European Review for Medical and Pharmacological Sciences 2022; 26: 1304-1319

Telomere length-related signature as a novel biomarker of prognosis and immune response in non-small cell lung cancer

X.-G. LIU, M. LI, S.-J. MAI, R.-J. CAI

Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou, China

Abstract. – OBJECTIVE: Telomere length-related genes (TLRGs) play an important role in multiple tumors; however, there is a lack of systematic reporting about their relevance in non-small cell lung cancer (NSCLC). This study investigated the relation between TLRG gene expression and the immunotherapeutic response of patients with NSCLC.

MATERIALS AND METHODS: Differentially expressed TLRGs in tumor tissues and normal tissues were screened using Gene Expression Omnibus (GEO) datasets. A univariate Cox regression analysis was performed to identify the optimal prognosis-related genes. A prognostic risk model was constructed by using least absolute shrinkage, selection operator, and multivariate Cox regression analysis results. The model was then evaluated by a Kaplan-Meier analysis, functional enrichment annotation, and a receiver operating characteristic curve analysis; after which, it was validated in the TCGA dataset. The model was used to predict immunotherapeutic response and drug sensitivity.

RESULTS: An 18-gene prognostic signature was developed and used to stratify NSCLC patients into a low- or high-risk group in GEO cohorts. Patients in the low-risk group had better survival possibilities than those in the high-risk group, and showed significantly higher overall survival times in the TCGA cohort. The risk score was identified as an independent prognostic factor, when compared with other clinical factors. ssGSEA scores showed that the risk model was mainly linked to cancer- and immune-related pathways. Importantly, the candidate risk model was linked to tumor immunity and predicted a patient's response to PDL-1 blockade immune therapy. Several potential drugs that might target this model were identified.

CONCLUSIONS: This study provides broad molecular signatures that can be used in further functional and therapeutic studies of the telomere system, and also represents an integrated approach for characterizing key protein complexes when creating a prognosis and identifying new targets for cancer immunotherapy.

Key Words:

Telomere length-related genes, Non-small cell lung cancer, Prognosis, Immune prognostic signature, Drug sensitivity.

Abbreviations

Non-small cell lung cancer (NSCLC), Telomere length (TL), Telomere length-related genes (TLRGs), Tumor microenvironment (TME), Relative telomere length (RTL), Gene Expression Omnibus (GEO), The Cancer Genome Atlas (TCGA), Differentially expressed telomere length-related genes (DE-TLRGs), Least absolute shrinkage and selection operator (LASSO), Receiver operating characteristic (ROC), Gene Set Variation Analysis (GSVA), Tumor mutation burden (TMB), Area under the ROC curve (AUC), Genomics of Drug Sensitivity in Cancer (GDSC).

Introduction

Lung cancer is a major cause of cancer-related death worldwide, and non-small cell lung cancer (NSCLC) accounts for ~ 85% of all lung cancer cases[1,2]. Despite the considerable progress and advances made in understanding lung cancer biology and multimodality treatments, the prognosis for lung cancer patients remains far from satisfactory due to the usual late diagnosis of the disease[3]. Some factors, such as tumor extension, performance status, and histological type can be used to predict a poor prognosis for NSCLC[4,5]. Up to now, a prediction of prognosis has mainly depended on the histopathologic diagnosis and tumor stage. However, several different disease outcomes have been observed in patients with similar clinical and pathological signatures, suggesting that the present clinical prognostic factors adopted may be insufficient to accurately predict the outcomes of NSCLC patients[6]. Consequently, there remains a need to discover new reliable

Corresponding Author: Ruijun Cai, MD; e-mail: cairuijunnfy@163.com

prognostic biomarkers that can improve current therapeutic strategies.

Telomeres are nucleoprotein complexes located at the ends of eukaryotic chromosomes. These specialized structures serve to protect the linear chromosome ends from end-to-end fusion, recombination, and degradation[7,8]. The association between telomere dysfunction, human carcinogenesis and disease progression has become widely recognized[9,10]. Several experimental studies have found shorter telomere lengths (TLs) in tumor tissues when compared with corresponding adjacent para-carcinoma tissues[9]. Furthermore, recent studies have reported that an abnormal alternation (shortening or lengthening) of TLs in peripheral blood leukocytes is significantly associated with an increased risk for multiple human malignancies, including lung cancer[10-15]. Several studies[16,17] have reported that NSCLC patients whose tumors had shorter TLs experienced a more rapid clinical progression of their disease. It is well documented that TL integrity is maintained by several TL-related components including shelterin, telomerase complex, and DNA repair proteins, which determine the balance between processes that shorten or lengthen the telomere[18]. However, the complicated relationship between telomer-related gene expression and tumorigenesis/prognosis requires further detailed analysis.

The immune escape of cancer cells is an important mechanism of tumorigenesis. Interestingly, TL alterations were found to regulate cancer immune functions and re-program the tumor microenvironment (TME)[19-21]. More specifically, TL shortening in leukocytes was reported to reduce the immune response[19]. An investigation of colon cancer patients with short relative telomere lengths (RTLs) showed that those patients had higher percentages of CD4+ T cells and the lower percentages of B cells among their peripheral blood mononuclear cells. Additionally, those patients had lower concentrations of plasma transforming growth factor-β1, suggesting that their immune functions had changed along with their RTLs[20]. A similar observation was reported in gastric cancer patients, where a short RTL enhanced a patient's immunosuppressive status[21]. These studies provided a glimpse into the roles played by RTL in cancer immunity. However, it has not been studied whether RTL acts to regulate the TME in NSCLC patients. When considering the great importance of the TME in immunotherapy, a comprehensive study is needed to determine the association between TLRGs and the local immune status of NSCLC patients. To gain a deeper understanding of how TLRGs participate in NSCLC progression, we obtained the expression profiles of 118 genes from The Gene Expression Omnibus (GEO) database. Our novel model is based on the optimal prognostic components of TLRGs, which were identified by univariate Cox regression analyses. A comprehensive analysis of multiple levels of data was subsequently performed on the risk model, with the variables being mRNA expression, patient survival, immune signature, and chemical compounds. Our integrated analysis of TLRGs provides an important theoretical resource for understanding TLRG biology, and might assist in developing an accurate prognosis for NSCLS patients and identifying therapeutic targets.

