1, ALOX5, DOCK4, CSF2RB, PTPN6, TNFSF12, HNRNPA2B1, NCL, and BIRC2). These genes were used to construct a prognostic signature, which was subsequently validated using an external cohort. Moreover, the immune profile analysis indicated that the low-risk group exhibited a distinct immune cell infiltration and relatively active status. Importantly, the prognostic signature was closely associated with PD-1, CTLA4, tumor mutation burden, and anti-cancer drug sensitivity. In summary, this study proposes a new prognostic signature for patients with LUAD based on M2 TAM and angiogenesis-related genes. The signature forecasts the prognosis of LUAD by an independent manner, reveals the potential molecular mechanisms involved in tumor immune-related functions, and offers appropriate clinical strategies for the treatment of patients with LUAD.

1. Introduction

Lung cancer is the second most common and deadliest cancer around the world, representing 18.0 % of all cancer deaths [1]. The number of patients with lung adenocarcinoma (LUAD) is increasing and has become the most prevalent subtype of lung cancer [2]. Despite promising advances in various drug treatments for LUAD so far, there is still a low 5-year survival rate for patients with LUAD. Recently, immunotherapy has become a prospective strategy for treating cancer, however only a proportion of patients with LUAD can

* Corresponding author.

** Corresponding author.

E-mail addresses: liuanb@mail2.sysu.edu.cn (A. Liu), liugq8@mail2.sysu.edu.cn (G. Liu), wangxh273@mail2.sysu.edu.cn (X. Wang), yandq8@mail.sysu.edu.cn (D. Yan), zhangjh33@mail.sysu.edu.cn (J. Zhang), weili8@mail.sysu.edu.cn (L. Wei).

¹ Anbang Liu and ¹Gengqiu Liu contributed equally to this work and share first authorship.

tolerate and benefit from treatment with immune checkpoint inhibitors (ICIs) [3]. Hence, it is necessary to develop appropriate biomarkers for predicting prognostic status and treatment response of LUAD [4].

The tumor microenvironment (TME) is known as the environment surrounding the tumor, including stromal cells, immune cells, and extracellular matrix, which are strongly associated with tumor progression and treatment outcomes [5]. Among them, tumor-associated macrophage (TAM) is the predominant portion of immune cells in TME and has a vital effect on tumor progression and treatment. Macrophages are used to be distinguished into two phenotypes: the pro-inflammatory M1 type and the anti-inflammatory M2 type [6,7]. M1 macrophages generate pro-inflammatory molecules, including iNOS, IL-12, IL-23 and TNF- α , which trigger inflammation and have anti-tumor effects. M2 macrophages generate anti-inflammatory cytokines, including TGF- β , Arg-1 and IL-10, which contribute to immunosuppressive and tumor promoting effects [8]. TAM typically exhibits M2-like properties, promoting tumor growth and metastasis to the periphery of solid tumors [9]. TAMs interact with various immune cells in the TME by inhibiting differentiated cluster CD8⁺ T cells, inducing dysfunction of natural killer (NK) and NK T cells, and amplifying T-regulatory cells (Tregs) to indirectly suppress effector T cells, thus accelerating tumorigenesis and progression by decreasing the number of anti-tumor immune cells [10]. In addition, angiogenesis has a significant impact on tumorigenesis, progression and metastasis in lung cancer [11,12]. TAMs are important drivers of angiogenesis, producing angiogenic growth factors, cytosolic proteases and matrix metalloproteinases, to create a suitable microenvironment for angiogenesis [13–15]. Therefore, there is a need to explore the potential association between M2 TAM and angiogenesis in patients with LUAD and to discover biomarkers for predicting prognosis.

Single-cell RNA sequencing (scRNA-seq) provides a basis for individualized therapy by effectively exploring the mechanisms of tumor heterogeneity and evolution [16,17]. In addition, scRNA-seq analysis of immune cells in the TME helps to dissect the molecular signature of immune cells, which provides new insights into cancer immunity [18,19]. The development of a genetic signature based on the molecular characteristics of immune cells may be an effective way to predict the treatment outcomes and prognoses of patients with cancer. In this study, a comprehensive analysis of scRNA-seq and bulk RNA sequencing for LUAD was conducted to identify M2 TAM marker genes and angiogenesis-related genes, and the signature of M2 TAM and angiogenesis was developed to predict the status of prognosis in LUAD. Then, the relationship between the signature based on both M2 TAM and angiogenesis-related genes and the immune cell infiltration landscape, immune checkpoint, tumor mutation burden (TMB) and chemotherapy drug sensitivity was further analyzed to assess the treatment response of LUAD.

2. Methods

2.1. Acquisition and quality control of scRNA-seq data

The single-cell transcriptome profiles of LUAD were obtained from the study by Bischoff et al. [20]. 10 normal lung samples and 10 fresh tumor tissue samples were collected from 12 patients with LUAD. The scRNA-seq data were analyzed using the “Seurat” software package [21], providing data filtering (cellular and genetic), normalization, and unified flow approximation and projection (UMAP). By filtering the transcriptomic data of the cells, 500–10,000 genes were detected, and 1000–100,000 unique molecular identifiers (UMIs) were calculated, with a fraction of hemoglobin reads <5 % and a fraction of mitochondrial reads <30 %. The normalization with 2000 variable features was then used for variance-stabilization of UMI counts [22], and the fraction of mitochondrial reads and the number of UMIs were regressed. The UMAP algorithm was used for cell clustering of the top 20 principal components (PCs). M2 TAM markers were adapted from the CellMarker database (<http://xteam.xbio.top/CellMarker/index.jsp>), while the other immune cell type markers and the main cell types markers were obtained from Habermann et al. [23] and Tata et al. [24] (Additional file 1: Table S1). The “CellChat” package was used to predict the communication relationship between immune cells [25]. The function “FindAllMarkers” was used to find out the marker genes of each immune cell type.

2.2. Acquisition of bulk RNA-sequencing data

The RNA-seq data, associated clinical parameters, and gene mutation data of LUAD were obtained from The Cancer Genome Atlas (TCGA) database, TPM was adopted as the format of bulk RNA-seq data. The LUAD dataset GSE72094, which included RNA-seq data and clinical characteristics, was obtained from the Gene Expression Omnibus (GEO) database as an external verification set. All samples with overall survival (OS) < 30 days were excluded from TCGA-LUAD and GSE72094. The GeneCards database (<https://www.genecards.org/>) provided 4476 genes involved in angiogenesis (Additional file 1: Table S2). The clinicopathological information of all patients was presented in Additional file 1: Table S3.

2.3. Acquisition of both M2 TAM and angiogenesis-related genes

In each TCGA-LUAD sample, the relative immune infiltration content of M2 macrophages was calculated using the R package “CIBERSORT” [26]. The genes most associated with the immune infiltration content of M2 macrophages were obtained by analyzing TCGA-LUAD RNA expression data using the “WGCNA” package [27]. The overall correlation of all samples was determined by clustering samples to exclude outliers. The soft threshold power β was selected given the lowest power, and the minimum number of genes per module was set to 30. Then, the most relevant modules for M2 macrophage content were identified by performing a correlation analysis between traits and modules. Finally, the obtained modular genes were intersected with angiogenesis-related genes (ARGs) and M2 TAM marker genes to screen out M2 TAM- and angiogenesis-related genes. The R package “clusterProfiler” was used to achieve the functional analysis of the intersected genes through the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene

Ontology (GO) [28].

2.4. Construction and validation of the prognostic signature

To establish the predictive signature of both M2 TAM and angiogenesis-related genes, the univariate Cox regression analysis was initially conducted. The least absolute shrinkage and selection operator (LASSO) was further used to take 10-fold cross validation through the “glmnet” package, the variable with non-zero coefficient used was lambda.min [29,30]. Lastly, the multivariate Cox regression analysis was used to identify an efficient prognostic signature of both M2 TAM and angiogenesis-related genes based on the following computational formula (1):

$$\text{Risk score} = \sum_{i=1}^n \text{Coef}(i) \times \text{Expr}(i)$$

Coef(i) and Expr(i) indicate the multivariate Cox regression coefficient for each gene and the corresponding expression levels, respectively. Furthermore, according to the median value of risk score, patients with LUAD were classified into two categories: high-risk and low-risk. Kaplan-Meier survival curves were used to compare the prognosis between the two risk groups. The “survivalROC” package was used to plot 1-,3-and 5-year receiver operator characteristic (ROC) curves for assessing the performance of the prognostic signature [31]. Additionally, a nomogram was developed for the purpose of predicting LUAD survival at 1, 3, and 5 years. The nomogram predictive performance was evaluated by a calibration curve, the time-dependent C-index and decision curve analysis (DCA) using the “rms”, “pec” and “rmda” packages, respectively. According to the KEGG database, gene set enrichment analysis (GSEA) 4.1.0 was used to investigate which pathways were most enriched in two risk groups [32]. Nominal p-value <0.05 and normalized enrichment score |NES| ≥1 were taken into account as our standard for statistical significance.

2.5. Analysis of immune cells and functions

To perform immune-related analysis of the prognostic signature, including infiltration of immune and stromal cells, the ESTIMATEScore, ImmuneScore, and StromalScore were calculated by the R package “estimate” [33], respectively. The infiltration score of immune cells was computed to explore the biological functions of immune-related pathways between two risk groups through the “GSVA” and “GSEABase” packages [34].

2.6. Analysis of somatic mutation and predictive efficacy of LUAD therapy

The exon length and mutation frequency in each patient were also calculated. The package “maftools” was used to generate waterfall plots and describe TMB values to identify gene mutation characteristics [35]. Patients with LUAD were classified into low- and high-TMB groups according to the median value of TMB, and the difference was compared in terms of the signature-gene expression. To investigate the potential of the signature as a therapeutic guide for LUAD, we employed IC50 as a reference to assess its predictive value in chemotherapy and targeted therapy. The R package “pRRophetic” was utilized to calculate the IC50 values of commonly used chemotherapeutic and targeted drugs for the treatment of LUAD [36].

