A prognostic nomogram for lung adenocarcinoma based on immune-infiltrating Treg-related genes: from bench to bedside

Xiaofei Wang#, Zengtuan Xiao#, Jialin Gong, Zuo Liu, Mengzhe Zhang, Zhenfa Zhang

Department of Lung Cancer Surgery, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China

Contributions: (I) Conception and design: Z Zhang, X Wang; (II) Administrative support: Z Zhang; (III) Provision of study materials or patients: X Wang; (IV) Collection and assembly of data: Z Xiao; (V) Data analysis and interpretation: X Wang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Zhenfa Zhang, MD. Department of Lung Cancer Surgery, Tianjin Medical University Cancer Institute and Hospital, Huanhu West Rd, Tianjin, China. Email: zhangzhenfa@tmu.edu.cn.

Background: Accumulating evidence suggests that lymphocyte infiltration in the tumor microenvironment is positively correlated with tumorigenesis and development, while the role of Tregs (regulatory T cells) has been controversial. Therefore, we attempted to discover the possible value of Tregs for lung adenocarcinoma (LUAD).

Methods: The gene-sequencing data of LUAD were applied from three Gene Expression Omnibus (GEO) datasets—GSE10072, GSE32863 and GSE43458; the corresponding fractions of tumor-infiltrating immune cells were extracted from the CIBERSORTx portal. Weighted gene coexpression network analysis (WGCNA) and protein-protein interaction (PPI) network analysis were conducted to identify the significant module and candidate genes related to Tregs. The role of candidate genes in LUAD was further verified using data from The Cancer Genome Atlas (TCGA) database. Finally, we constructed a nomogram model to predict the prognosis of LUAD by plotting Kaplan-Meier (K-M), receiver operating characteristic (ROC) and calibration curves, which elucidated the performance of the nomogram.

Results: In total, 10,047 genes in 333 samples (196 tumor and 137 normal samples) from the GEO database were included. By WGCNA and PPI analysis, we identified a significant black module and 36 candidate genes related to Treg. Next, the candidate genes were verified using TCGA data by Cox regression analysis to screen 13 hub genes that stratified LUAD patients into low- or high-risk groups. Low-risk patients showed a significantly longer overall survival (OS) than high-risk patients (3-year OS: 70.2% *vs.* 35.2%; 5-year OS: 36.6% *vs.* 0; P=1.651E-09), and the areas under the ROC curves (AUCs) showed good (3-year AUC: 0.733; 5-year AUC: 0.777). Next, we constructed a survival nomogram combining the hub genes and clinical parameters; the low-risk patients still showed a favorable prognosis compared with that of the high-risk patients (P=7.073E-13), and the AUCs were better (3-year AUC: 0.763; 5-year AUC: 0.873).

Conclusions: We revealed the role of immune-infiltrating Treg-related genes in LUAD and constructed a prognostic nomogram, which may help clinicians make optimal therapeutic decisions and help patients obtain better outcomes.

Keywords: Tregs (regulatory T cells); lung adenocarcinoma (LUAD); Gene Expression Omnibus (GEO); TCGA; WGCNA

Introduction

Lung cancer is the leading cause of malignancy-related death worldwide, with approximately 2.2 million new cases and 1.9 million deaths worldwide in 2017 (1), despite great progress in diagnosis and therapeutics over the years. As the most prevailing histological type of lung cancer, lung adenocarcinoma (LUAD) always derives from the acinar cells of the lung periphery and has a poor prognosis. This cause may be mainly due to the high heterogeneity of LUAD and advanced stage of patients when diagnosed (2). However, the current TNM (tumor size, lymph nodes and distant metastasis) staging system tends to be insufficient to accurately characterize and stage tumors at an early stage and even after surgery (3,4), features that are necessary to design an optimal initial treatment plan and offer adequate postoperative adjuvant therapy. Therefore, more complementary methods to enhance TNM staging and identify the behavior of LUAD are needed.

In recent years, our view of cancer has changed drastically. Tumors are no longer regarded as simple malignant masses or cells but as a complex tumor microenvironment: tumor cells recruit other infiltrating immune cell subpopulations to constitute a self-sufficient biological unit (5). The composition of the tumor microenvironment varies in different patients and even in the same type of cancer, such as different fractions of B cells, NK cells, M1/M2 macrophages, granulocytes, mast cells, CD8+ T cells, CD4+ helper T cells, and regulatory T cells, which determine tumor characteristics and the patient prognosis (6-8). Previous studies showed that the expansion and accumulation of suppressive Tregs always caused the development, metastasis and recurrence of multiple malignancies including lung cancer (9-12). Some studies found that FoxP3+CD4+Tregs infiltrating correlated negatively with the survival of small cell lung cancer (SCLC) (13,14). Other researches on non-small cell lung cancer (NSCLC) revealed that whether in peripheral blood or intratumor, high level of Tregs was associated with high metastasis and low survival rates (15-17). However, a study published in 'Nature Immunology' from Ferreira et al., highlighted the role of type 1 Tregs in enhancing the immunity barrier in peripheral tissues, which challenged the classical view of Tregs in immunosuppression (18). Given that the role of Tregs (regulatory T cells) in tumors has been controversial, we attempted to explore the potential value of Tregs for LUAD.

In the present study, we integrated three LUAD sequencing datasets, GSE10072, GSE32863, and GSE43458, from the Gene Expression Omnibus (GEO) database and extracted the relevant fractions of 22 immune-infiltrating cells from the CIBERSORTx portal. Next, weighted gene coexpression network analysis (WGCNA) and protein-protein interaction (PPI) network analysis were conducted to identify the most significant module and candidate genes related to Tregs. The candidate genes were then further validated using data from The Cancer Genome Atlas (TCGA) database, and 13 hub genes were screened. The correlation between hub genes and Tregs was tested using Spearman's method. Finally, we constructed a nomogram model combining the hub genes and clinical parameters, which showed a better performance to predict the risk of LUAD. The flow diagram of this study is shown in *Figure 1*. We present the following article in accordance with the TRIPOD reporting checklist (available at http://dx.doi.org/10.21037/tlcr-20-822).

Methods

Data source and processing

The LUAD sequencing data were applied from three GEO (http://www.ncbi.nlm.nih.gov/geo/) datasets, GSE10072, GSE32863, and GSE43458. We used the sav and limma packages of R to perform batch calibration and data normalization. When a gene corresponds to multiple probes, the mean value is taken as the final expression value.

Estimation of immune infiltrating cells

Using the sequencing data, we estimated the fractions of 22 tumor-infiltrating immune cells using CIBERSORTx (https://cibersortx.stanford.edu/), an online tool that imputes gene expression profiles by a deconvolution algorithm and provides an estimated abundance of known cell types within a mixed cell population (19).