Materials and Methods

Datasets Acquisition and Preprocessing

The expression profile data and detailed survival data of NSCLC patients were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/, Accession Number: GSE30219, GSE31210, GSE37745, GSE50081) for further analysis.

Supplementary Figure 1 shows the workflow for establishing the risk score model.

RNA-seq data and corresponding clinicopathological information for the external validation cohort were obtained from The Cancer Genome Atlas (TCGA, https://portal.gdc.cancer.gov/) **(Supplementary Table I)**. The "ComBat" algorithm of the sva Package[22] was adopted to eliminate batch effects. The expression data in the TCGA dataset were log2 FPKM transformed prior to analysis. All analyses in this study were performed using the R (version 3.6.2) and R Bioconductor packages. In addition, patients with advanced urinary tract transitional cell carcinoma or metastatic urothelial cancer treated with immunotherapy were derived from two cancer studies[23,24] (http://research-pub.gene.com/IMvigor210CoreBiologies; https://github.com/hammerlab/multi-omic-urothelial-antipdl1). The relationship between TLRG_scores and immunotherapy prognosis was analyzed using the Survival R packages. The statistical significance of survival differences between the low and high groups was determined using the log-rank test.

Identification of Differentially Expressed Telomere Length-Related Genes (DE-TLRGs)

A total of 118 TLRGs were selected from the published literature[25], detailed information is provided in **Supplementary Table II**. Genes that were differentially expressed in NSCLC and normal samples in the GEO cohort were identified using the "limma" package (https://bioconductor.org/packages/release/bioc/). A p-value < 0.05 was considered to indicate a significant difference.

Development and Evaluation of the DE-TLRG Prognostic Model

A univariate Cox regression analysis was performed on the DE-TLRGs in the GEO cohort to determine the prognostic value of those candidate genes in NSCLC. To prevent gene omissions, a p-value<0.05 was set as the threshold, and 49 survival-related genes were screened for further study. The R package "glmnet" was used to perform a least absolute shrinkage and selection operator (LASSO) Cox regression analysis to narrow down the candidate genes. A total of 20 DE-TLRGs were included in a multifactor Cox regression analysis, and a risk model comprising 18 DE-TLRGs was finally established. The prognostic risk score model was established as follows: Risk Score = , where coef () indicates the coefficients counted by the multivariate Cox regression model, and represents a gene expression level[26]. A score was then calculated for each patient with NSCLC. Each patient was stratified into a high-risk (high risk score) or low-risk (low risk score) group based on the median value of their risk score, and OS differences between the two subgroups were assessed in a Kaplan–Meier analysis and by the two-sided log-rank test. The prognostic performance of the model was measured by a time dependent receiver operating characteristic (ROC) curve analysis. A 1-, 3-, and 5-year ROC curve analysis was performed using the "survival" and "survival ROC" R packages. Next, the performance of the TLRG-based risk score model was validated in the TCGA cohort using the same method.

Independence of the Candidate Risk Model

To determine whether the candidate prognostic model could be used independently, and was even superior to other predictors (e.g., age, sex, and stage) in predicating the prognosis of patients with NSCLC, multivariate Cox proportional hazards regression analyses were performed with the training sets and validation sets.

Establishment and Validation of a Predictive Nomogram

The predictive probabilities of the nomogram and other parameters (age, sex, risk score, and stage) for OS were generated. Correction curves based on the Hosmer-Lemeshow test were used to determine the uniformity between the clinical outcome and model prediction outcome of each patient.

Biological Pathway Analysis

To distinguish the difference in biological pathways between the two clusters, a Gene Set Variation Analysis (GSVA) was performed using the R package, "gsva." The gene sets "c2.cp.kegg.v7.1.symbols.gmt" and "h.all.v7.2" downloaded from the Molecular Signature Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) were selected as the reference gene sets.

Evaluation of the Immunological Characteristics of the TME

A list of immunomodulators and immune cell numbers was obtained from a recent publication by TCGA immune response working group[27,28]. The immunomodulators consisted of a group of immunoregulatory genes involving antigen presentation factor, ligand, and receptor, which are all critical for cancer immunotherapy. Spearman correlation coefficients between expression levels listed in the TLRG's publication and levels of immunomodulator mRNA expression were calculated. To determine the immune cell landscape of NSCLC, the R package "gsva" was used to perform a ssGSEA for the purpose of evaluating the level of immune infiltration (recorded as the ssGSEA score) in a single sample based on the expression of an immune cell-specific marker. We also analyzed the Spearman correlation coefficients between TLRG's expression levels and levels of immune infiltration. Finally, differences between the levels of immune cell infiltration in the high- and low-risk groups were demonstrated. Marker genes related to immune cell type were obtained from an article published by Charoentong et al[28].

Exploration of the Model in Immunotherapeutic Treatment

The tumor mutation burden (TMB) score of each patient in TCGA lung cancer cohort was cal-

culated using the maftools R package[29]. The TIDE algorithm was employed to predict the likelihood of an immunotherapeutic response in LUAD and LUSC patients in the TCGA cohort[30]. Gene expression data prior to a PDL-1 immune checkpoint blockade and overall survival data were gathered from two cancer studies (patients with advanced urinary tract transitional cell carcinoma or metastatic urothelial cancer treated with PDL-1 immune checkpoint blockade therapy)[23,24].