2.7. Sample gathering and real-time reverse transcription-quantitative polymerase chain reaction

Tumor and adjacent normal tissue samples were collected from 10 patients with LUAD. A summary of the detailed clinical information was listed in Additional file 1: Table S4. Total RNA of human tissues was extracted using TRIzol® Reagent. RNA was reversely transcribed into complementary DNA through Evo M-MLV RT premix. The SYBR-Green Pro Taq HS qPCR Kit II (Accurate Biology) was used to perform qPCR for a real-time PCR detection system and all the primers of PCR were displayed in Additional file 1: Table S5. The relative mRNA expression of genes was normalized to that of β-actin and presented by the $2^{-\Delta\Delta Ct}$ method.

2.8. Statistical analysis

All statistical analyses were conducted using R software (version 4.0.5). If not otherwise stated, statistical significance was set as p < 0.05. The difference between the two groups was compared by the Wilcoxon signed-rank test or t-test. The difference among three or more groups was compared through the Kruska-Wallis test. The survival analysis was conducted using the Kaplan-Meier method through the R package “survminer” and “survival”. The predictive efficacy was assessed based on the calculation of the area under the curves (AUCs).

3. Results

3.1. Identification of main cell types

A general flowchart of this study was observed in Fig. 1. The number of genes (nFeature), the seq-count per cell (nCount), and the percentage of mitochondrial genes (pMT) were presented in Vlnplot (Fig. 2A). 26,296 genes were obtained from 114,489 cells after

quality control. After filtering, UMAPs were color-coded by tissue and sample types according to the top 20 PCs of all single-cell transcriptomes (Fig. 2B and C). To further classify the main cell types, differentially expressed marker genes were identified from the 17 clusters and visualized using the UMAP algorithm (Fig. 2D). The expression levels of stromal, immune, and epithelial marker gene were analyzed on the basis of single-cell transcriptomes (Fig. 2E). In total, 4454 stromal, 88,335 immunes, and 21,700 epithelial single-cell transcriptomes were covered by manual annotation (Fig. 2F). Fig. 2G showed the count proportion of the main cell types in each sample.

3.2. Identification of M2 TAM marker genes

A total of 43,378 immune cells from tumor tissues were obtained by isolating immune cells from normal and tumor tissues. To classify the immune cell types in tumor tissues, differentially expressed marker genes were identified from 9 clusters and visualized using the UMAP algorithm (Fig. 3A). The expression of immune cell marker genes was analyzed (Fig. 3B), and M2 TAM, mast cells, T cells, NK cells, dendritic cells, and B cells were identified by manual annotation (Fig. 3C). The proportions of immune cell types in each tissue sample were shown in Fig. 3D. Fig. 3E and F showed the communication relationship between different immune cell types. Then, marker genes of each immune cell type were detected in tumor samples (Fig. 3G), 1 and 555 M2 TAM marker genes were obtained

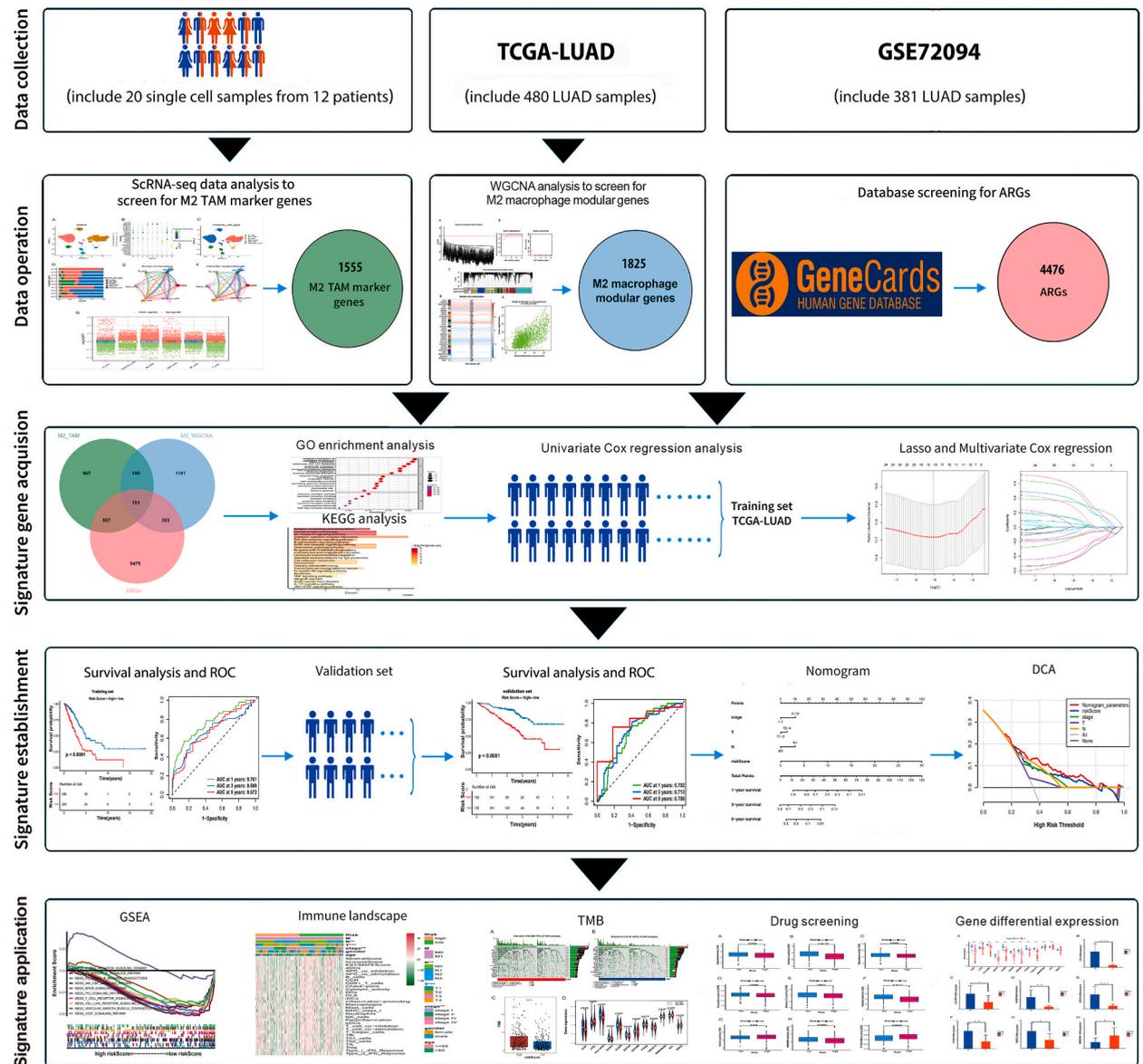


Fig. 1. The flow chart of our research.

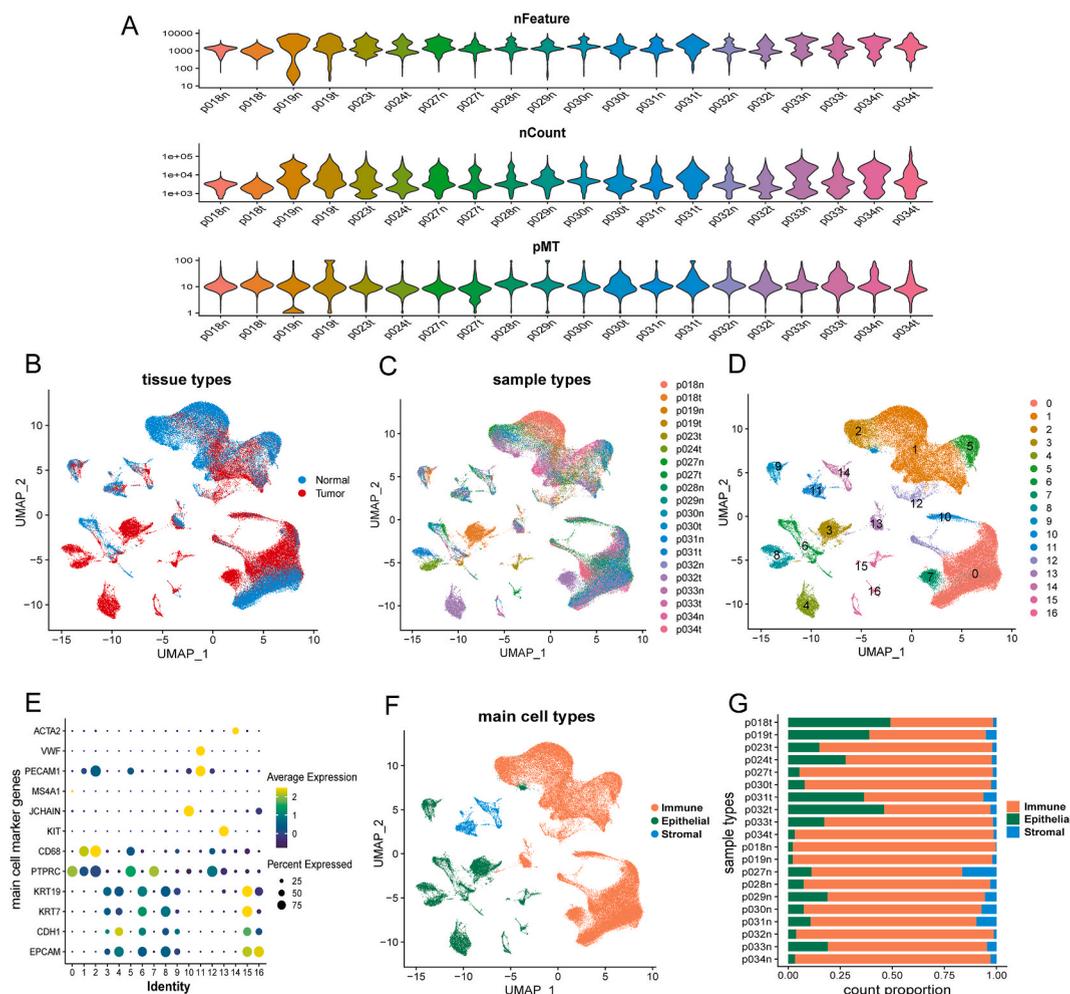


Fig. 2. Processing of the scRNA-seq data and distinguishing main cell types. (A) Violin plot presented the nFeature, nCount, and pMT according to the scRNA-seq data of the LUAD cell samples. (B, C) Tissue types and sample types in accordance with the top 20 PCs of each single-cell transcriptome using the UMAP algorithm. (D) Visualization of differentially expressed marker genes of 17 clusters. (E) Bubble plot to express immune, epithelial and stromal marker genes of 17 clusters. (F) Detailed annotation of main cell types through the UMAP algorithm (G) Bar plot for the count proportion of main cell types in each sample.