Construction of the coexpression network and module-trait relationships

The expression values of the 10,048 genes of the LUAD samples were used to construct a weight coexpression network employing the R package "WGCNA", a biological method used to integrate genes with coexpression into the same module. The correlations between the modules and sample traits are calculated to screen the models with a high
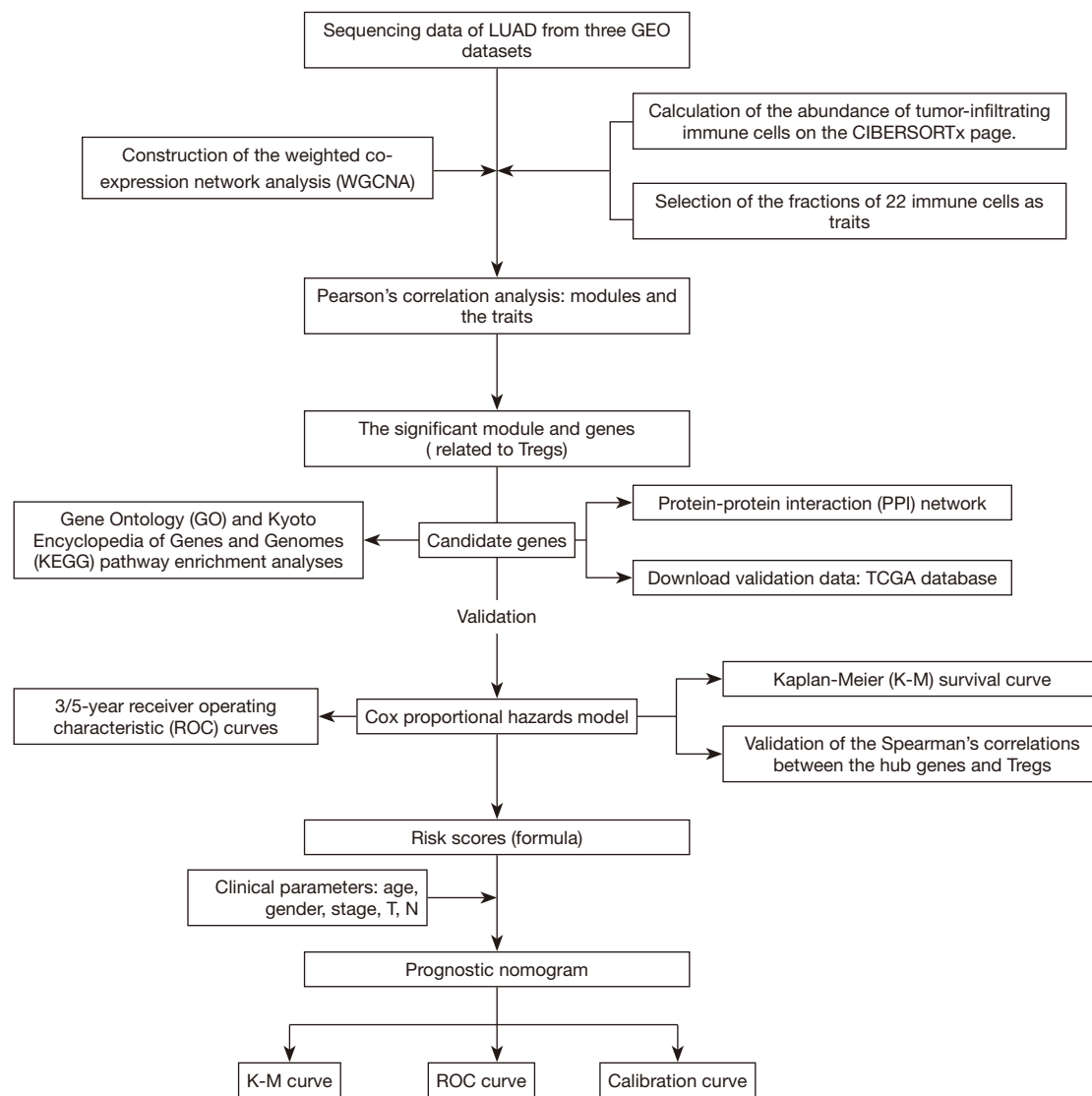
```
Sequencing data of LUAD from three GEO
datasets
                                    │
                                    │──────────> Calculation of the abundance of tumor-infiltrating
                                    │            immune cells on the CIBERSORTx page.
Construction of the weighted co-    │
expression network analysis (WGCNA) │──────────> Selection of the fractions of 22 immune cells as
                                    │            traits
                                    ↓
Pearson's correlation analysis: modules and
the traits
                                    ↓
The significant module and genes
( related to Tregs)
                                    │
Gene Ontology (GO) and Kyoto        │──────────> Protein-protein interaction (PPI) network
Encyclopedia of Genes and Genomes <─── Candidate genes
(KEGG) pathway enrichment analyses  │──────────> Download validation data: TCGA database
                                    │
                              Validation
                                    ↓
3/5-year receiver operating            │──────────> Kaplan-Meier (K-M) survival curve
characteristic (ROC) curves <───── Cox proportional hazards model
                                       │──────────> Validation of the Spearman's correlations
                                       │            between the hub genes and Tregs
                                    ↓
Risk scores (formula)
                                    │
Clinical parameters: age, ──────────>│
gender, stage, T, N                 │
                                    ↓
Prognostic nomogram
                                    │
              ┌─────────────────────┼─────────────────────┐
              ↓                      ↓                      ↓
         K-M curve              ROC curve            Calibration curve
```

Figure 1 Flow diagram of the study.

correlation with traits, and the genes in the modules are analyzed to identify target genes (20).

Here, we used the fractions of 22 immune-infiltrating cells as sample traits and chose an optimal soft threshold power (β) to build a scaleless network when setting the index of scale-free topologies as 0.90. Next, we assigned genes with similar expression patterns to the same module (minimum size =30) using the "dynamic tree cutting" algorithm. Moreover, we estimated the correlation of the module eigengenes with the infiltrating level of the 22 immune-infiltrating cells to screen the significance of the modules by Pearson's test. Finally, we selected the "Tregs (regulatory T cells)" subtype of interest

and the module with the highest correlation with Tregs was selected for further study.

Construction of the PPI network and identification of candidate genes

From the significant module, we obtained Treg-related module genes, with which the PPI network was developed using Search Tool for the Retrieval of Interacting Genes (STRING; https://string-db.org/). Next, the PPI network was presented using Cytoscape (version 3.7.2), which is a free app for visualizing sophisticated networks and integrating

them with attribute data. The "CytoHubba" module of Cytoscape is a plug-in that recognizes hub genes in a network based on the properties of nodes in a network (21), from which we screened the candidate genes related to Tregs.

Verification of candidate genes using TCGA data and screening of the hub genes

To further verify the role of candidate genes in LUAD, we applied gene sequencing and the corresponding completely clinical data (375 LUAD and 48 normal samples) at the TCGA (https://portal.gdc.cancer.gov/) portal, from which we extracted the expression values of the Treg-related genes. The expression values of LUAD and normal tissues were averaged using the "mean" function. Next, we used $\log_2$ transform to normalize all the average expression values. Analysis of the statistically significant differences between the LUAD and normal expression data was conducted using the Wilcoxon signed-rank test built into R (version 3.6.3; https://www.r-project.org/), defining the threshold of |log(fold change)| no less than 1 and the false discovery rate (FDR) less than 0.05. Finally, we implemented a Cox proportional hazards model to screen the Treg-related hub genes and their coefficients, from which the patients were assigned a high- or low-risk score. The risk scores were calculated using the following formula: *Risk score*$=\sum_1^n e_g * c_g$ (e_g is the expression value of gene *g* in a sample; *n* is the number of independent indicators, and c_g refers to the regression coefficient of gene *g* in the Cox proportional hazards model).

Validation of Treg-related hub genes

The Kaplan-Meier (K-M) curve was illustrated to estimate the differences in overall survival (OS) between the low- and high-risk groups using the log-rank test to analyze the statistical significance. Moreover, we implemented receiver operating characteristic (ROC) curves to evaluate the accuracy of grouping (low/high risk). Spearman's correlation between the Treg infiltration level and expression of hub genes was calculated using the data from Tumor Immune Estimation Resource (TIMER2.0; http://timer.cistrome. org/), and the results were visualized using the "ggstatsplot" package of R.

Enrichment analysis of hub genes related to Tregs

To identify tumor-related molecular mechanisms of the

hub genes related to Tregs, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed in R using the packages org.Hs.eg.db, Cluster Profiler, enrichplot, and ggplot2 with both P values and q-values less than 0.05.

Construction and validation of the prognostic nomogram model for LUAD

To apply the Treg-related genes better clinically, we constructed a prognostic nomogram model for LUAD, combining the risk score with traditional clinical parameters (age, gender, stage, T, N and M). A nomogram is an effective tool that formulates the scoring criteria for all the variables in the regression equation according to their regression coefficients. Next, each patient receives a summed score, which can be converted into the probability of the outcome time of each patient through the function (22). We then performed ROC, calibration and K-M curve analyses to elucidate the performance of the nomogram.

Statistical analysis

All statistical analyses and graphics were generated using the R and Perl packages. A Cox proportional hazards model was applied to identify survival genes related to Tregs. K-M curve analysis was performed to show the differences in OS between the low- and high-risk groups, using the log-rank test to estimate the significance of the differences. The calibration curve, ROC curve and area under the curve (AUC) values were used to determine the efficacy of the model. A P value less than 0.05 was defined as statistically significant.