Results

Identification of DE-TRLGs in the Normal and Tumor Tissues of Patients with fNSCLC

The expression levels of 118 pyroptosis-related genes from 34 normal and 896 tumor tissues were compared in the GEO dataset. Those comparisons showed that 80 genes were differentially expressed; among which, 18 genes were downregulated and 62 were upregulated **(Supplementary Table III)**. The differential expression of those genes is shown in a volcano map (Figure 1A).

Development of a Prognostic Gene Model for the GEO Cohort

A total of 896 NSCLC samples from corresponding patients who had complete survival information were screened. A univariate Cox regression analysis was performed for the primary screening of DE-TRLG-associated prognostic genes in the GEO training set **(Supplementary Table IV)**. A total of 49 DE-TRLGs were found to have independent prognostic characteristics for NSCLC ($p < 0.05$) (Figure 1B). After conducting a LASSO Cox regression analysis, a 20-gene signature was established based on the optimal λ value (Figure 1C, D). Subsequently, a multivariate Cox ratio hazard regression analysis was performed to identify independent prognostic factors. A total of 18 DE-TRLGs showed a significant correlation with OS in the training queue and were used to develop a model for evaluating the prognostic risk for NSCLC patients. To further investigate the interactions of these pyroptosis-related genes, a linear regression analysis was performed. Those results demonstrated the co-occurrence (mutually exclusive) of 18 parameters; among which, NEK2 and AURKB, NEK2 and NLK, NEK2 and HNRNPC, NEK2 and RAP1A, NEK2 and HSP90AA1, RAP1A and POLA2, and RAP1A and AURKB exhibited significant correlations (Figure 1E). Subsequently, a protein-protein interaction network was established using the STRING database and Cytoscape software[31]. The minimum required interaction score for the protein-protein interaction network analysis was set at 0.9 (the highest confidence); the top 10 hub genes identified are shown in Figure 1F.

Based on the median value of the risk scores, all the candidate samples were assigned low- and high-risk groups **(Supplementary Table V)**. Patients in the low-risk group (n = 448) had a better prognosis than those in the high-risk group (n = 448) (Figure 2A). A significant difference in OS time between the low- and high-risk groups was detected ($p < 0.0001$, Figure 2C). To further determine the sensitivity and specificity of the prognostic model, a time dependent ROC analysis was performed. The result demonstrated a good predictive ability for the 1-, 3- and 5-year OS rates, with area under the ROC curve (AUC) values of 0.691, 0.69, and 0.673, respectively (Figure 2E).

Evaluation of the Prognostic Signature in the TCGA Cohorts

In order to prove that the prognostic model had a similar predictive capability in different populations, the TCGA cohort was utilized as an independent external validation set. All eligible NSCLC patients in the TCGA cohort were divided into a low-risk group (n = 495) and high-risk group (n = 495), respectively **(Supplementary Table VI)**. Patients with a high-risk score had shorter survival times and higher death rates than those with a low-risk score (Figure 2B). Furthermore, a Kaplan–Meier analysis also showed a significant difference in the survival rate between the low and high-risk subgroups ($p < 0.0001$, Figure 2D). An ROC curve analysis of the validation cohort indicated that the risk model generated by the TCGA validation set also had a strong predictive capacity (AUC = 0.612 for 1-year, 0.636 for 3-year, and 0.611 for 5-year survival) (Figure 2F).

Determination of the DE-TRLG Signature as an Independent Prognostic Factor

To further examine whether the risk score could serve as an independent prognostic factor, a multivariable Cox regression analyses was performed. The result showed that the risk score was an independent prognostic factor (HR = 0.57, 95% CI: 0.47–0.7 and HR: 0.58, 95% CI: 0.47–0.71) for patients with NSCLC in the GEO cohort (Figure 3A) and TCGA cohort, respectively (Figure 3B). The concordance index of the risk score and AUC