(Additional file 1: [Table S6](#)).

3.3. Screening of genes linked to M2 macrophage by WGCNA

Genes linked to M2 macrophage were identified by WGCNA and 5 outliers were removed from TCGA-LUAD (Fig. 4A). The optimal soft-threshold power was determined to be 4 (Fig. 4B), and 31 modules were generated (Fig. 4C). Correlation analysis indicated that the green module had a significant association with high M2 macrophage content (correlation = 0.30, $P < 0.001$) (Fig. 4D). Moreover, a clear positive relationship was observed between module membership and gene significance in the green module (correlation = 0.65, $P < 0.001$) (Fig. 4E). Therefore, 1825 genes in the green module were selected for further analysis (Additional file 1: [Table S7](#)).

3.4. Screening of both M2 TAM and angiogenesis-related prognostic genes

Through the intersection of 4476 ARGs, 1825 M2 macrophage modular genes and 1555 M2 TAM marker genes, 191 candidate genes associated with both M2 TAM and angiogenesis were obtained (Fig. 5A, Additional file 1: [Table S8](#)). As shown in the KEGG analysis, 23 pathways were displayed in a bar plot, such as B cell receptor, NOD-like receptor, Toll-like receptor, and NF-kappa B signaling pathways (Fig. 5B). GO analysis revealed that the biological process (BP) category included activation of leukocyte in the immune response, regulation of the immune effector process, and positive regulation of cytokine production, etc. The cell component (CC) category included endocytic vesicle, cell-substrate junction, and the external side of the plasma membrane, etc. The molecular function (MF) category included immune receptor activity, cytokine receptor binding, and integrin binding, etc. (Fig. 5C). Following the results

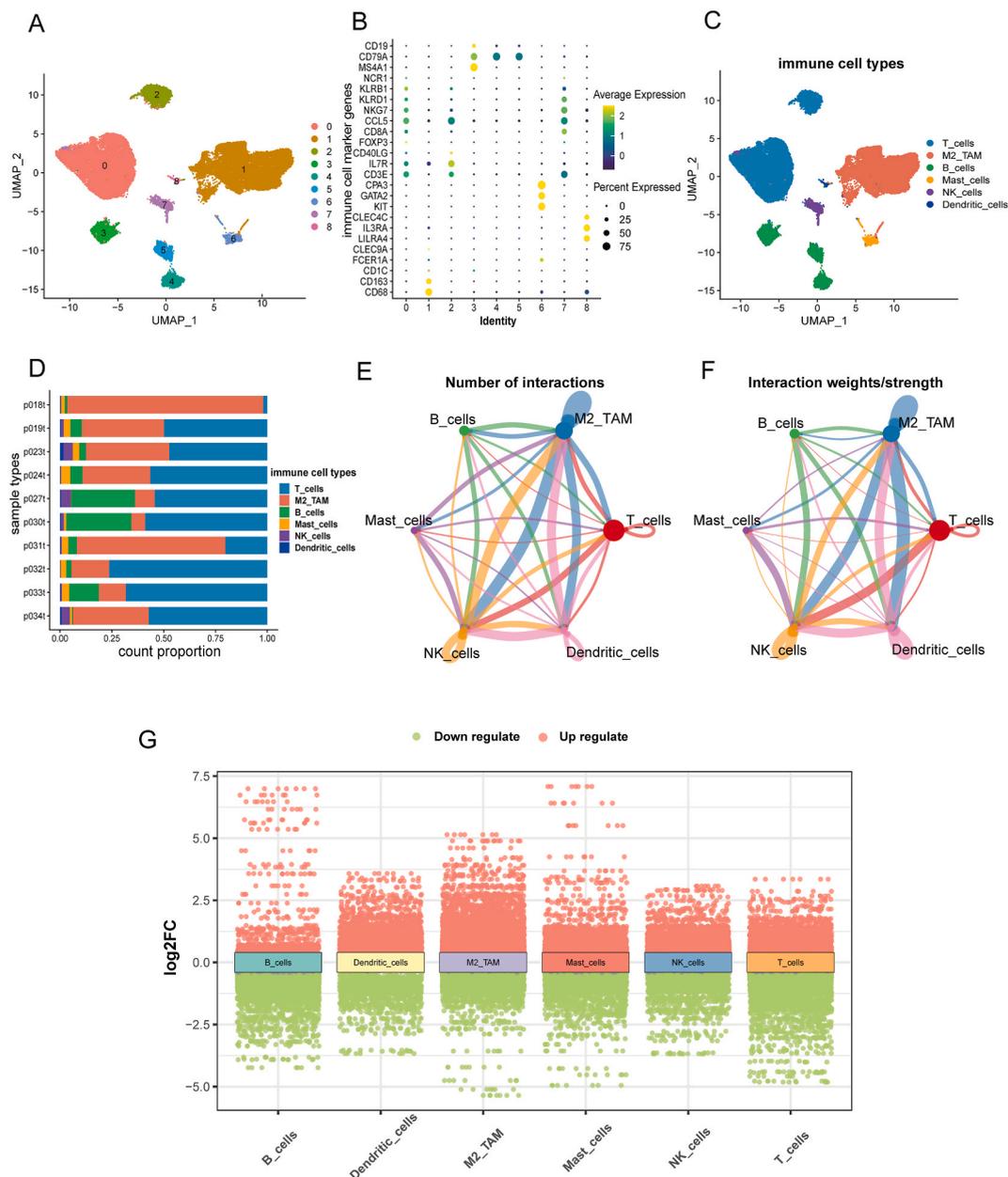


Fig. 3. Identification of M2 TAM marker genes in tumor samples. (A) Immune cell visualization of 9 clusters based on the UMAP algorithm. (B) Bubble plot to express immune cell marker genes of 9 clusters. (C) Detailed annotation of immune cell types through the UMAP algorithm (D) Bar plot for the count proportion of immune cell types in all tumor samples. (E, F) Communication relationship between the immune cell types. (G) Marker genes of every immune cell type in tumor samples.

of the univariate Cox regression analysis, a total of 24 genes were found to be associated with the prognosis of LUAD (Fig. 5D). The LASSO and the multivariate Cox regression analysis identified 12 genes related to both M2 TAM and angiogenesis (OLR1, CTSL, HLA-DPB1, NUPR1, ALOX5, DOCK4, CSF2RB, PTPN6, TNFSF12, HNRNPA2B1, NCL, and BIRC2) and constructed a prognostic signature (Additional file 2: Fig. S1, Additional file 1: Table S9).

3.5. Construction of the prognostic signature related to both M2 TAM and angiogenesis

Given the designed formula (2), risk score = $(-0.213 \times \text{OLR1 expression}) + (0.368 \times \text{CTSL expression}) + (0.229 \times \text{HLA-DPB1 expression}) + (-0.472 \times \text{NUPR1 expression}) + (0.190 \times \text{ALOX5 expression}) + (-0.584 \times \text{DOCK4 expression}) + (-0.243 \times \text{CSF2RB expression}) + (-0.656 \times \text{PTPN6 expression}) + (0.449 \times \text{TNFSF12 expression}) + (0.683 \times \text{HNRNPA2B1 expression}) + (0.506 \times \text{NCL}$

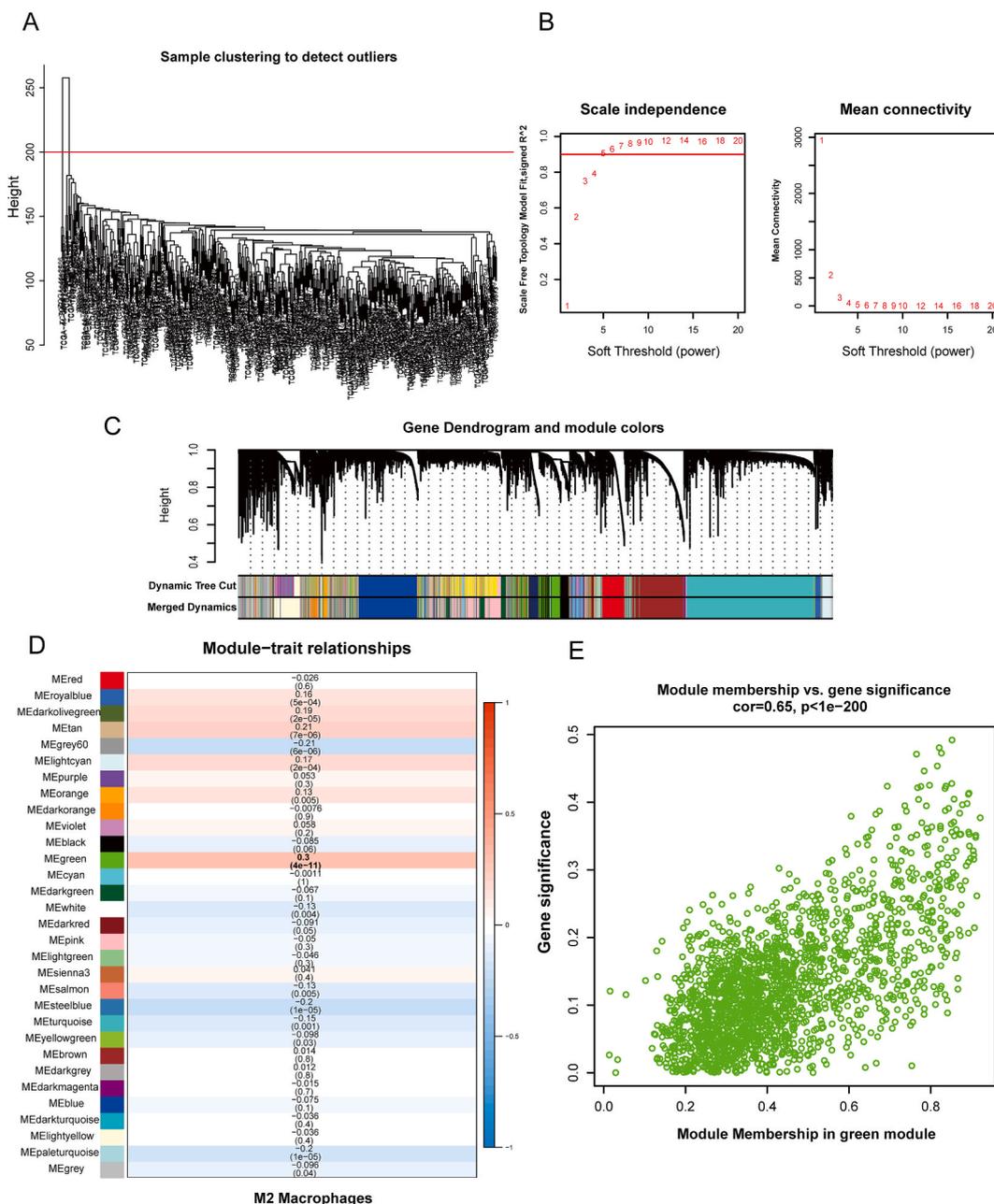


Fig. 4. Screening M2 macrophage related genes through WGCNA. (A) 480 TCGA-LUAD samples were clustered, and 5 outlier samples were removed. (B) WGCNA package showed 4 was chosen as the soft threshold power. (C) Gene dendrogram showed 31 modules were generated. (D) In correlation analysis between modules and traits, the green module was considered the most related module for M2 macrophages. (E) The significantly positive association between gene significance and module membership within the green module.