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Results

Gene sequencing data and estimation of immune-infiltrating cells

We acquired the gene expression data of 10,048 genes from 196 LUAD and 137 normal tissues from the GEO database and calculated the abundance of 22 immune-infiltrating cells for each sample using the CIBERSORTx portal. Next, the fractions of 22 immune-infiltrating cells with gene expression data were selected as traits of WGCNA.
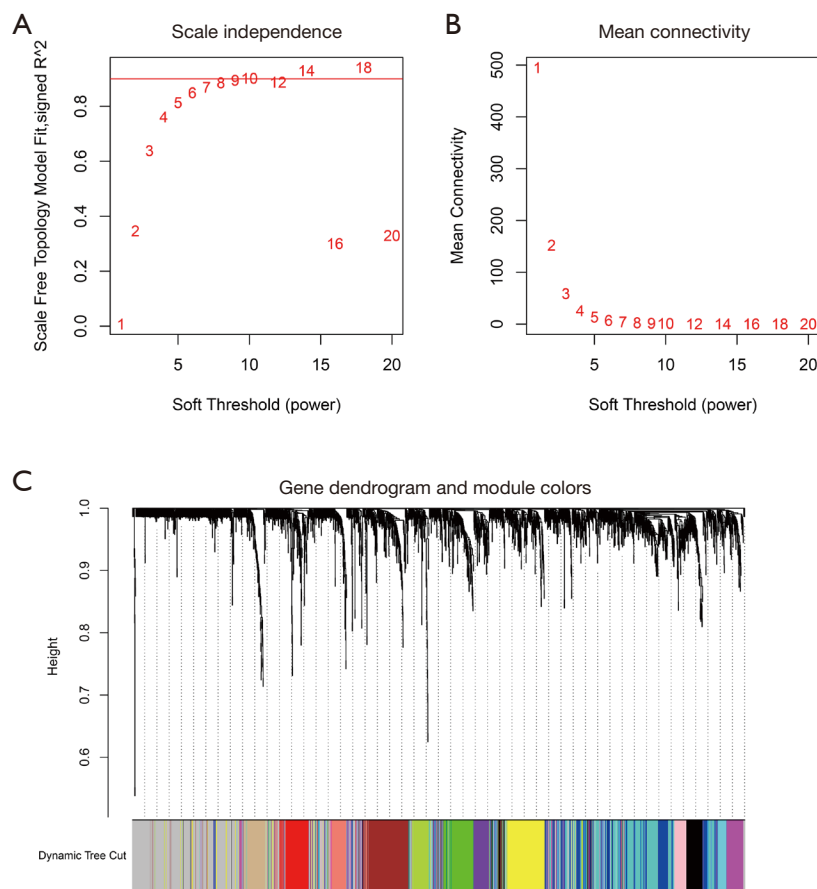
Figure 2 Selection of the appropriate soft threshold (power) and construction of the hierarchical clustering tree. (A) Selection of the soft threshold made the index of scale-free topologies reach 0.90. (B) Analysis of the average connectivity of 1–20 soft threshold power. (C) Treg-related genes with similar expression patterns were merged into the same module using a dynamic tree-cutting algorithm, creating a hierarchical clustering tree.

Coexpression network and protein-protein interaction network

The expression data of the 10,048 genes with the abundance of 22 immune-infiltrating cells were used to build the coexpression network. To build the scaleless network, we chose the appropriate soft threshold power (β=10) because it was the first power value to make the index of scale-free topologies reach 0.90 (*Figure 2A,B*). Treg-related genes with similar expression patterns were incorporated into the same module using a dynamic tree-cutting algorithm (module size =30), making a hierarchical clustering tree with 12 modules (*Figure 2C*). As shown in *Figure 3A*, the black module was highly correlated with regulatory T cells (Tregs) (R^2=0.52, P=1e-10) among the twelve modules. Because we were interested in Tregs, we selected 111 Treg-related genes (Table S1) in the black module with P<0.05 for further study.

Next, we developed the PPI network (*Figure 3B*) using 111 module genes and eventually screened 36 candidate genes related to Tregs using the "CytoHubba" module of Cytoscape with all scores no less than 10.

Verification of the role of Treg-related genes in LUAD

To further verify the role of 36 candidate genes in LUAD, we extracted the gene sequencing and corresponding clinical data from the TCGA datvabase, comprising 375 LUAD and 48 normal samples. Using the Cox proportional hazards model, we eventually screened 13 hub genes (*CCNB2, ECT2, RAD51AP1, UBE2C, CENPE, TOP2A, TYMS,*

172

Wang et al. A Treg-infiltrating nomogram for lung adenocarcinoma
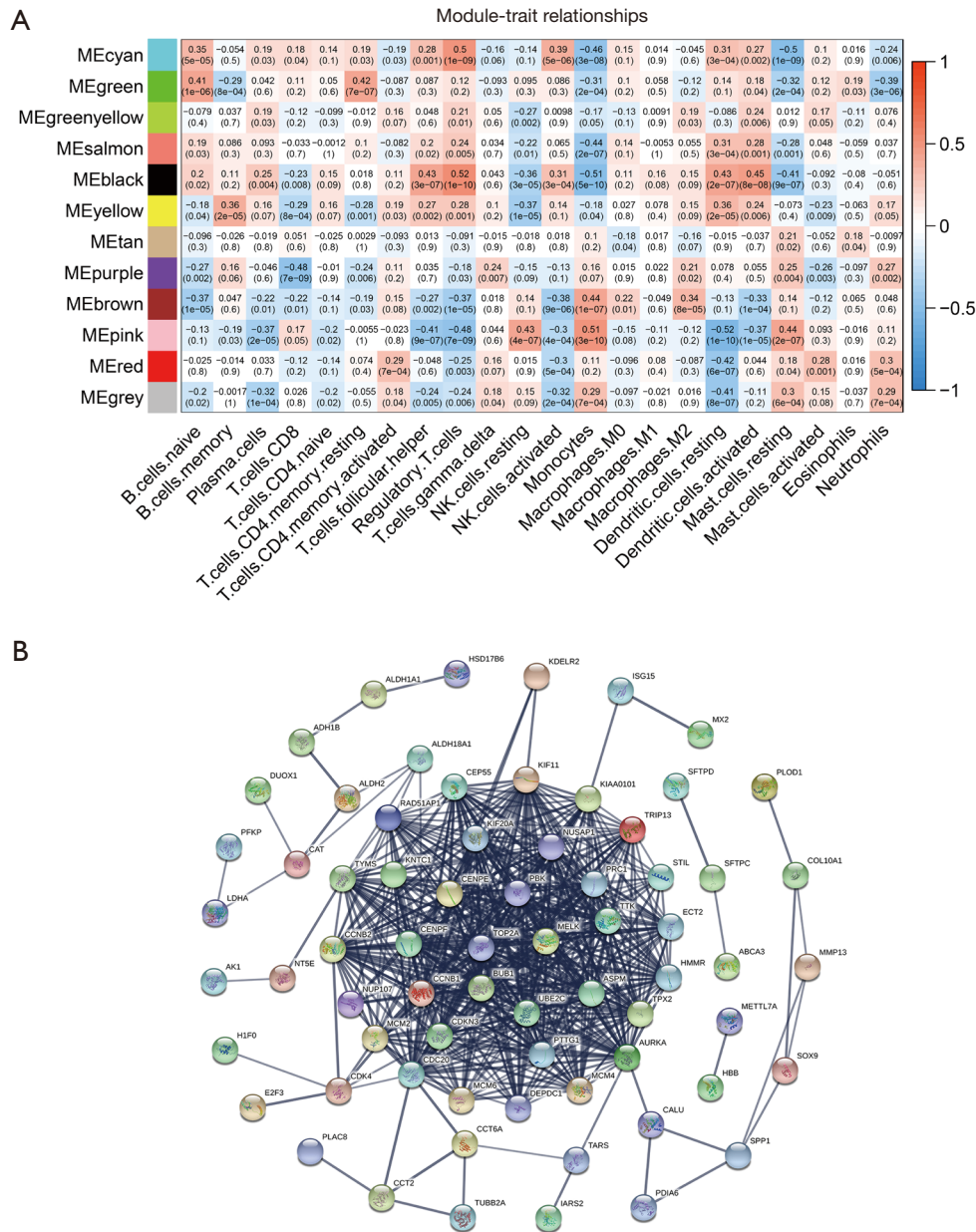
A

Module-trait relationships

B

Figure 3 Heatmap of the correlations between the modules and immune-infiltrating cells (traits). (A) Within every square, the number on the top refers to the coefficient between the cell infiltrating level and corresponding module, and the bottom is the P value. (B) The protein-protein interaction network of Treg-related genes.