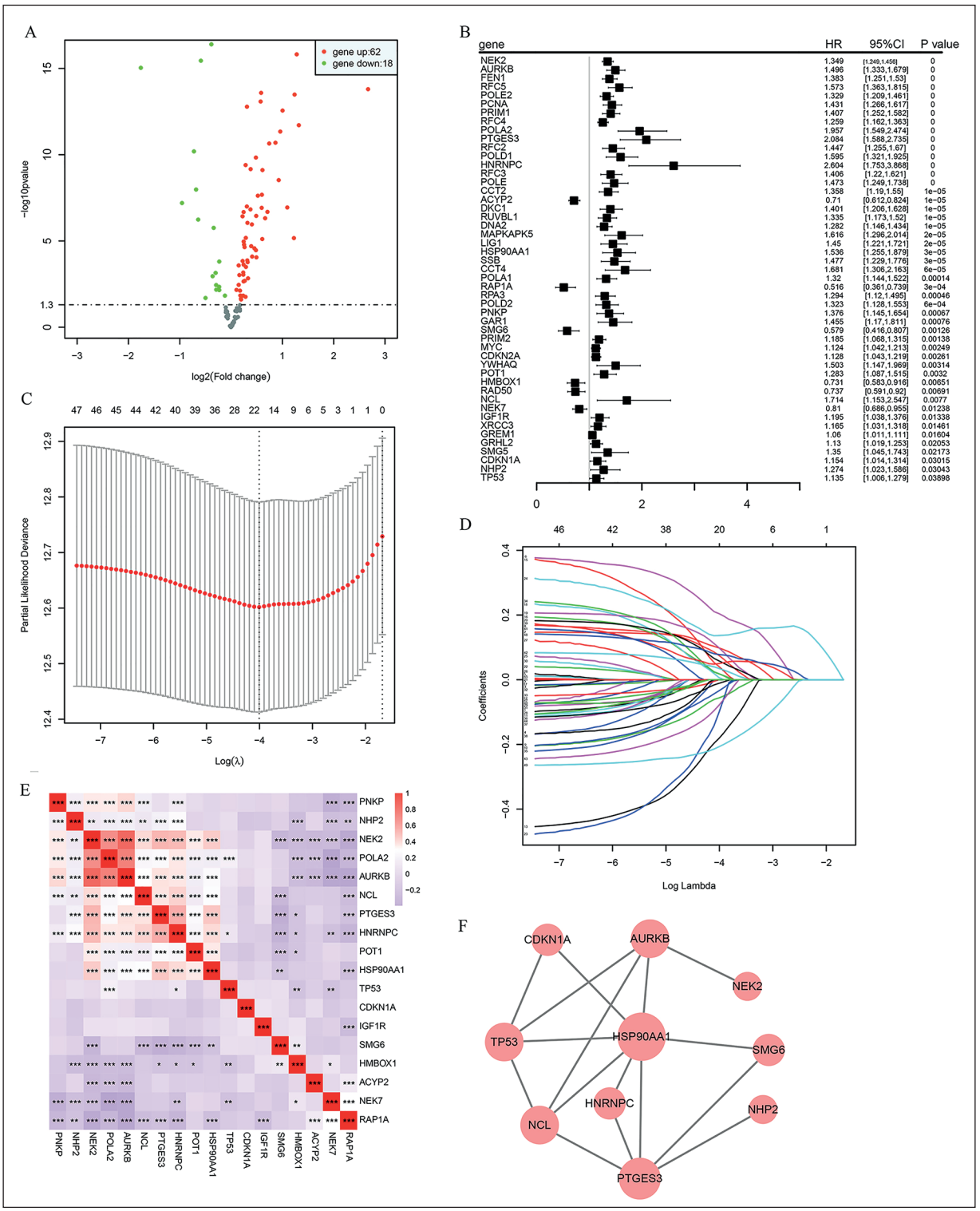

Figure 1. Construction of a prognostic model based on telomere-related genes. **A,** Volcano map of differentially expressed telomere-related genes in the normal samples and tumor samples. **B,** Univariate Cox regression analysis for differentially expressed telomere-related genes. The horizontal box plot represents the 95% confidence interval of the HR value. The green dashed line perpendicular to coordinate 1 is used to distinguish whether a gene is a protective factor or risk factor. The HR, 95% CI, and specific *p*-value are also displayed. **C-D,** Least absolute shrinkage and selection operator analysis; Log (λ) is the tuning parameter related to survival time. **E,** Analysis of gene expression correlation of the model-building genes. **F,** Analysis of the protein interaction network of model-building genes.