expression) + (0.747 × BIRC2 expression). Subsequently, high-risk and low-risk groups were distinguished based on the median value of the risk score (Fig. 6A). As the score climbed, the likelihood of death increased (Fig. 6B). Survival analysis indicated that compared to those at high risk, low-risk patients had a significantly longer OS (Fig. 6C, $p < 0.001$). The AUCs of survival at 1, 3, and 5 years were 0.761, 0.669 and 0.672, respectively, manifesting relatively good predictive performance (Fig. 6D). Compared to other clinicopathological factors, the risk score (AUC = 0.645) and stage (AUC = 0.649) had almost the same predictive power (Fig. 6E). Univariate Cox regression analysis revealed that stage, T, N, M and risk score had a close association with the OS of LUAD patients (Fig. 6F). The risk score and stage were identified as two major independent predictors by the multivariate Cox regression analysis (Fig. 6G).

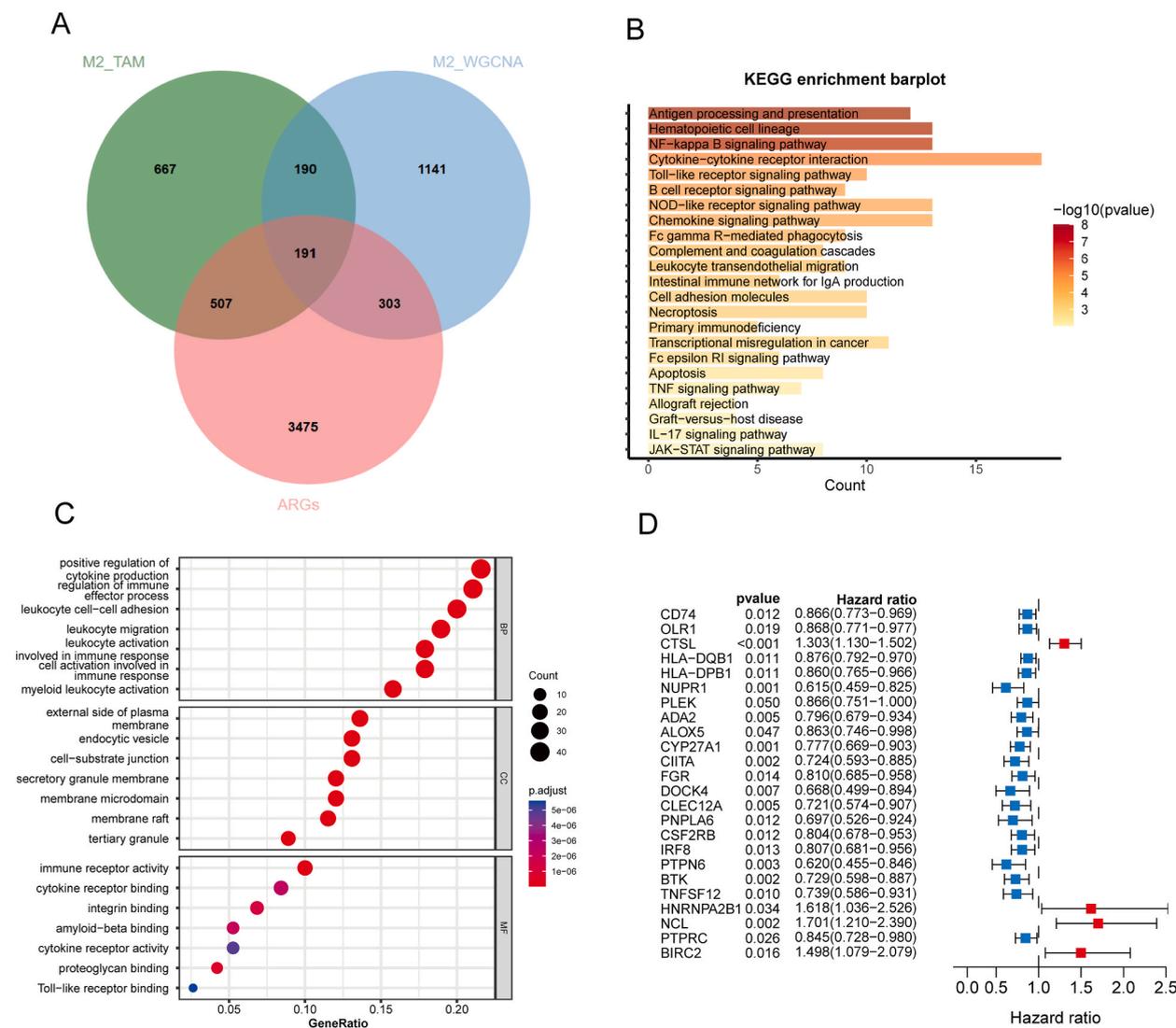


Fig. 5. Screening of both angiogenesis and M2 TAM-related prognostic genes. (A) Acquisition of 191 candidate genes associated with both angiogenesis and M2 TAM through Venn diagram. (B, C) KEGG and GO analysis of both angiogenesis and M2 TAM-related genes, respectively. (D) A univariate Cox regression analysis identified that 24 genes were related to the LUAD prognosis.

3.6. External validation of the prognostic signature

External validation was further taken to confirm the prognostic signature's reliability. In the same way, all samples were grouped based on the median value of the risk score and the increased risk score also enhanced the likelihood of death in the validation set (Fig. 7A and B). The prognosis in the low-risk group was worse than that in the high-risk group (Fig. 7C, $p < 0.001$). In the validation cohort, the AUCs of 1-year, 3-year and 5-year survival were 0.702, 0.713 and 0.788, respectively (Fig. 7D). The risk score emerged as the most robust predictor among all factors, as evidenced by its highest AUC value (AUC = 0.759), as depicted in Fig. 7E. In the univariate and multivariate Cox regression analysis, the risk score, gender and stage were deemed to be independent predictors (Fig. 7F and G). On the basis of these findings, it is proposed that M2 TAM and angiogenesis-related prognostic signature was reliable predictors of LUAD prognosis.

3.7. The relationship between clinicopathological characteristics and the prognostic signature

The data of 328 TCGA-LUAD patients, including all expression data, survival data and clinicopathological information, were obtained to examine the connection between the risk score and patient prognosis. Based on different clinicopathological factors, these samples were classified in various clinical subgroups. The results indicated that the prognostic signature constructed both M2 TAM and angiogenesis-related genes as a risk factor was a relatively better predictor for LUAD across various clinicopathological subgroups