KIF20A, *STIL*, *CDKN3*, *PRC1*, *AURKA*, *HMMR*) and their coefficients. Of the 13 hub genes, 4 (*STIL*, *CCNB2*, *RAD51AP1*, *TOP2A*) were considered risk genes (HRs: 0.2895–0.9717; all P<0.05), and their overexpression may lead to a worse prognosis. Additionally, 4 genes (*HMMR*, *ECT2*, *TYMS*, and *PRC1*) may serve as protective genes

(HRs: 1.0030–2.6878; all P<0.05), and their overexpression may lead to a better outcome for LUAD. The results of Cox regression analyses are shown in detail in *Table 1*. The individual risk score of each LUAD patient was calculated according to our *risk score* formula. Based on the median risk score, the LUAD patients were assigned to the low-

Table 1 Univariate and multivariate cox regression analyses of Treg-related genes in lung adenocarcinoma

Genes	Univariate analysis		Multivariate analysis		
	HR (95% CI)	P	HR (95% CI)	P	Coef
CCNB2	1.2865 (1.1052, 1.4975)	0.0011	0.5618 (0.3248, 0.9717)	0.0391	−0.5766
CENPF	1.3786 (1.1801, 1.6106)	0.0001			
ECT2	1.3876 (1.1780, 1.6344)	0.0001	1.4995 (1.0674, 2.1066)	0.0195	0.4051
TPX2	1.2902 (1.1383, 1.4624)	0.0001			
RAD51AP1	1.2278 (1.0348, 1.4568)	0.0187	0.5962 (0.3927, 0.9052)	0.0152	−0.5172
UBE2C	1.2097 (1.0799, 1.3551)	0.0010	1.1744 (0.938, 1.4706)	0.1610	0.1608
KIF11	1.4513 (1.2191, 1.7276)	0.0000			
CDC20	1.2925 (1.1298, 1.4786)	0.0002			
CENPE	1.6591 (1.3094, 2.1023)	0.0000	1.7341 (0.9894, 3.0393)	0.0545	0.5505
CEP55	1.3132 (1.1228, 1.5359)	0.0007			
TOP2A	1.1981 (1.0513, 1.3654)	0.0067	0.6710 (0.4736, 0.9507)	0.0248	−0.3989
TYMS	1.4399 (1.2047, 1.7209)	0.0001	1.5270 (1.0734, 2.1721)	0.0186	0.4233
KIF20A	1.5165 (1.2591, 1.8266)	0.0000	1.4245 (0.9433, 2.1511)	0.0925	0.3538
ASPM	1.5214 (1.2596, 1.8375)	0.0000			
BUB1	1.3451 (1.1377, 1.5903)	0.0005			
TRIP13	1.2308 (1.0550, 1.4360)	0.0083			
STIL	1.4508 (1.1313, 1.8605)	0.0034	0.5235 (0.2895, 0.9465)	0.0322	−0.6472
CDKN3	1.3521 (1.1623, 1.5730)	0.0001	1.3149 (0.9192, 1.8809)	0.1339	0.2738
PTTG1	1.4207 (1.1964, 1.6871)	0.0001			
MCM4	1.2973 (1.0940, 1.5385)	0.0028			
MCM6	1.3489 (1.0960, 1.6602)	0.0047			
DEPDC1	1.4407 (1.2008, 1.7286)	0.0001			
NUSAP1	1.3527 (1.1465, 1.5960)	0.0003			
MCM2	1.2728 (1.0766, 1.5049)	0.0047			
PRC1	1.4413 (1.2203, 1.7022)	0.0000	1.6418 (1.0030, 2.6873)	0.0486	0.4958
KNTC1	1.3367 (1.0618, 1.6826)	0.0135			
AURKA	1.2579 (1.0781, 1.4677)	0.0036	0.7174 (0.5122, 1.0049)	0.0534	−0.3321
PAICS	1.4356 (1.1493, 1.7933)	0.0014			
MELK	1.3452 (1.1507, 1.5726)	0.0002			
HMMR	1.4166 (1.1914, 1.6844)	0.0001	1.4713 (1.0176, 2.1274)	0.0401	0.3862
TTK	1.3397 (1.1229, 1.5984)	0.0012			
PBK	1.2934 (1.1217, 1.4914)	0.0004			
CCNB1	1.3997 (1.1978, 1.6355)	0.0000			

HR, hazard ratio; Coef, regression coefficient of genes in the multivariate Cox regression analysis.
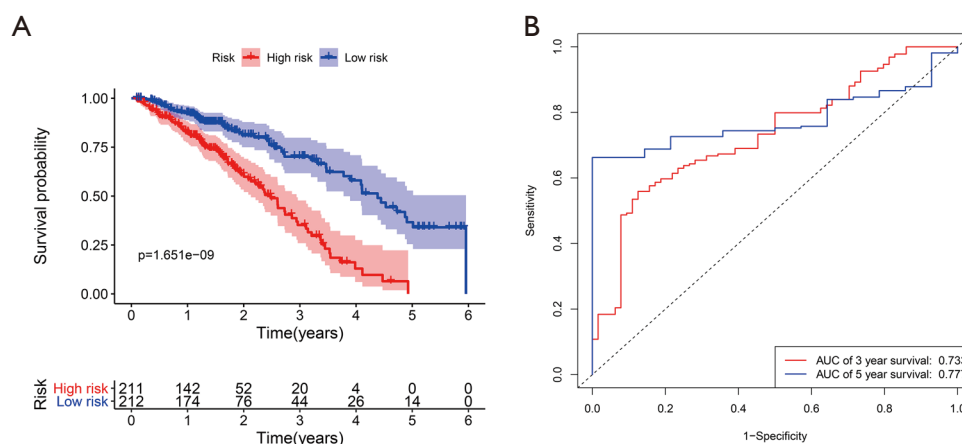
A

B

Figure 4 K-M and ROC curves based on the risk score model. (A) K-M curve of the high-risk (red) and low-risk (blue) LUAD patients. (B) Three-year (red) and five-year (blue) ROC curves of the risk score model.

and high-risk groups.

K-M survival curve analysis was performed to determine the difference between the two groups. The median survival time of the low-risk patients was 4.38 years, while that of the high-risk patients was 2.48 years (*Figure 4A*). The low-risk patients had a significantly better OS than the high-risk patients (3-year OS: 70.2% *vs.* 35.2%; 5-year OS: 36.6% *vs.* 0; P=1.651E-09), and the areas under the ROC curve (AUCs) were good (3-year AUC: 0.733; 5-year AUC: 0.777) (*Figure 4B*).

Spearman's correlation between the Treg infiltration level and expression level of 13 hub genes were illustrated using the R package "ggstatsplot" (*Figure 5*). Of the 13 hub genes, we found that the expression levels of *AURKA*, *CCNB2*, *CDKN3*, *ECT2*, *HMMR*, *KIF20A*, *PRC1*, *UBE2C*, *RAD51AP1*, *TOP2A* and *TYMS* all had a positive correlation with Treg infiltration in the tumor microenvironment. By contrast, the correlation between the gene expression (of CENPE and STIL) and the infiltration level was negative.

Prognostic nomogram model for LUAD: construction and validation

To apply Treg-related genes to clinical use, we constructed a nomogram model, combining the risk score with traditional clinical indicators, to predict the prognosis of LUAD. Given the high correlation between pathologic staging of M and T/N/stage, we finally included five clinical indicators (age, gender, stage, T and N) in the model (*Figure 6*).

To test the discriminability of the model, we implemented the calibration curve, which is used to assess the accuracy of the (disease) risk model in predicting the probability of an individual outcome event in the future and reflects the degree of consistency between the predicted model risk and actual occurrence risk (23). In the study, the calibration curves showed that the predicted survival rate was consistent with the actual incidence rate within 3/5 years (*Figure 7A,B*). Moreover, we plotted the K–M curve, which showed good discriminating ability of the nomogram (P=7.073e–13) (*Figure 8A*), and the AUC was improved (3-year AUC: 0.763; 5-year AUC: 0.873) (*Figure 8B*).