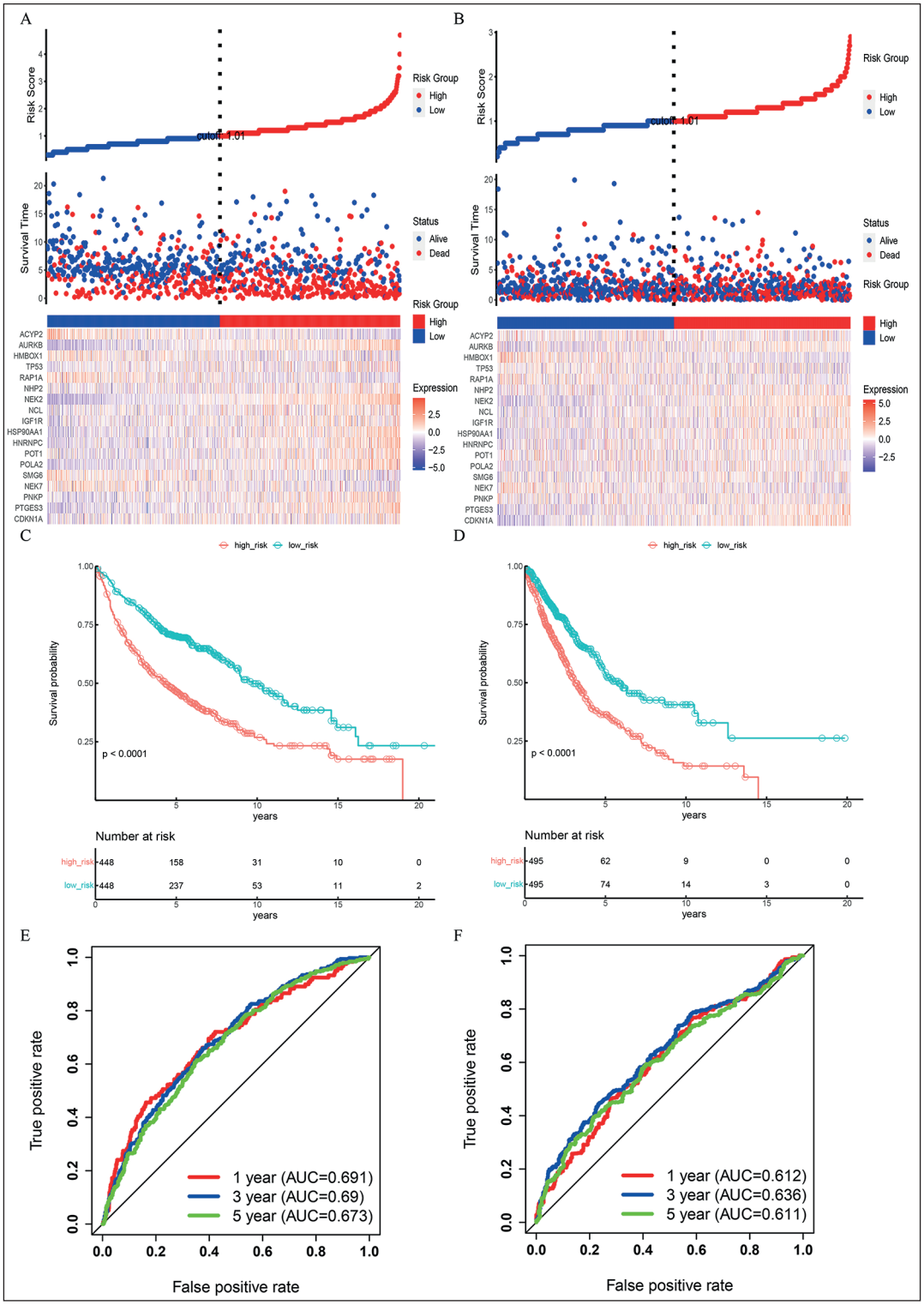

Figure 2. Prognosis analysis of patients in the high and low risk groups in two cohorts. **A-B,** The top part is the GEO or TCGA patient's risk score distribution predicted by the telomere-related gene model. The middle part is the GEO or TCGA patient's survival time and survival status under different risk groups. The bottom part is a gene expression heat map of the genes used to construct the model between different risk groups in the GEO or TCGA cohorts. **C-D,** Kaplan-Meier survival curve of the GEO or TCGA patients in the high and low risk groups. **E-F,** The ROC curves of the risk model used to predict a patient's prognosis in the GEO cohorts and TCGA cohorts.

were evaluated to better illustrate the uniqueness and susceptibility of risk scores. The concordance index and AUC of the risk score were always greater than those of other clinicopathological characteristics along with prolonged time, indicating that the prognostic risk model of 18 DE-TRLGs for NSCLC was comparatively dependable (Figure 3C-F).

Construction and Evaluation of the Prognostic Nomogram

In order to develop a clinically applicable method for predicting the prognosis of NSCLC patients, a prognostic nomogram was established to predict the survival probability at 1, 3, and 5 years based on the GEO training set. Four independent parameters (age, gender, risk score, and stage) were incorporated into the prediction model (Figure 4A). The Correlation diagram also demonstrated excellent agreement between prediction and observation for the 1-, 3- and 5-year OS probabilities of the patients (Figure 4B-D).

Function Analysis of the Low and High Risk Groups

In addition to telomere maintenance, TRLGs have also been reported to regulate multiple biological processes[32-35]. To investigate the biological pathways associated with the TRLG-based model, a GSVA analysis was performed to confirm the signaling pathways characteristic of the high- and low- subgroups **(Supplementary Tables VII and VIII)**. That analysis revealed both positive and negative correlation pathways in a context-depen-