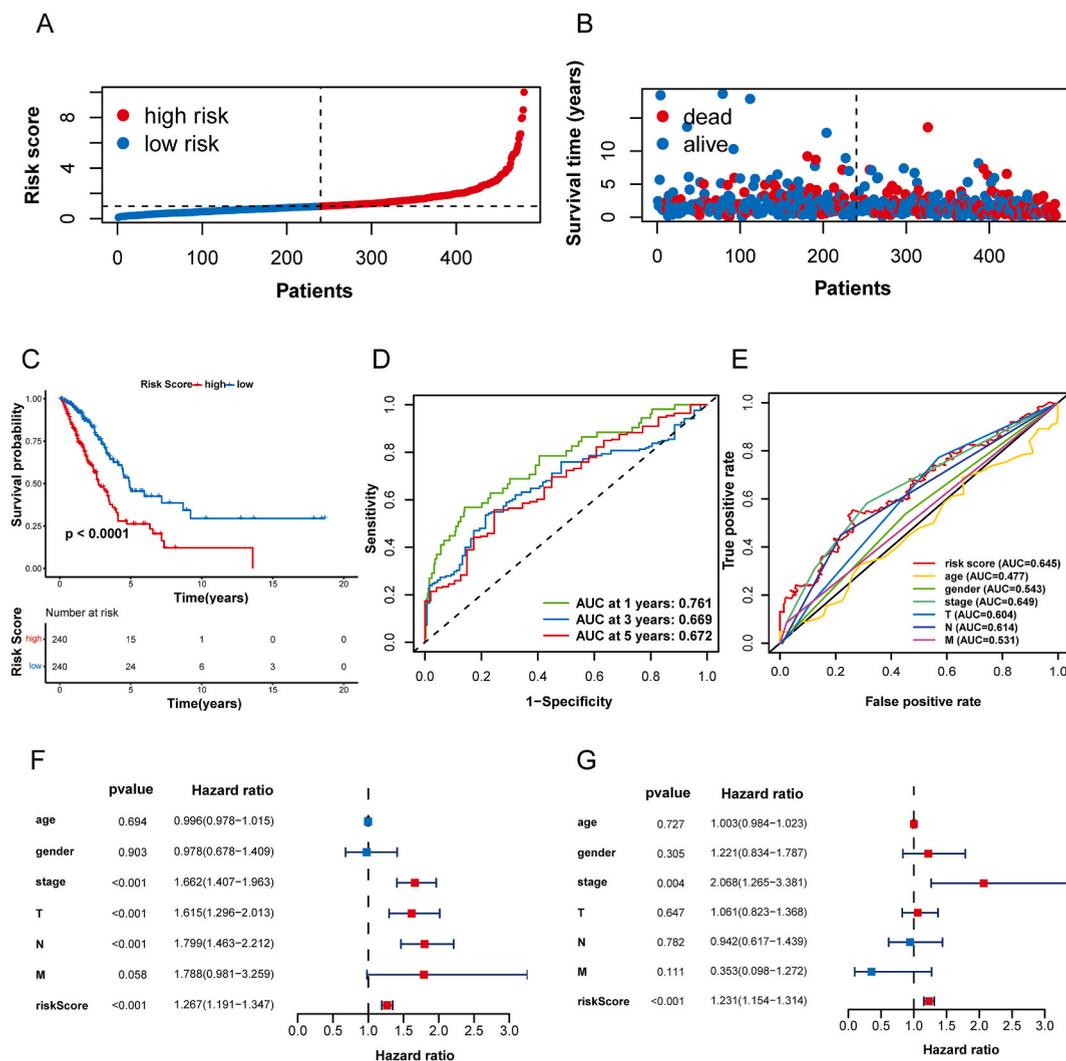


Fig. 6. Construction of prognostic model based on TCGA data. (A) Distribution of the risk score in the training cohort. (B) A scatterplot according to the OS per patient and risk score in the training cohort. (C) Survival curves of patients with LUAD in the training cohort. (D) The ROC curves for the training cohort at 1-, 3- and 5-year. (E) The ROC curve of the prognostic signature and clinicopathological factors in the training cohort. (F, G) The univariate and multivariate Cox regression analyses of prognostic signature and clinicopathological factors in the training cohort.

(Fig. 8A). The nomogram which involved the stage, T, N and the risk score, was further constructed to separately predict LUAD survival for 1-year, 3-year and 5-year (Fig. 8B). To assess the predictive efficacy of this signature, calibration curves were conducted and displayed fairly consistent predictive survival rates and actual survival rates (Fig. 8C). The time-dependent C-index showed that compared to the risk score and the clinicopathological information, the predictive result of the nomogram model was the most consistent with the reality (Fig. 8D). DCA curves revealed that the nomogram provided the optimal net clinical benefit (Fig. 8E). It suggested that the nomogram may be a useful method for predicting patient prognosis in clinical practice. GSEA further conducted an enrichment analysis of the prognostic signature to investigate the potential functions between the two risk groups (Fig. 8F). The results showed that some pathways were related to tumor-associated pathways, including p53, JAK/STAT and mTOR signaling pathways. Moreover, these results revealed the low-risk group had a close connection to angiogenesis and tumor-immune related pathways, such as the VEGF, T cell receptor, FC epsilon RI, and B cell receptor signaling pathways. In the aforementioned findings, it was verified that M2 TAM and angiogenesis-related genes were associated with tumor immune microenvironment (TIME) in our signature and further immune-related analyses were needed.

3.8. Analyses of TIME and immunotherapy

For the purpose of investigating the connection between clinicopathological factors, the risk score and TIME, heatmap results indicated that N, T and stage showed a statistical difference between the two risk groups, and there were more active immune-related

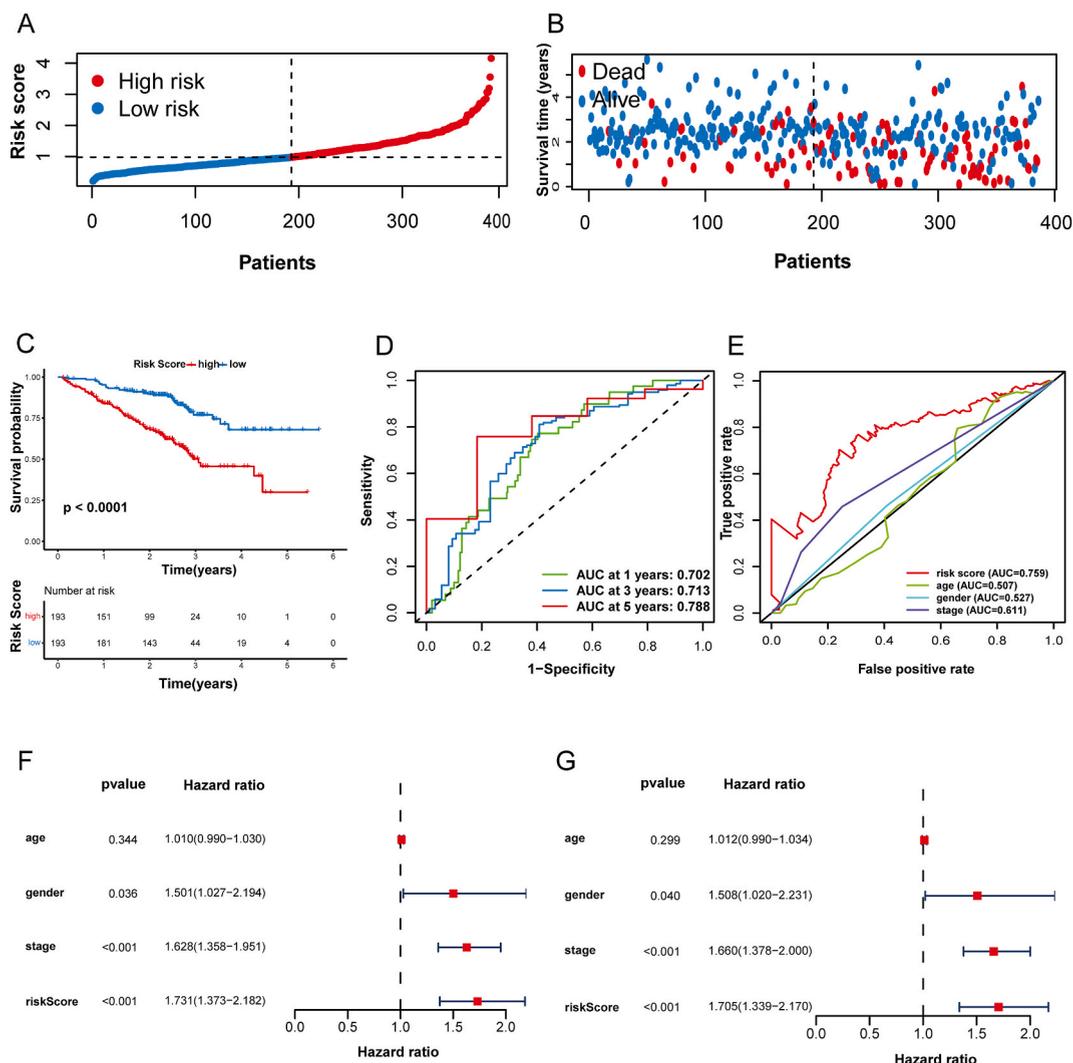


Fig. 7. External validation of prognostic model. (A) Distribution of risk score in the validation cohort. (B) A scatterplot according to the OS of each patient and risk score in the validation cohort. (C) Survival curves of patients with LUAD in the validation cohort. (D) The ROC curves for the test cohort at 1-year, 3-year and 5-year. (E) The ROC curve of the prognostic signature and clinicopathological factors in the validation cohort. (F, G) The univariate and multivariate Cox regression analyses of prognostic signature and clinicopathological factors in the validation cohort.

functions within the low-risk group (Fig. 9A). The three violin plots indicated that the StromalScore, ImmuneScore and ESTIMATE-Score were presented separately within the two risk groups. Noticeably, low-risk patients had significantly higher scores in all three types of scores (Fig. 9B–D). For a further investigation of the association between the prognostic signature and immune cells and functions, the ssGSEA enrichment scores were evaluated for several immune cell subgroups and the immune-related pathways. The results found that most immune cell infiltration happened in the low-risk group, including the neutrophils, Tregs, tumor-infiltrating lymphocyte (TIL), macrophages, dendritic cells (IDCs), T helper cells, mast cells, dendritic cells (DCs), immature B cells, plasmacytoid dendritic cells (pDCs) and activated dendritic cells (aDCs) (Fig. 9E). There were 6 immune-related pathways significantly linked to the signature, including chemokine receptor (CCR), T cell co-inhibition, check-point, T cell co-stimulation, human leukocyte antigen (HLA) and type II Interferon (IFN) response (Fig. 9F). Furthermore, it was discovered that the genes linked to the immune checkpoint were highly expressed in most patients with low risk, particularly, such as TNFSF18, ICOS, BTNL2, HAVCR2, CD80, BTLA, PDCD1 and CTLA4, etc. (Fig. 9G). These results provided an opportunity to identify new targets for LUAD immunotherapy.