GO and KEGG pathway enrichment analyses

To identify molecular mechanisms of the candidate genes in LUAD, GO and KEGG pathway enrichment analyses were performed (*Table 2*). GO analysis includes 3 categories: biological processes (BP), cellular components (CC) and molecular function (MF). We found that the top enriched terms were (mitotic) nuclear division, spindle organization, organelle fission and chromosome segregation in BP; spindle, kinetochore and chromosome/centromeric region in CC; protein serine/threonine kinase activity, DNA-dependent ATPase activity and catalytic activity, acting on DNA in MF (*Figure 9A*). For KEGG enrichment pathways, the Treg-related hub genes were mostly enriched in the cell cycle, oocyte meiosis, DNA replication, p53 signaling pathway, human T-cell leukemia virus 1 infection, cellular
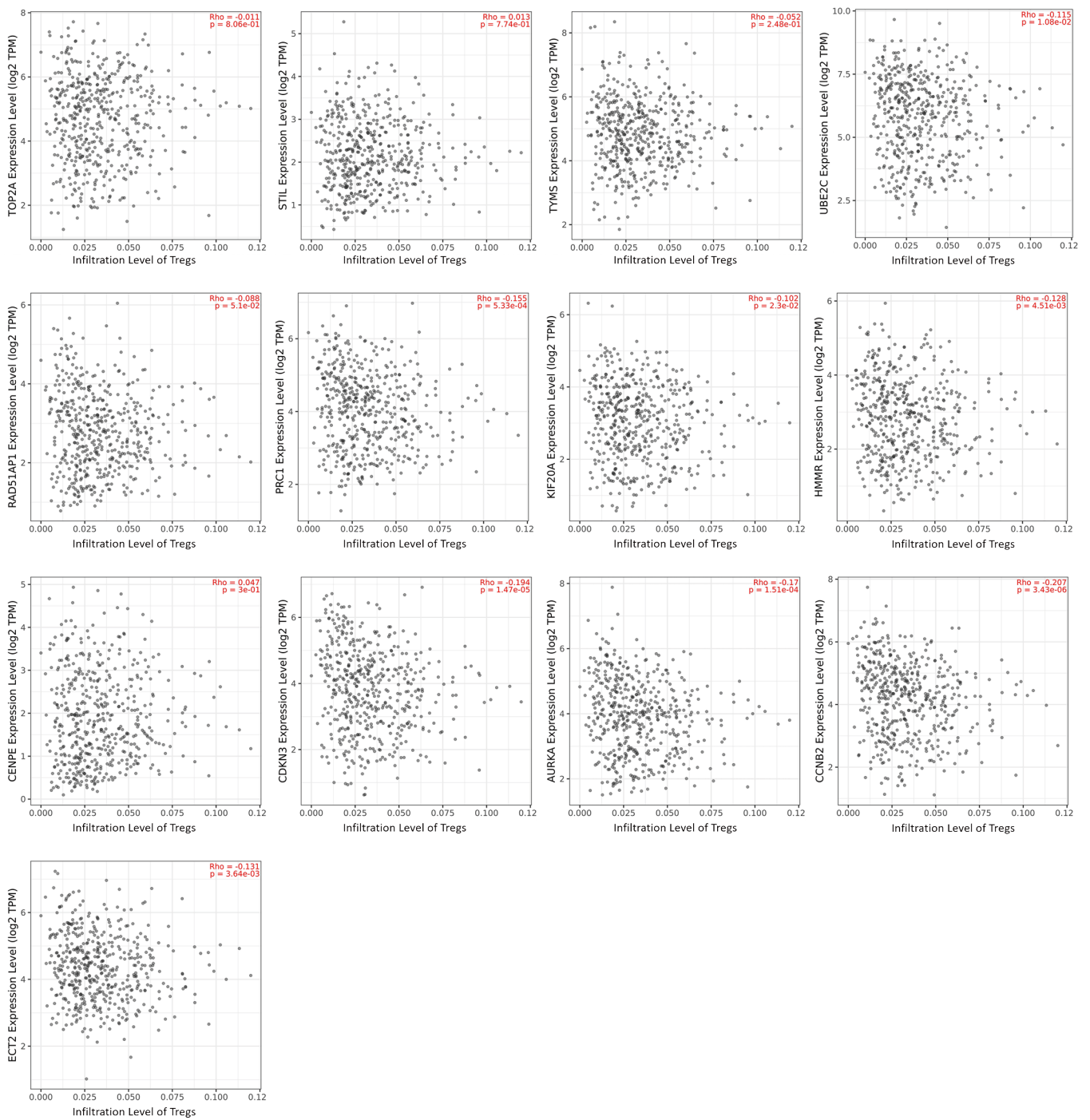
Figure 5 Spearman's correlations between 13 candidate genes and the infiltration level of Tregs (the "Rho" in the pictures indicates the Spearman's rank correlation coefficient, and "p" indicates the P value).
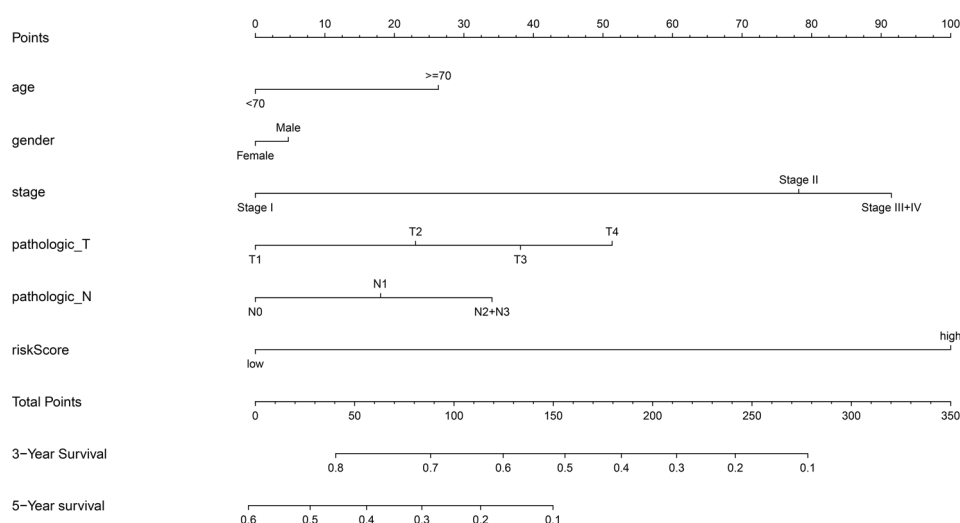
Figure 6 Prognostic nomogram for lung adenocarcinoma. According to the 6 variables (age, gender, stage, pathologic_T, pathologic_N and riskScore) in the model, 6 corresponding "points" values can be obtained, and the "total points" can be calculated by summing them. Therefore, the 3-/5-year survival rate of patients can be predicted
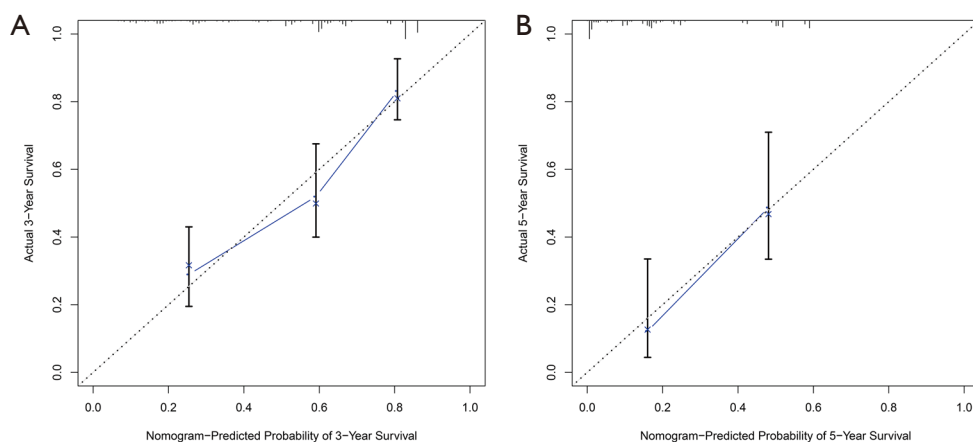
Figure 7 Calibration curve of the nomogram model at the 3-/5-year survival. Good concordance was obtained at the 3-year (A) and 5-year (B) year survivals of the nomogram-predicted probability with the actual survival.

senescence and ubiquitin-mediated proteolysis (*Figure 9B*).