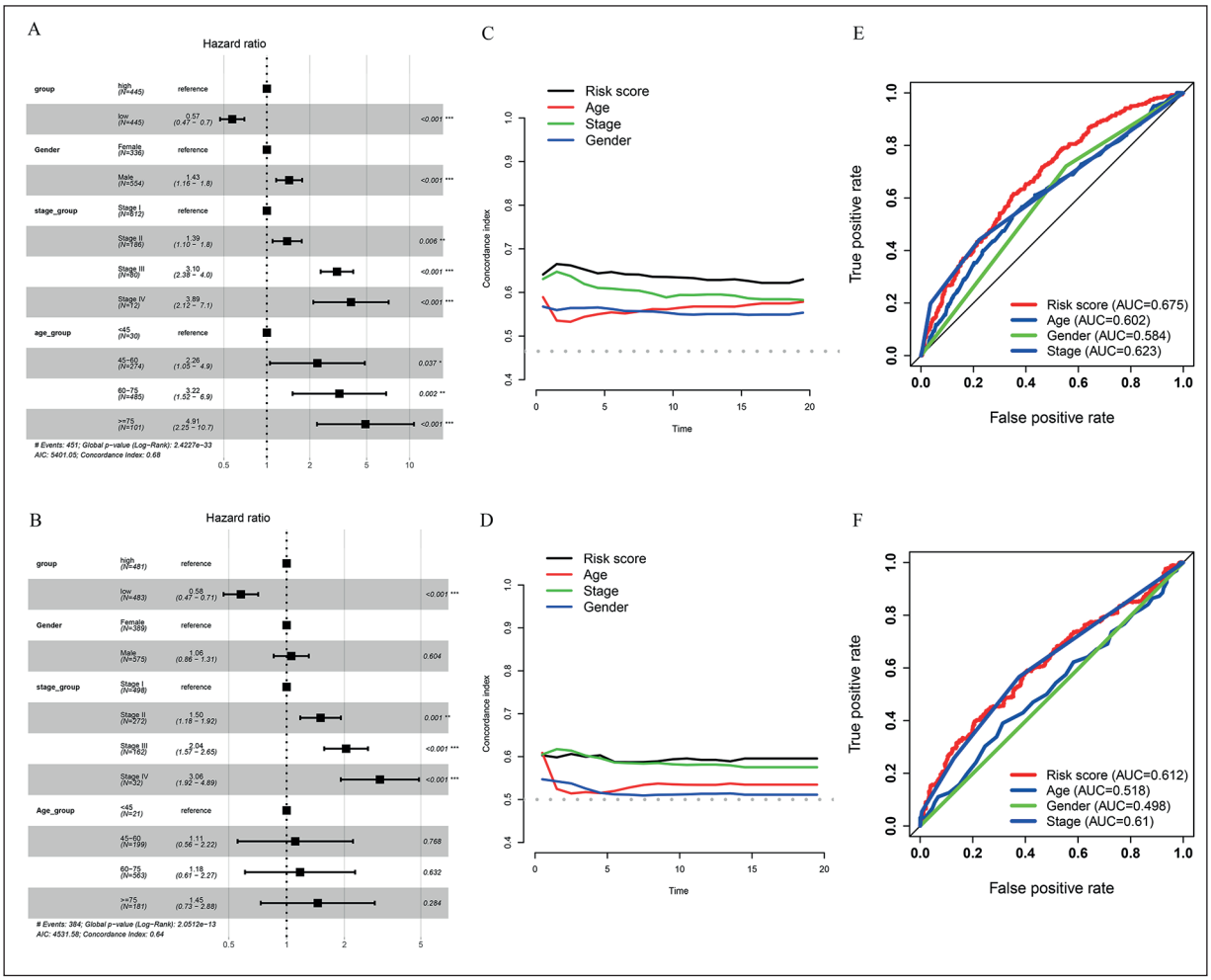

Figure 3. Identification of independent prognostic factors and a comparison of prognostic ability with clinical factors. **A-B,** A multivariate Cox regression model analysis, which included gender, age, pathological stage, and risk score in the GEO cohort and TCGA cohort. **C-D,** Concordance indexes of the risk scores and clinical characteristics in the GEO cohort and TCGA cohort. **E-F,** Multi-index ROC curve comparison, including gender, age, pathological stage, and risk score in the GEO cohort and TCGA cohort.

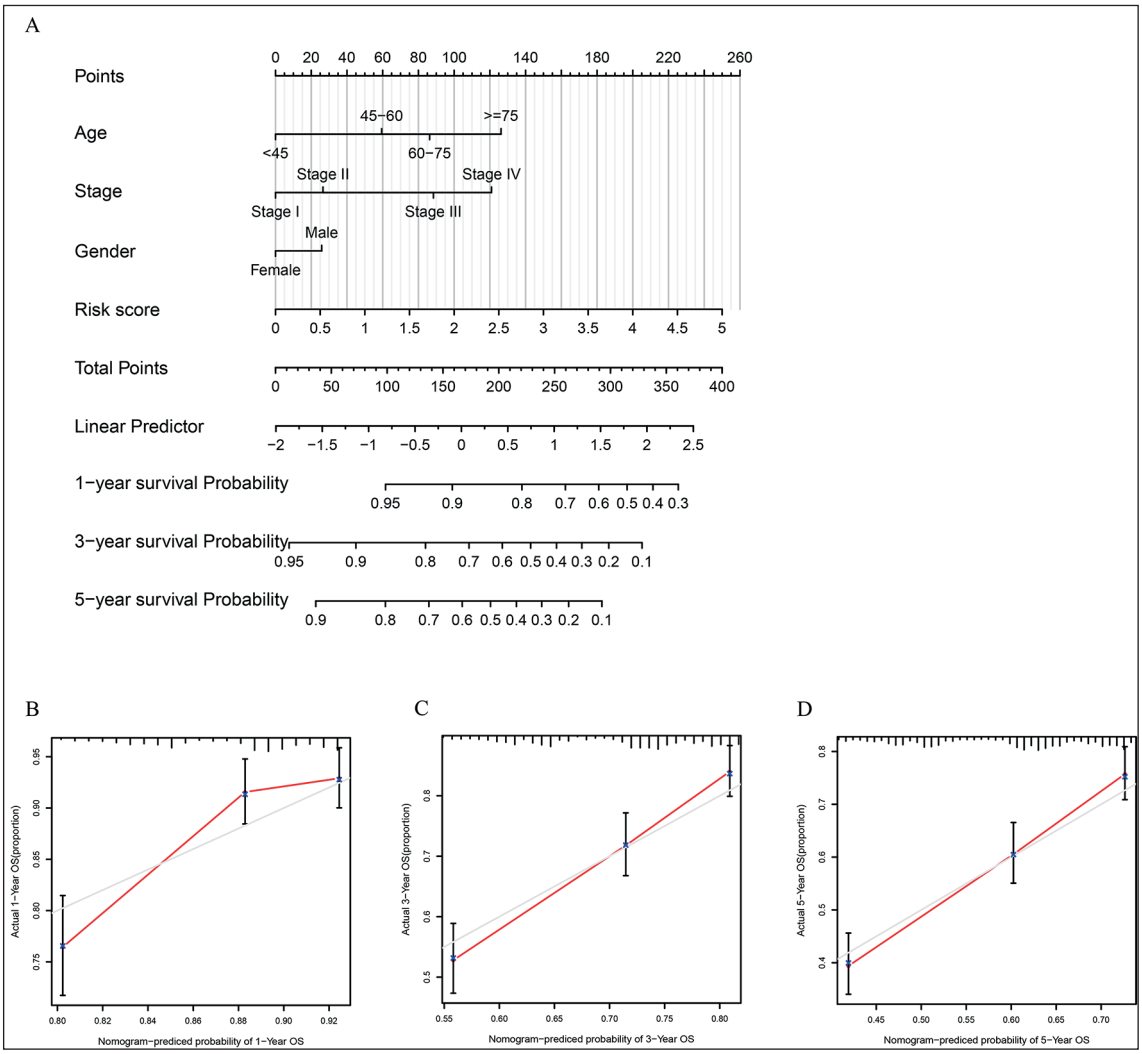

Figure 4. Construction and evaluation of a prognostic nomogram. **A,** The nomogram predicts the probability of 1-, 3-, and 5-year OS. **B–D,** The calibration plot of the nomogram predicts the probability of 1-, 3-, and 5-year OS.