3.9. The relationship between somatic mutation and the prognostic signature

Additional file 2: Fig. S2 presented the overall mutation profile of TCGA-LUAD. We further investigated the genetic mutations and found that the top-5 mutant genes, such as TP53, MUC16, TTN, RYR2 and CSMD3, were the most frequently mutated genes in the high-risk group, while the top-5 mutant genes were TP53, CSMD3, MUC16, TTN and LRP1B in the low-risk group (Additional file 2:

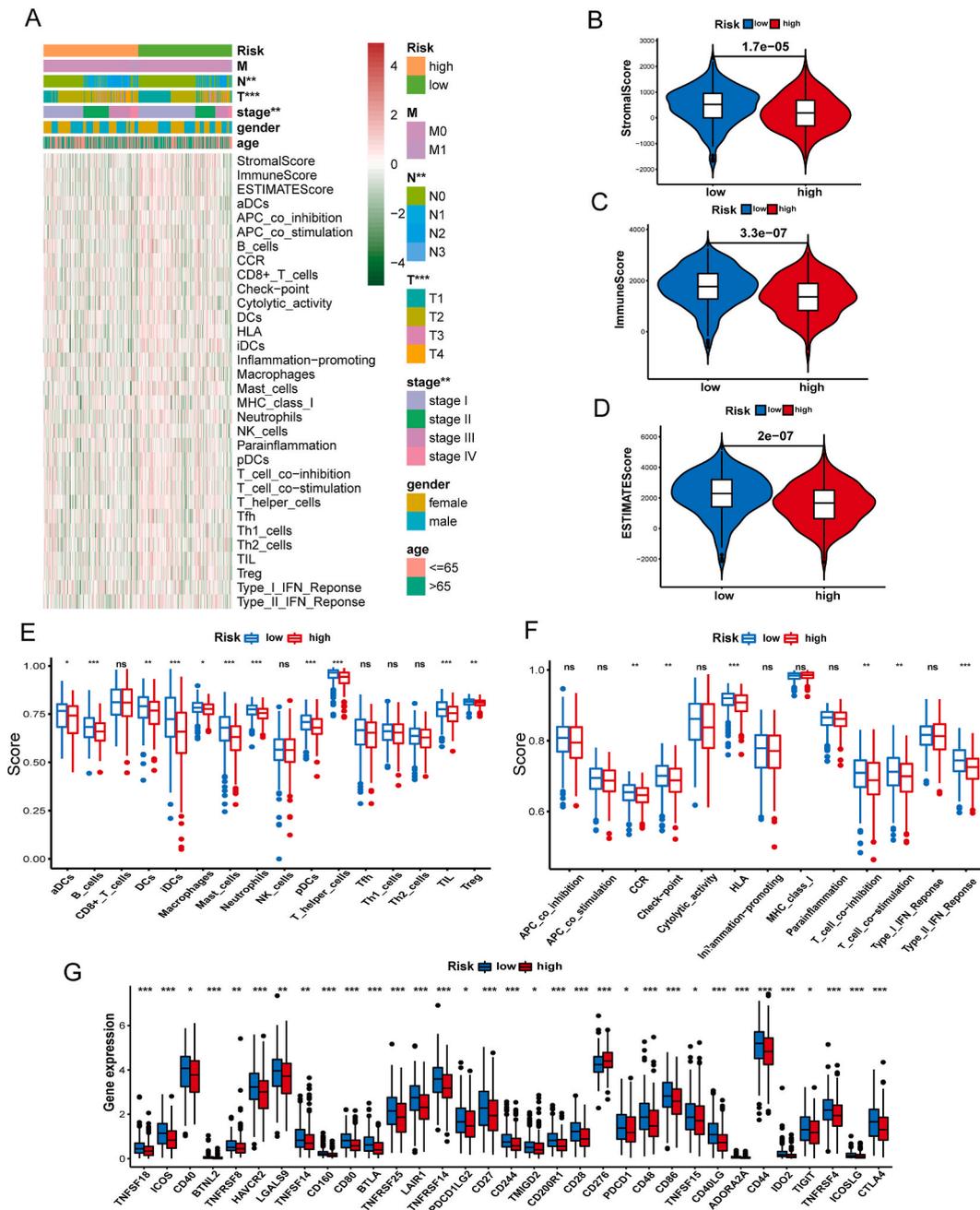


Fig. 9. Analyses of TIME and immunotherapy. (A) Distribution heatmap of the risk score, clinicopathological factors, immune cells and immune-related functions. (B–D) Violin plots of StromalScore, ImmuneScore, and ESTIMATEScore between the two risk groups. (E, F) Enrichment scores of immune cells and immune-related pathways between the two risk groups. (G) The expression of immune checkpoint between two risk groups. ns, not significant. *p < 0.05, **p < 0.01, ***p < 0.001.

cisplatin, etoposide, docetaxel, gemcitabine, mitomycin.C and paclitaxel (Fig. 10A–F). Regarding the IC50 of targeted drugs, such as Axitinib, BIBW2992 and PF.02341066, significant differences were observed between the two risk groups (Fig. 10G–I). Therefore, both M2 TAM and angiogenesis-related prognostic signature could be applied to predict the efficacy of LUAD therapy and providing clinical treatment guidance.

3.11. Measurement of signature-genes expression in tissues

When the prognostic signature of both M2 TAM and angiogenesis-related genes were established, the expression of signature-genes

was further explored in the TCGA-LUAD samples. As observed in Fig. 11A, the mRNA expression levels of OLR1, CTSL, HLA-DPB1, NUPR1, ALOX5, DOCK4, CSF2RB, PTPN6 and TNFSF12 were significantly downregulated in tumor samples, whereas HNRNPA2B1 and NCL showed an obvious upregulation in tumor samples and BIRC2 was not significantly different between adjacent normal tissues and tumor tissues. Furthermore, in 10 samples obtained from LUAD patients, it was found that mRNA expression levels of 4 genes (OLR1, NUPR1, PTPN6 and NCL) were not significantly different or contrary to the gene expression trends of TCGA-LUAD in adjacent normal tissues and tumor tissues (Additional file 2: Fig. S4). However, it is noted that the mRNA expression levels of 7 genes, including CTSL, HLA-DPB1, ALOX5, DOCK4, CSF2RB, TNFSF12 and HNRNPA2B1, were significantly different and in accordance with the gene expression trends of TCGA-LUAD (Fig. 11B–H).

4. Discussion

Immunotherapy has gained popularity as a potent therapeutic approach for cancer treatment. Recently, the effectiveness of ICIs has garnered interest in immunotherapy for lung cancer [37]. However, identifying patients with LUAD who would benefit from immunotherapy remains a significant challenge. Recent studies have shown that the scRNA-seq technology contributes significantly to the analysis of the tumor heterogeneity and different cell subpopulations, which is necessary to identify prospective treatment targets [38]. The research analyzed the single-cell RNA and bulk RNA sequencing to explore 12 both M2 TAM and angiogenesis-related genes (OLR1, CTSL, HLA-DPB1, NUPR1, ALOX5, DOCK4, CSF2RB, PTPN6, TNFSF12, HNRNPA2B1, NCL and BIRC2) in LUAD and constructed a prognostic signature. Additionally, a nomogram was subsequently established for clinically predicting prognosis. In various validation methods, including calibration plots, the time-dependent C-index and DCA, proved that the nomogram had a higher

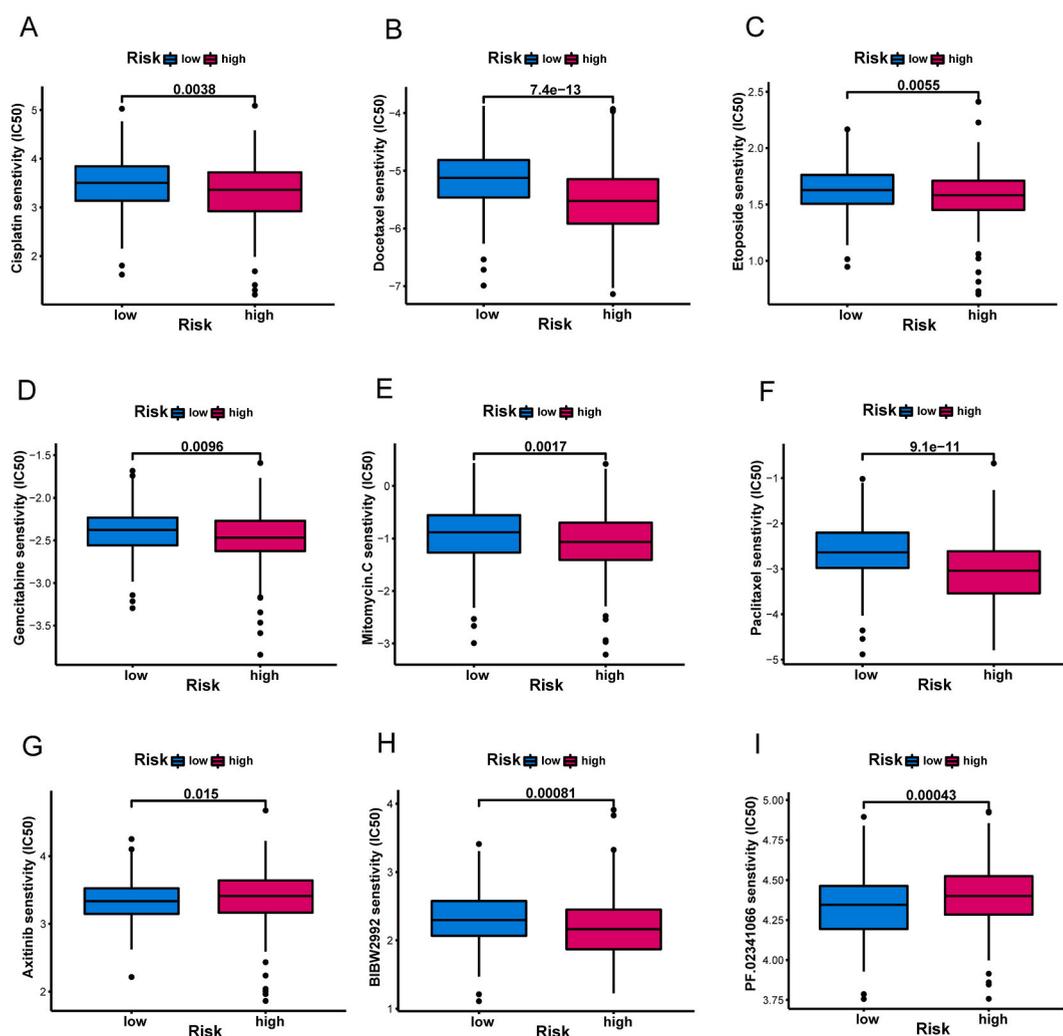


Fig. 10. The sensitivity analysis of common chemotherapeutic and targeted drugs. (A–F) The relationship between the prognostic signature and chemotherapy drugs, including cisplatin, docetaxel, mitomycin.C, etoposide, gemcitabine and paclitaxel. (G–I) The relationship between the prognostic signature and targeted drugs, including Axitinib, BIBW2992 and PF.02341066.