Discussion

As the predominant histological phenotype of lung cancer, LUAD has a poor prognosis with a 5-year survival rate less than 25% (24,25), likely because of its undetected pathogenesis and complicated patterns of invasive growth, such as lymphovascular invasion, pleural invasion, and aerogenous invasion (26,27). Histopathological analysis

revealed that the infiltration of inflammatory cells and lymphocytes is an important activity of the tumor microenvironment that may impact tumorigenesis, invasion, metastasis, and prognosis. Therefore, increasing emphasis has been placed on immune infiltration cell targeting compared with direct tumor cell killing (28). In recent years, immunotherapy for LUAD has advanced rapidly and has been markedly beneficial to patients (29). The 2018 Nobel Prize in physiology and medicine was awarded to Professor James P. Allison, and Professor Tasuku Honjo for
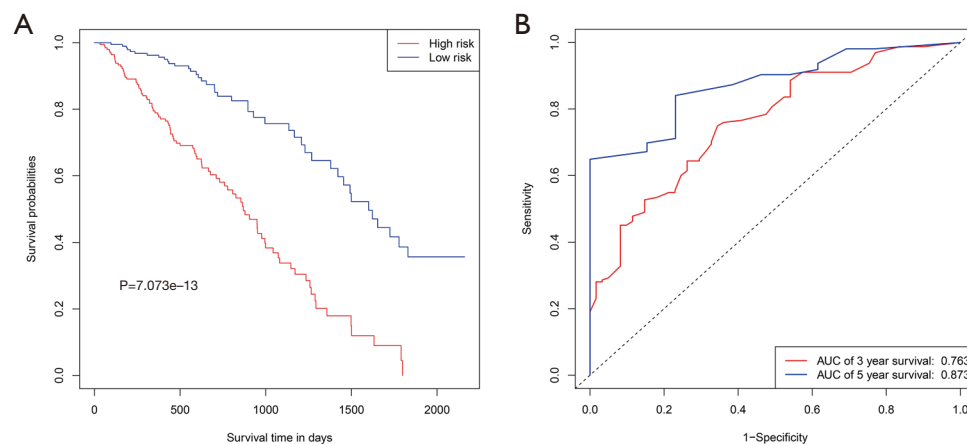
Figure 8 K-M and ROC curves based on the nomogram model. (A) K-M curve of high-risk (red) and low-risk (blue) LUAD patients. (B) Three-year (red) and five-year (blue) ROC curves of the risk score model.

their contribution to immune checkpoint therapy. To date, antibodies against PD-1/L1—nivolumab, pembrolizumab, cemiplimab, atezolizumab, durvalumab and avelumab—have been approved by the U.S. Food and Drug Administration (FDA) for first-line and/or late-stage treatment of 17 cancers (including NSCLC) (30,31). Pembrolizumab has been recommended for non-squamous cell carcinoma patients with high PD-L1 expression levels (tumor proportion score ≥50%) (32). However, the intratumor or inter-tumor heterogeneity and the non-standardized cut-off values for PD-1/L1, tumor mutation burden (TMB) and other independent immune-related biomarkers are far from being efficient (33-36). Thus, more potential and effective biomarkers are required.

In this study, we integrated three different GEO datasets (GSE10072/GSE32863/GSE43458) and obtained 196 LUAD and 137 normal samples. First, we acquired the abundances of 22 tumor-infiltrating immune cells using a deconvolution algorithm with the help of the CIBERSORTx portal, and then we identified candidate modules (black) and 36 candidate genes highly related to Tregs using the method of weighting gene coexpression network and PPIs. To verify the Treg-related genes, we generated the expression data of Treg-related genes as well as the clinical parameters from the TCGA database and then implemented a Cox proportional hazards model to calculate a risk score for each LUAD patient. The model performed well in that the low-risk patients had significantly longer 3- and 5-year survival times than the high-risk patients. Moreover, to further apply the Treg-related risk score in the clinic, we constructed a prognostic

nomogram for LUAD, integrating the Treg-related risk score with traditional clinical parameters (age, gender, stage, T and N). The AUC value and calibration curve indicated that the nomogram performed better. Additionally, GO enrichment analysis revealed that the prognostic Treg-related genes were mainly enriched in BP involving cell proliferation, such as mitosis, chromosome separation, and DNA replicase activity. For the KEGG pathway analysis, Treg-related genes were mostly involved in the cell cycle, DNA replication and ubiquitin-mediated proteolysis, which were similar to the GO enriched terms for the cell cycle. This may reflect the reason for Treg enrichment in the tumor microenvironment, a finding that was consistent with that in previous studies (37-39). Moreover, the p53 and FoxO signaling pathways were included in the KEGG enriched pathways. P53 is a tumor suppressor that monitors the cell cycle, maintains genomic stability by participating in DNA repair and is coexpressed with angiogenic genes such as Smad4 to inhibit tumor angiogenesis (40,41). FoxO is a nuclear protein subfamily that mediates the inhibition of insulin or insulin-like growth factor to further influence cell cycle regulation, energy metabolism, protein stability, oxidative stress, apoptosis, and immunity (42,43).

Akimova *et al.* found that the number and inhibitory function of Treg intratumor were significantly higher than those in blood, lungs and lymph nodes by single-cell studies (37,38). Conventional research has highlighted the protective role of Tregs in alleviating inflammation in autoimmune diseases (44). Xie *et al.* showed that Tregs recruited to tumors played a role as "accomplices" in helping tumor cells escape immunological surveillance (45).

Transl Lung Cancer Res 2021;10(1):167-182 | http://dx.doi.org/10.21037/tlcr-20-822

178

Wang et al. A Treg-infiltrating nomogram for lung adenocarcinoma

Table 2 GO and KEGG pathway enrichment analysis of candidate genes in the most significant terms

Terms	ID	Description	Gene ratio	P_{adjust}	Gene ID	Count
Biological process	GO:0000280	Nuclear division	15/35	4.15E-14	CENPF/TPX2/UBE2C/KIF11/CDC20/CENPE/ TOP2A/ASPM/TRIP13/NUSAP1/PRC1/KNTC1/ AURKA/TTK/CCNB1	15
	GO:0048285	Organelle fission	15/35	8.22E-14	CENPF/TPX2/UBE2C/KIF11/CDC20/CENPE/ TOP2A/ASPM/TRIP13/NUSAP1/PRC1/KNTC1/ AURKA/TTK/CCNB1	15
	GO:0140014	Mitotic nuclear division	13/35	8.22E-14	CENPF/TPX2/UBE2C/KIF11/CDC20/CENPE/ TRIP13/NUSAP1/PRC1/KNTC1/AURKA/TTK/ CCNB1	13
	GO:0007088	Regulation of mitotic nuclear division	11/35	6.79E-13	CENPF/UBE2C/KIF11/CDC20/CENPE/TRIP13/ NUSAP1/KNTC1/AURKA/TTK/CCNB1	11
	GO:0051783	Regulation of nuclear division	11/35	2.50E-12	CENPF/UBE2C/KIF11/CDC20/CENPE/TRIP13/ NUSAP1/KNTC1/AURKA/TTK/CCNB1	11
	GO:0007051	Spindle organization	10/35	2.06E-11	TPX2/KIF11/CDC20/CENPE/ASPM/STIL/ PRC1/AURKA/TTK/CCNB1	10
	GO:1902850	Microtubule cytoskeleton organization involved in mitosis	9/35	1.01E-10	TPX2/KIF11/CDC20/CENPE/STIL/NUSAP1/ PRC1/TTK/CCNB1	9
	GO:0007059	Chromosome segregation	11/35	2.72E-10	CENPF/ECT2/CDC20/CENPE/TOP2A/BUB1/ TRIP13/NUSAP1/PRC1/TTK/CCNB1	11
	GO:0051983	Regulation of chromosome segregation	8/35	6.96E-10	CENPF/ECT2/CDC20/CENPE/BUB1/TRIP13/ TTK/CCNB1	8
	GO:0098813	Nuclear chromosome segregation	10/35	6.96E-10	CENPF/ECT2/CDC20/CENPE/BUB1/TRIP13/ NUSAP1/PRC1/TTK/CCNB1	10
Cellular component	GO:0005819	Spindle	13/35	8.73E-13	CENPF/ECT2/TPX2/KIF11/CDC20/CENPE/ KIF20A/ASPM/PRC1/KNTC1/AURKA/TTK/ CCNB1	13
	GO:0000922	Spindle pole	9/35	2.76E-10	CENPF/TPX2/KIF11/CDC20/ASPM/PRC1/ KNTC1/AURKA/CCNB1	9
	GO:0030496	Midbody	8/35	2.02E-08	CENPF/ECT2/CENPE/CEP55/KIF20A/ASPM/ PRC1/AURKA	8
	GO:0098687	Chromosomal region	9/35	1.62E-07	CENPF/CENPE/BUB1/MCM4/MCM6/MCM2/ KNTC1/TTK/CCNB1	9
	GO:0072686	Mitotic spindle	6/35	3.66E-07	ECT2/TPX2/KIF11/CENPE/ASPM/AURKA	6
	GO:0005876	Spindle microtubule	5/35	5.84E-07	KIF11/CENPE/PRC1/KNTC1/AURKA	5
	GO:0000776	Kinetochore	6/35	1.48E-06	CENPF/CENPE/BUB1/KNTC1/TTK/CCNB1	6
	GO:0005813	Centrosome	9/35	1.48E-06	CCNB2/CENPF/CDC20/CEP55/ASPM/STIL/ AURKA/HMMR/CCNB1	9
	GO:0042555	MCM complex	3/35	7.85E-06	MCM4/MCM6/MCM2	3
	GO:0000775	Chromosome, centromeric region	6/35	7.85E-06	CENPF/CENPE/BUB1/KNTC1/TTK/CCNB1	6