dent setting. Cell proliferation-related pathways such as E2F targets, G2M checkpoint, and MYC targets were found to be enriched in the high-risk group, confirming the fact that TRLGs are essential regulators of cell growth and the cell cycle (Figure 5A, B). Some metabolic-related pathways, for example fatty acid metabolism and histidine metabolism, were mostly grouped in the low-risk score subgroup, which was consistent with previous observations that RAP1 inhibits fat accumulation[35] (Figure 5A, B). Because accumulating evidence indicated that the TRLG-based model was also associated with tumor immunity[19-21,34], we next focused on immune pathways. Intriguingly, we found a substantial number of immune-related pathways that were enriched in the two clusters; these pathways included inflammatory response, allograft rejection, interferon gamma response, and IL6-JAK-STAT3 signaling (Figure 5A, B).

Estimation of the Tumor Immune Microenvironment and Cancer Immunotherapy Response via the TRLG Risk Model

To better understand the molecular link between our TRLG-based model and tumor immunity, we calculated the correlations of the TRLGs with a panel of immunomodulators that play crucial roles in immunotherapy[27,28]. We observed that the expression of most immunomodulators was highly correlated with the expression of TRLGs (Figure 6C). By examining the correlations be-

tween the RLGs-based model and numbers of tumor infiltrating immune cells, an ssGSEA algorithm was adopted (**Supplementary Table IX**). The expression of most TRLGs was negatively

Figure 5. KEGG gene set and cancer-related gene set enrichment analysis of low and high risk groups. **A,** Heat maps of GSVA enrichment results showing the activation status of biological pathways under different risk groups. **B,** A bar graph showing the results of GSVA enrichment of cancer-related gene sets under different risk groups. LogFC > 0 indicates that the gene set was enriched in the high-risk group.

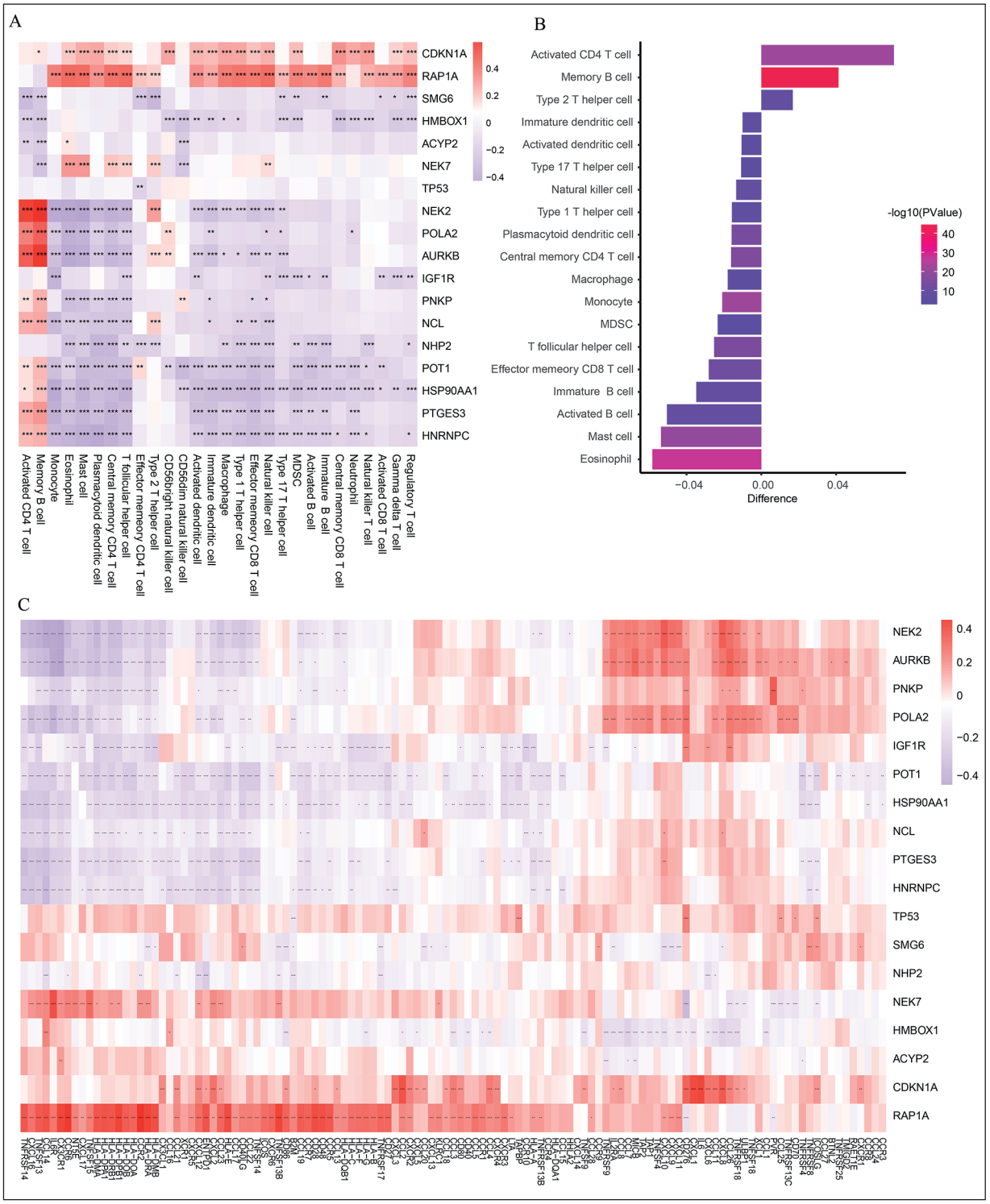

Figure 6. The tumor microenvironment status of GEO non-small cell lung cancer. **A,** Correlation analysis of expression of telomere-related genes and immunomodulator-related genes. **B,** Comparison of immune cell infiltration in different risk groups. **C,** Correlation between a patient's telomere-related gene expression and amount of mmune cell infiltration.