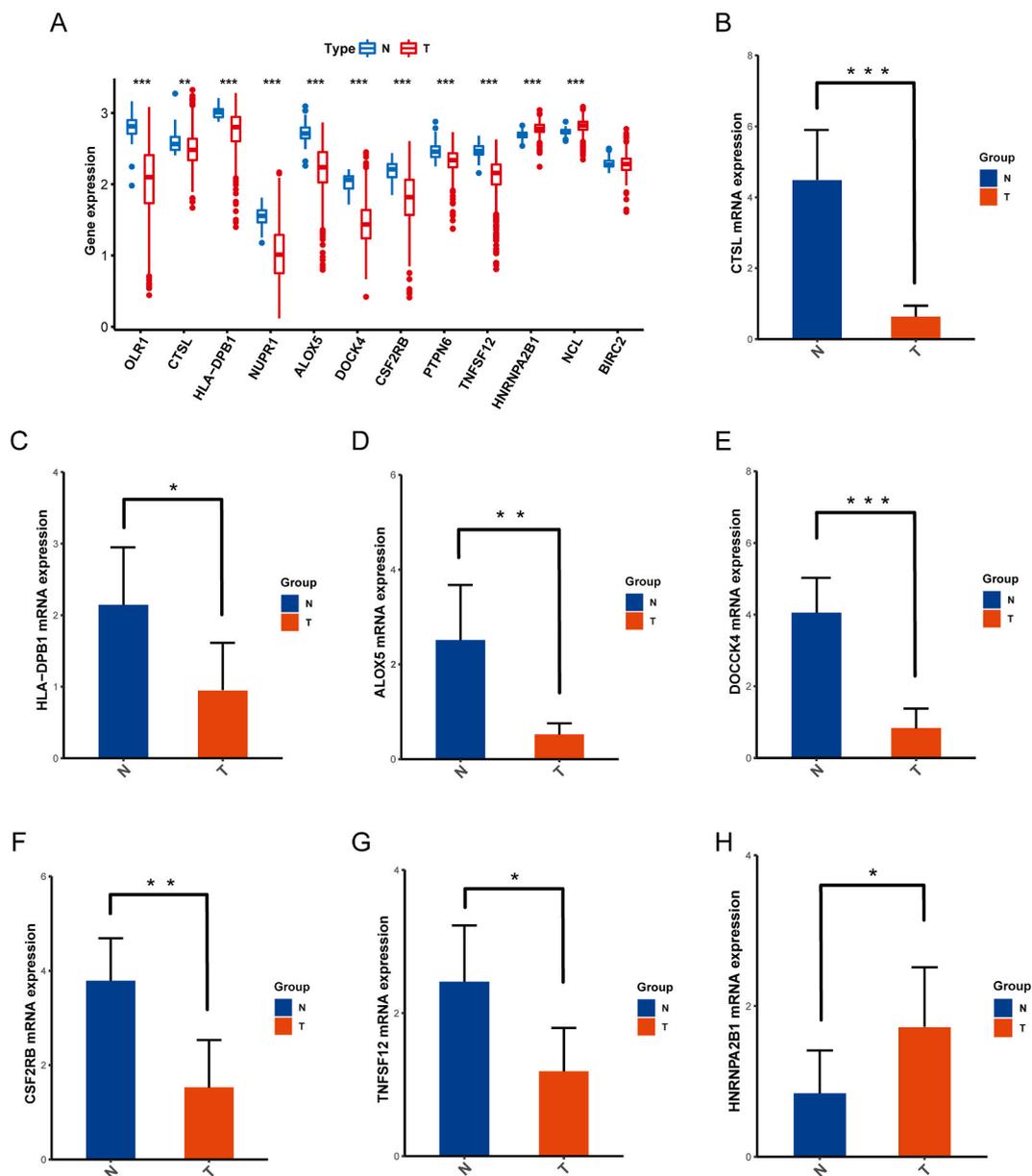


Fig. 11. Measurement of the signature-gene expression in tissues. (A) Expression of the signature genes in TCGA-LUAD samples. (B–H) mRNA expression of CTSL (B), HLA-DPB1 (C), ALOX5 (D), DOCK4 (E), CSF2RB (F), TNFSF12 (G) and HNRNPA2B1 (H) in 10 pairs of collected samples. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

predictive accuracy. Therefore, this nomogram can better establish the individual prognostic status of patients with LUAD and guide personalized treatment. Furthermore, PCR results indicated that the mRNA expression levels of CTSL, HLA-DPB1, ALOX5, DOCK4, CSF2RB, TNFSF12, and HNRNPA2B1 were significantly different between normal and tumor tissues, which may indicate that the 7 genes contribute more to the prognostic model of LUAD.

In addition, our study depicts the potential biological functional characteristics of the different risk groups. GSEA enrichment analysis showed that the high and low risk groups were associated with the mTOR signaling pathway, p53 signaling pathway and JAK/STAT signaling pathway, etc. Among them, p53 is one of the most thoroughly studied tumor suppressors, and it was found previously that patients with P53 mutations have a higher likelihood of undergoing immune escape and exhibiting a poor prognosis [39]. The therapy which targets the p53 pathway may be significantly influenced by the TME in solid tumors [40]. The mTOR signaling pathway is a popular target in anti-tumor therapy research. Recently, novel findings regarding mTOR inhibitors have progressed into clinical studies, and various drugs have been found to have high activity in association with mTOR inhibitors [41]. Following an in-depth study of the JAK/STAT signaling pathway in recent years, its continuous activation has been closely connected to not only the development

and metastasis of lung cancer, but also the occurrence of drug resistance [42]. Notably, the low-risk group was mainly related to angiogenesis and tumor immune pathways. Among them, the VEGF signaling pathway contributes especially to the tumor angiogenesis process. VEGF and its cognate receptors are generally considered to be the most critical regulators of angiogenesis. The stimulation of the VEGF signaling pathway leads to neo-tumor angiogenesis and branch cell pathway formation, which can promote rapid tumor growth and metastatic potential [43]. The recognized role of VEGF in promoting tumor angiogenesis and human cancer pathogenesis has led to the rational design and development of drugs that selectively target this pathway. Moreover, the T cell receptor signaling pathway and B cell receptor signaling pathway presented obvious enrichment within the low-risk group, while greater enrichment of the immune-related pathways suggested a better immune response status.

Since TIME contributes significantly to the anti-tumor response and influences prognosis [44], the connection between the signature and TIME was explored. First of all, immune scores were obviously higher in the low-risk group than in the high-risk group. Next, the immune cell infiltration level presented a higher proportion of Tregs, B cells, TIL, T helper cells, pDCs, DCs, iDCs, neutrophils, and aDCs in the low-risk group, indicating that these patients were likely to stay in a relatively active state of anti-tumor immune response. It has been reported that TIL, T helper cells, B cells and neutrophils eliminate tumor cells in the antitumor immune environment [45]. During inflammation in normal tissues, aDCs induce protective CD8⁺ T-cell responses [46], whereas in the presence of mature DCs (activated or inactivated), an immune response against lung cancer is required to organize cytotoxic T-cells, which are associated with good therapeutic efficacy and clinical prognostic outcomes [47]. Consistently, in our study, patients in the low-risk group exhibited a higher number of infiltrating cells and tended to have better OS.

Previous research has suggested that ICIs are potential therapeutic target for lung cancer [48]. ICIs target regulatory pathways in the T cells to enhance anti-tumor immune responses rather than producing direct cytotoxic effects on tumor cells. Our findings presented that the common immune checkpoint-related genes (TNFSF18, ICOS, BTNL2, HAVCR2, CD80, BTLA, PDCD1 and CTLA4) were highly expressed in the low-risk group, and PDCD1 and CTLA4 have been confirmed as important immunotherapeutic targets [49]. CTLA-4 and PD-1 negatively regulate T cell activity at different stages of the immune response [50]. Higher expression of immune checkpoint molecules may be associated with increased sensitivity to ICIs treatment, indicating that these tumors are in a state of pre-activated immunity [51]. Considering the individual effectiveness of ICIs and the limitations caused by hyperprogressive disease [52], our signature is particularly important as a validated biomarker for predicting ICIs efficacy in patients with LUAD.

Several studies have shown the predictive value of TMB for the response to immunotherapy and patient prognosis, therefore, we also investigated the differences in TMB between two risk groups. Our study showed that higher TMB occurred in the high-risk group, suggesting that this group may be associated with a poorer prognosis [53]. Huang et al. found that even rejuvenation of CD8⁺ T cells by anti-PD-1 immunotherapy may be clinically ineffective if the TMB is high [54]. TMB could predict the response to immunotherapy, but some lower clinical benefits were also observed in cancers with high TMB. It further explains why patients in the high-risk group were associated with a poorer prognosis [49].

To guide LUAD treatment, pharmacovigilance analyses have been performed in various risk groups. Nine anti-cancer drugs were compared between the low- and high-risk groups, including cisplatin, etoposide, gemcitabine, docetaxel, mitomycin.C, paclitaxel, Axitinib, BIBW2992 and PF.02341066. These findings presented that the high-risk group was sensitive to the six chemotherapeutic agents and could be referred to for the clinical selection of chemotherapeutic agents. In addition, the sensitivity of targeted drugs differed significantly between the two risk groups. Particularly, axitinib as a potent and selective inhibitor of VEGFR-1, -2, and -3 (classic angiogenic pathways), is seen to be one of the newest and most effective anti-angiogenic tyrosine kinase inhibitors that have now been assessed for the treatment of non-small cell lung cancer with controlled toxicity. It is not surprising that our signature is related to axitinib and other conventional chemotherapeutic drugs because it was constructed based on M2 TAM and angiogenesis-related genes. Specific chemotherapeutic drugs induce the release of tumor antigens and cofactors to engage macrophages in an effective cancer immune cycle, a process known as immunogenic cell death [55]. Specific anti-cancer drugs can also reverse the polarization of TAM, thereby increasing their response to treatment [56]. These findings can guide the subsequent clinical treatment of patients with LUAD and provide innovative insights into the development of new drugs.

However, some limitations are still inevitable in our study. Firstly, the number of scRNA-seq samples, data published in public databases, and samples collected from patients are limited and not comprehensive, possibly resulting in some potential bias. Secondly, the molecular mechanisms of signature genes require further experimental verification. Thirdly, the drug sensitivity need to be further confirmed using cellular assays. Therefore, it is necessary to further validate our results with multicenter, prospective, large-sample double-blind trials.

5. Conclusions

In summary, a novel prognostic signature composed of 12 M2 TAM and angiogenesis-related genes was formulated and validated through the integrated analysis of the single-cell RNA and bulk RNA sequencing. It could be a valid prognostic biomarker, potentially predicts patient response to immunotherapy, the sensitivity to chemotherapy and targeted drugs in patients with LUAD. However, further detailed experiments are required to confirm the validity of this prognostic signature.