Table 2 (*continued*)

Table 2 (*continued*)

Terms	ID	Description	Gene ratio	P_{adjust}	Gene ID	Count
Molecular function	GO:0004674	Protein serine/threonine kinase activity	6/35	9.86E-03	*BUB1/AURKA/CDK4/MELK/TTK/PBK*	6
	GO:0008094	DNA-dependent ATPase activity	3/35	2.04E-02	*TOP2A/MCM4/MCM6*	3
	GO:0016538	Cyclin-dependent protein serine/threonine kinase regulator activity	2/35	2.04E-02	*CDK4/CCNB1*	2
	GO:0003697	Single-stranded DNA binding	3/35	2.07E-02	*RAD51AP1/MCM4/MCM6*	3
	GO:0004003	ATP-dependent DNA helicase activity	2/35	3.66E-02	*MCM4/MCM6*	2
	GO:0004386	Helicase activity	3/35	3.80E-02	*MCM4/MCM6/MCM2*	3
	GO:0004712	Protein serine/threonine/tyrosine kinase activity	2/35	4.11E-02	*AURKA/TTK*	2
	GO:0003678	DNA helicase activity	2/35	4.11E-02	*MCM4/MCM6*	2
	GO:0140097	Catalytic activity, acting on DNA	3/35	4.11E-02	*TOP2A/MCM4/MCM6*	3
	GO:0008026	ATP-dependent helicase activity	2/35	4.62E-02	*MCM4/MCM6*	2
KEGG pathways	hsa04110	Cell cycle	10/17	9.47E-15	*CCNB2/CDC20/BUB1/PTTG1/MCM4/MCM6/MCM2/CDK4/TTK/CCNB1*	10
	hsa04114	Oocyte meiosis	6/17	1.56E-07	*CCNB2/CDC20/BUB1/PTTG1/AURKA/CCNB1*	6
	hsa04914	Progesterone-mediated oocyte maturation	4/17	4.57E-05	*CCNB2/BUB1/AURKA/CCNB1*	4
	hsa03030	DNA replication	3/17	5.39E-05	*MCM4/MCM6/MCM2*	3
	hsa04115	p53 signaling pathway	3/17	4.47E-04	*CCNB2/CDK4/CCNB1*	3
	hsa05166	Human T-cell leukemia virus 1 infection	4/17	9.68E-04	*CCNB2/CDC20/PTTG1/CDK4*	4
	hsa04218	Cellular senescence	3/17	4.00E-03	*CCNB2/CDK4/CCNB1*	3
	hsa04068	FoxO signaling pathway	2/17	3.06E-02	*CCNB2/CCNB1*	2
	hsa04120	Ubiquitin mediated proteolysis	2/17	3.46E-02	*UBE2C/CDC20*	2
	hsa00670	One carbon pool by folate	1/17	4.15E-02	*TYMS*	1

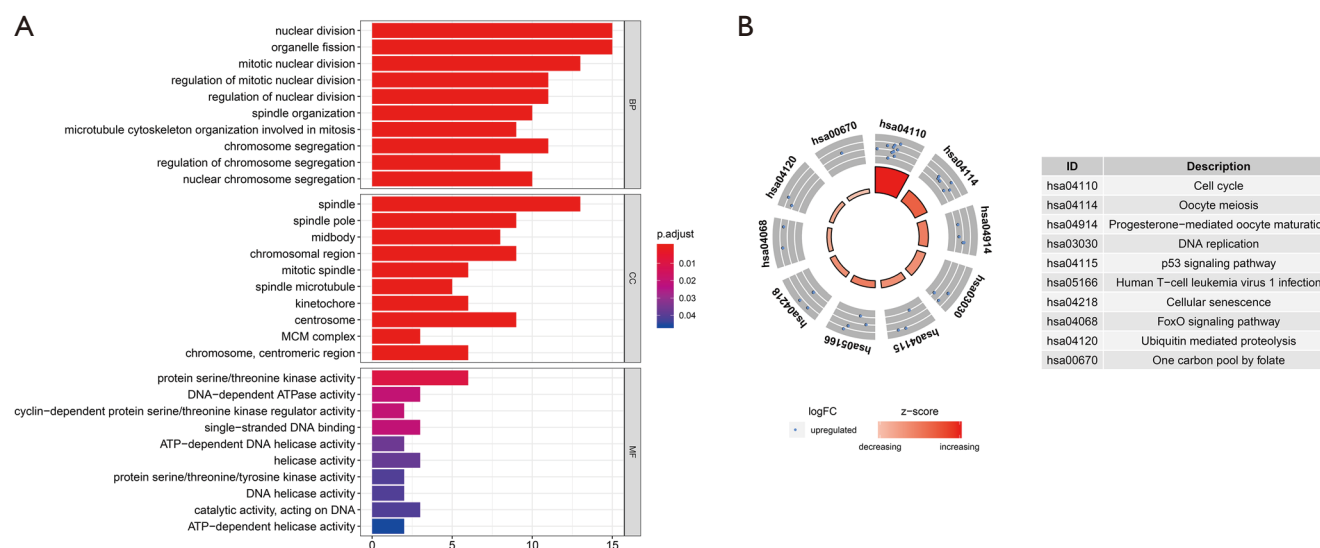GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure 9 Gene Ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses. (A) Enriched GO terms. The changing colors from blue to red elucidate the P$_{adjust}$ value increasing, and the length of the bar indicates the numbers of gene enrichment terms. (B) Enriched KEGG pathways. The depth of red indicates the size of the Z value, and the number of blue points indicates the number of enriched genes.

Shimizu *et al.* and Marshall *et al.* revealed that the enrichment of Tregs in tumors usually indicates a poor prognosis (16,46). However, the latest discovery of Ferreira *et al.* found that Tregs promoted the differentiation of CD8$^+$ effector memory T cells into tissue-resident memory T cells by providing the necessary cytokines, yielding more effective antitumor immunity (18). Therefore, studying the role of Treg cell infiltration into tumor tissue may provide a new perspective for immunotherapy or prognosis of LUAD.

Infiltrating Tregs in the tumor microenvironment play a potentially important role, which has been partially confirmed and applied in our study. However, this study has certain limitations. Although we drew the conclusion through multiple sequencing data and across different databases, more studies are needed to verify our results. Additionally, our results were obtained at the transcriptome level, and more proteomics level validation and clinical trials are needed to accelerate the clinical application.