Ethics and consent statement

This study was approved by the Human Subjects Committee of the Seventh Affiliated Hospital of Sun Yat-sen University (Ethical Number: KY-2023-051-02). All patients provided their informed consent to participate in the study and for their data to be published.

Funding

This work received support from the following funding source: the Medical and Health Research Project of Bao'an District Medical Association (Grant No: BAYXH2023050).

Data availability statement

The datasets from 4 database (CellMarker, TCGA, Genecards, GEO) are publicly available.

CRedit authorship contribution statement

Anbang Liu: Writing – original draft, Visualization, Validation, Software, Resources, Data curation, Conceptualization. **Gengqiu Liu:** Writing – original draft, Resources, Methodology, Investigation, Formal analysis. **Xiaohuai Wang:** Visualization, Methodology, Data curation. **Dongqing Yan:** Visualization, Software, Investigation. **Junhang Zhang:** Writing – review & editing, Supervision, Conceptualization. **Li Wei:** Supervision, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to 10 patients who participated in this medical research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e34784>.

References

- [1] H. Sung, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, *CA Cancer J Clin* 71 (3) (2021) 209–249.
- [2] F.R. Hirsch, et al., Lung cancer: current therapies and new targeted treatments, *Lancet* 389 (10066) (2017) 299–311.
- [3] J.J. Havel, D. Chowell, T.A. Chan, The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy, *Nat. Rev. Cancer* 19 (3) (2019) 133–150.
- [4] R.D. Paucek, D. Baltimore, G. Li, The cellular immunotherapy revolution: arming the immune system for precision therapy, *Trends Immunol.* 40 (4) (2019) 292–309.
- [5] M. Binnewies, et al., Understanding the tumor immune microenvironment (TIME) for effective therapy, *Nat Med* 24 (5) (2018) 541–550.
- [6] A. Mantovani, et al., Tumour-associated macrophages as treatment targets in oncology, *Nat. Rev. Clin. Oncol.* 14 (7) (2017) 399–416.
- [7] J.P. Väyrynen, et al., The prognostic role of macrophage polarization in the colorectal cancer microenvironment, *Cancer Immunol. Res.* 9 (1) (2021) 8–19.
- [8] T. Lawrence, G. Natoli, Transcriptional regulation of macrophage polarization: enabling diversity with identity, *Nat. Rev. Immunol.* 11 (11) (2011) 750–761.
- [9] A. Mantovani, et al., Cancer-related inflammation, *Nature* 454 (7203) (2008) 436–444.
- [10] A.J. Petty, Y. Yang, Tumor-associated macrophages: implications in cancer immunotherapy, *Immunotherapy* 9 (3) (2017) 289–302.
- [11] C. Viallard, B. Larrivée, Tumor angiogenesis and vascular normalization: alternative therapeutic targets, *Angiogenesis* 20 (4) (2017) 409–426.
- [12] S.A. Stacker, et al., Lymphangiogenesis and lymphatic vessel remodelling in cancer, *Nat. Rev. Cancer* 14 (3) (2014) 159–172.
- [13] L. Cassetta, J.W. Pollard, Targeting macrophages: therapeutic approaches in cancer, *Nat. Rev. Drug Discov.* 17 (12) (2018) 887–904.
- [14] K. Kessenbrock, V. Plaks, Z. Werb, Matrix metalloproteinases: regulators of the tumor microenvironment, *Cell* 141 (1) (2010) 52–67.
- [15] C. Bailey, et al., Chemokine expression is associated with the accumulation of tumour associated macrophages (TAMs) and progression in human colorectal cancer, *Clin. Exp. Metastasis* 24 (2) (2007) 121–130.
- [16] K.Y. Ma, et al., Single-cell RNA sequencing of lung adenocarcinoma reveals heterogeneity of immune response-related genes, *JCI Insight* 4 (4) (2019).
- [17] A. Maynard, et al., Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing, *Cell* 182 (5) (2020) 1232–1251.e22.
- [18] R. Zhong, et al., Immune cell infiltration features and related marker genes in lung cancer based on single-cell RNA-seq, *Clin. Transl. Oncol.* 23 (2) (2021) 405–417.
- [19] T. Wu, et al., Immune contexture defined by single cell technology for prognosis prediction and immunotherapy guidance in cancer, *Cancer Commun.* 39 (1) (2019) 21.
- [20] P. Bischoff, et al., Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma, *Oncogene* 40 (50) (2021) 6748–6758.
- [21] Y. Hao, et al., Integrated analysis of multimodal single-cell data, *Cell* 184 (13) (2021) 3573–3587.e29.
- [22] C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression, *Genome Biol.* 20 (1) (2019) 296.
- [23] A.C. Habermann, et al., Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis, *Sci. Adv.* 6 (28) (2020) eaba1972.
- [24] P.R. Tata, J. Rajagopal, Plasticity in the lung: making and breaking cell identity, *Development* 144 (5) (2017) 755–766.
- [25] S. Jin, et al., Inference and analysis of cell-cell communication using CellChat, *Nat. Commun.* 12 (1) (2021) 1088.
- [26] A.M. Newman, et al., Robust enumeration of cell subsets from tissue expression profiles, *Nat. Methods* 12 (5) (2015) 453–457.
- [27] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, *BMC Bioinf.* 9 (2008) 559.
- [28] G. Yu, et al., clusterProfiler: an R package for comparing biological themes among gene clusters, *OMICS* 16 (5) (2012) 284–287.

- [29] R. Tibshirani, The lasso method for variable selection in the Cox model, *Stat. Med.* 16 (4) (1997) 385–395.
- [30] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, *J Stat Softw* 33 (1) (2010) 1–22.
- [31] P.J. Heagerty, T. Lumley, M.S. Pepe, Time-dependent ROC curves for censored survival data and a diagnostic marker, *Biometrics* 56 (2) (2000) 337–344.
- [32] A. Subramanian, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, *Proc Natl Acad Sci U S A* 102 (43) (2005) 15545–15550.
- [33] K. Yoshihara, et al., Inferring tumour purity and stromal and immune cell admixture from expression data, *Nat. Commun.* 4 (2013) 2612.
- [34] S. Hänzelmann, R. Castelo, J. Guinney, GSEA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.* 14 (2013) 7.
- [35] A. Mayakonda, et al., Maftools: efficient and comprehensive analysis of somatic variants in cancer, *Genome Res.* 28 (11) (2018) 1747–1756.
- [36] P. Geeleher, N. Cox, R.S. Huang, pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels, *PLoS One* 9 (9) (2014) e107468.
- [37] A. Steven, S.A. Fisher, B.W. Robinson, Immunotherapy for lung cancer, *Respirology* 21 (5) (2016) 821–833.
- [38] E. Papalexi, R. Satija, Single-cell RNA sequencing to explore immune cell heterogeneity, *Nat. Rev. Immunol.* 18 (1) (2018) 35–45.
- [39] G. Blandino, S. Di Agostino, New therapeutic strategies to treat human cancers expressing mutant p53 proteins, *J. Exp. Clin. Cancer Res.* 37 (1) (2018) 30.
- [40] I. Amelio, G. Melino, Context is everything: extrinsic signalling and gain-of-function p53 mutants, *Cell Death Discov* 6 (2020) 16.
- [41] Z. Zou, et al., mTOR signaling pathway and mTOR inhibitors in cancer: progress and challenges, *Cell Biosci.* 10 (2020) 31.
- [42] X. Yang, et al., [Research advances of JAK/STAT signaling pathway in lung cancer], *Zhongguo Fei Ai Za Zhi* 22 (1) (2019) 45–51.
- [43] D.J. Hicklin, L.M. Ellis, Role of the vascular endothelial growth factor pathway in tumor growth and angiogenesis, *J. Clin. Oncol.* 23 (5) (2005) 1011–1027.
- [44] J.M. Pitt, et al., Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy, *Ann. Oncol.* 27 (8) (2016) 1482–1492.
- [45] J.M. Taube, et al., Implications of the tumor immune microenvironment for staging and therapeutics, *Mod. Pathol.* 31 (2) (2018) 214–234.
- [46] P.F. McKay, et al., Identification of potential biomarkers of vaccine inflammation in mice, *Elife* 8 (2019).
- [47] J. Goc, et al., Dendritic cells in tumor-associated tertiary lymphoid structures signal a Th1 cytotoxic immune contexture and license the positive prognostic value of infiltrating CD8+ T cells, *Cancer Res.* 74 (3) (2014) 705–715.
- [48] J. Yang, et al., Immune checkpoint blockade as a potential therapeutic target in non-small cell lung cancer, *Expert Opin Biol Ther* 16 (10) (2016) 1209–1223.
- [49] B.D. Qin, X.D. Jiao, Y.S. Zang, Tumor mutation burden to tumor burden ratio and prediction of clinical benefit of anti-PD-1/PD-L1 immunotherapy, *Med. Hypotheses* 116 (2018) 111–113.
- [50] E.I. Buchbinder, A. Desai, CTLA-4 and PD-1 pathways: similarities, differences, and implications of their inhibition, *Am. J. Clin. Oncol.* 39 (1) (2016) 98–106.
- [51] Z. Zeng, et al., Identification and application of a novel immune-related lncRNA signature on the prognosis and immunotherapy for lung adenocarcinoma, *Diagnostics* 12 (11) (2022).
- [52] R. Ferrara, et al., Hyperprogressive disease in patients with advanced non-small cell lung cancer treated with PD-1/PD-L1 inhibitors or with single-agent chemotherapy, *JAMA Oncol.* 4 (11) (2018) 1543–1552.
- [53] M. Yarchoan, A. Hopkins, E.M. Jaffee, Tumor mutational burden and response rate to PD-1 inhibition, *N. Engl. J. Med.* 377 (25) (2017) 2500–2501.
- [54] Z.R. Chalmers, et al., Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden, *Genome Med.* 9 (1) (2017) 34.
- [55] D.G. DeNardo, B. Ruffell, Macrophages as regulators of tumour immunity and immunotherapy, *Nat. Rev. Immunol.* 19 (6) (2019) 369–382.
- [56] A. Mantovani, et al., Macrophages as tools and targets in cancer therapy, *Nat. Rev. Drug Discov.* 21 (11) (2022) 799–820.