Conclusions

We provide insights into the roles of Treg-related genes in LUAD and constructed a promising nomogram, which may help clinicians formulate more adequate treatment and follow-up plans.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at http://dx.doi.org/10.21037/tlcr-20-822

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tlcr-20-822). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International

References

1. Fitzmaurice C, Abate D, Abbasi N, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol 2019;5:1749-68.

2. Senosain MF, Massion PP. Intratumor Heterogeneity in Early Lung Adenocarcinoma. Front Oncol 2020;10:349.

3. He R, Zuo S. A Robust 8-Gene Prognostic Signature for Early-Stage Non-small Cell Lung Cancer. Front Oncol 2019;9:693.

4. Wang X, Yao S, Xiao Z, et al. Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes. J Transl Med 2020;18:149.

5. Peltanova B, Raudenska M, Masarik M. Effect of tumor microenvironment on pathogenesis of the head and neck squamous cell carcinoma: a systematic review. Mol Cancer 2019;18:63.

6. Bremnes RM, Busund LT, Kilvaer TL, et al. The Role of Tumor-Infiltrating Lymphocytes in Development, Progression, and Prognosis of Non-Small Cell Lung Cancer. J Thorac Oncol 2016;11:789-800.

7. Whiteside TL. The tumor microenvironment and its role in promoting tumor growth. Oncogene 2008;27:5904-12.

8. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science 2011;331:1565-70.

9. Beyer M, Schultze JL. Regulatory T cells in cancer. Blood 2006;108:804-11.

10. Ganesan AP, Johansson M, Ruffell B, et al. Tumor-infiltrating regulatory T cells inhibit endogenous cytotoxic T cell responses to lung adenocarcinoma. J Immunol 2013;191:2009-17.

11. Domagala-Kulawik J, Osinska I, Hoser G. Mechanisms of immune response regulation in lung cancer. Transl Lung Cancer Res 2014;3:15-22.

12. Vinay DS, Ryan EP, Pawelec G, et al. Immune evasion in cancer: Mechanistic basis and therapeutic strategies. Semin Cancer Biol 2015;35 Suppl:S185-98.

13. Koyama K, Kagamu H, Miura S, et al. Reciprocal CD4+ T-cell balance of effector CD62Llow CD4+ and CD62LhighCD25+ CD4+ regulatory T cells in small cell lung cancer reflects disease stage. Clin Cancer Res 2008;14:6770-9.

14. Wang W, Hodkinson P, McLaren F, et al. Small cell lung cancer tumour cells induce regulatory T lymphocytes, and patient survival correlates negatively with FOXP3+ cells in tumour infiltrate. Int J Cancer 2012;131:E928-37.

15. Petersen RP, Campa MJ, Sperlazza J, et al. Tumor infiltrating Foxp3+ regulatory T-cells are associated with recurrence in pathologic stage I NSCLC patients. Cancer 2006;107:2866-72.

16. Shimizu K, Nakata M, Hirami Y, et al. Tumor-infiltrating Foxp3+ regulatory T cells are correlated with cyclooxygenase-2 expression and are associated with recurrence in resected non-small cell lung cancer. J Thorac Oncol 2010;5:585-90.

17. Erfani N, Mehrabadi SM, Ghayumi MA, et al. Increase of regulatory T cells in metastatic stage and CTLA-4 over expression in lymphocytes of patients with non-small cell lung cancer (NSCLC). Lung Cancer 2012;77:306-11.

18. Ferreira C, Barros L, Baptista M, et al. Type 1 Treg cells promote the generation of CD8(+) tissue-resident memory T cells. Nat Immunol 2020;21:766-76.

19. Rusk N. Expanded CIBERSORTx. Nat Methods 2019;16:577.

20. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 2008;9:559.

21. Jayaraman A, Jamil K, Khan HA. Identifying new targets in leukemogenesis using computational approaches. Saudi J Biol Sci 2015;22:610-22.

22. Mattes MD, Weber WA, Foster A, et al. A Predictive Model for Lymph Node Involvement with Malignancy on PET/CT in Non-Small-Cell Lung Cancer. J Thorac Oncol 2015;10:1207-12.

23. Huang YQ, Liang CH, He L, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol 2016;34:2157-64.

24. Li Y, Tao L, Cai W. Profiles of Immune Infiltration and Prognostic Immunoscore in Lung Adenocarcinoma. Biomed Res Int 2020;2020:5858092.

25. Zhang P, Li S, Lv C, et al. BPI-9016M, a c-Met inhibitor, suppresses tumor cell growth, migration and invasion of lung adenocarcinoma via miR203-DKK1. Theranostics 2018;8:5890-902.

26. Shih AR, Mino-Kenudson M. Updates on spread through air spaces (STAS) in lung cancer. Histopathology 2020;77:173-80.

27. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol 2015;10:1243-60.

28. Pilipow K, Darwich A, Losurdo A. T-cell-based breast cancer immunotherapy. Semin Cancer Biol 2020. [Epub ahead of print]. doi: 10.1016/j.semcancer.2020.05.019.

29. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. N Engl J Med 2016;375:1823-33.

30. Topalian SL, Taube JM, Pardoll DM. Neoadjuvant checkpoint blockade for cancer immunotherapy. Science 2020;367:eaax0182.

31. Santarpia M, Aguilar A, Chaib I, et al. Non-Small-Cell Lung Cancer Signaling Pathways, Metabolism, and PD-1/PD-L1 Antibodies. Cancers (Basel) 2020;12:1475.

32. Hanna NH, Schneider BJ, Temin S, et al. Therapy for Stage IV Non-Small-Cell Lung Cancer Without Driver Alterations: ASCO and OH (CCO) Joint Guideline Update. J Clin Oncol 2020;38:1608-32.

33. Greillier L, Tomasini P, Barlesi F. The clinical utility of tumor mutational burden in non-small cell lung cancer. Transl Lung Cancer Res 2018;7:639-46.

34. Dempke WCM, Fenchel K, Dale SP. Programmed cell death ligand-1 (PD-L1) as a biomarker for non-small cell lung cancer (NSCLC) treatment-are we barking up the wrong tree? Transl Lung Cancer Res 2018;7:S275-9.

35. Bai X, Wu DH, Ma SC, et al. Development and validation of a genomic mutation signature to predict response to PD-1 inhibitors in non-squamous NSCLC: a multicohort study. J Immunother Cancer 2020;8:e000381.

36. Keenan TE, Burke KP, Van Allen EM. Genomic correlates of response to immune checkpoint blockade. Nat Med 2019;25:389-402.

37. Akimova T, Zhang T, Negorev D, et al. Human lung tumor FOXP3+ Tregs upregulate four "Treg-locking" transcription factors. JCI Insight 2017;2:e94075.

38. Ward-Hartstonge KA, Kemp RA. Regulatory T-cell heterogeneity and the cancer immune response. Clin Transl Immunology 2017;6:e154.

39. Baratelli F, Lee JM, Hazra S, et al. PGE(2) contributes to TGF-beta induced T regulatory cell function in human non-small cell lung cancer. Am J Transl Res 2010;2:356-67.

40. Tollini LA, Jin A, Park J, et al. Regulation of p53 by Mdm2 E3 ligase function is dispensable in embryogenesis and development, but essential in response to DNA damage. Cancer Cell 2014;26:235-47.

41. Fan G, Tu Y, Chen C, et al. DNA methylation biomarkers for hepatocellular carcinoma. Cancer Cell Int 2018;18:140.

42. Peixoto MS, da Silva Junior FC, de Oliveira Galvao MF, et al. Oxidative stress, mutagenic effects, and cell death induced by retene. Chemosphere 2019;231:518-27.

43. Wu Y, Peng H, Cui M, et al. CXCL12 increases human neural progenitor cell proliferation through Akt-1/FOXO3a signaling pathway. J Neurochem 2009;109:1157-67.

44. Miyara M, Ito Y, Sakaguchi S. TREG-cell therapies for autoimmune rheumatic diseases. Nat Rev Rheumatol 2014;10:543-51.

45. Xie M, Wei J, Xu J. Inducers, Attractors and Modulators of CD4(+) Treg Cells in Non-Small-Cell Lung Cancer. Front Immunol 2020;11:676.

46. Marshall EA, Ng KW, Kung SH, et al. Emerging roles of T helper 17 and regulatory T cells in lung cancer progression and metastasis. Mol Cancer 2016;15:67.