correlated with the level of immune cell infiltration (Figure 6A). We also found that immune activated cells, such as central memory CD4 T cells, effector memory CD8 T cells, and activated B cells, were clustered in the low-risk group (Figure 6B, **Supplementary Table X**), while the layout

of the high-risk group showed an immunosuppression pattern. When taken together, these data indicated that TRLGs were notably associated with cancer immunity, including the regulation of immunomodulators and the numbers of tumor infiltrating immune cells in NSCLC.

A major effort has been made to identify biomarkers (e.g., TMB and PD-L1 protein levels) that can be used to predict a patient's response to immunotherapy[36-38]. Considering that the TRLGs-based model in this study appeared to be associated with the immune microenvironment of the tumor, we examined the ability of the TRLGs-based model to predict the response of patients to ICB therapy. To achieve this goal, the TIDE algorithm was used to predict the response of NSCLC patients in the TCGA cohort to immunotherapy. The results showed that the TMB value of the high-risk group was higher than that of the low-risk group, and the prediction of immunotherapy effect in the high-risk group was also better than that in the low-risk group (Figure 7A, B). We also verified our model in another type of cancer. Two immunotherapy cohorts (the IMvigor210 cohort[23] and the bladder cancer cohort[24]) were retrieved from the public resource. We found that patients in the low-risk group exhibited significant clinical benefits and had a markedly prolonged overall survival time in both anti-PD-L1 cohorts (Figure 8A, B). The 348 patients in the IMvigor210 cohort showed different degrees of response to a PD-L1 blocker, including complete response, partial response, stable disease, and progressive disease. As expected, the complete response patients had much lower risk scores than patients with other types of responses (Figure 8C). A Fisher test performed on the two subgroups also revealed a distinctly better therapeutic response in the low-risk group (Figure 8D). In addition, the TMBs and neoantigen burdens in the low-risk group greatly exceeded those in the high-risk group (Figure 8E, F), which may partially explain the survival advantage and greater benefit of ICB treatment in the low-risk group. In summary, the collective data suggested that the established risk model could serve as indicator for predicting a patient's response to anti-PD-L1 immunotherapy.

Identification of Potential Compounds Targeting the TRLG Risk Model

To search for candidate compounds targeting the candidate model, we used the pRRophetic algorithm to predict the IC_{50} values of compounds obtained from the Genomics of Drug Sensitivity in Cancer (GDSC) website in the GEO NSCLC cohort. Based on the GDSC database, we identified 110 highly relevant chemicals between the high-risk and low-risk groups. The high-risk group was more sensitive to 53 of the compounds, while the low-risk group was more sensitive to the other compounds **(Supplementary Table XI)**. Figure 9 shows the top 10 compounds that

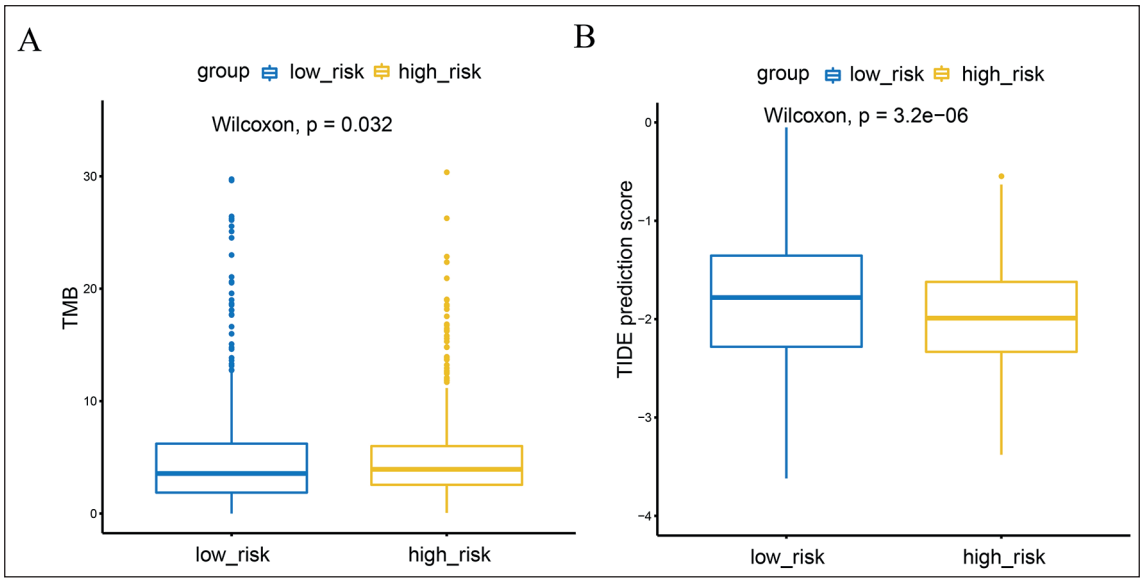

Figure 7. Prediction of risk score for immunotherapy response in the TCGA cohort. **A,** Comparison of TMB scores between high and low risk groups of patients in the TCGA cohort. **B,** Comparison of TIDE prediction scores between high and low risk groups of patients in TCGA.

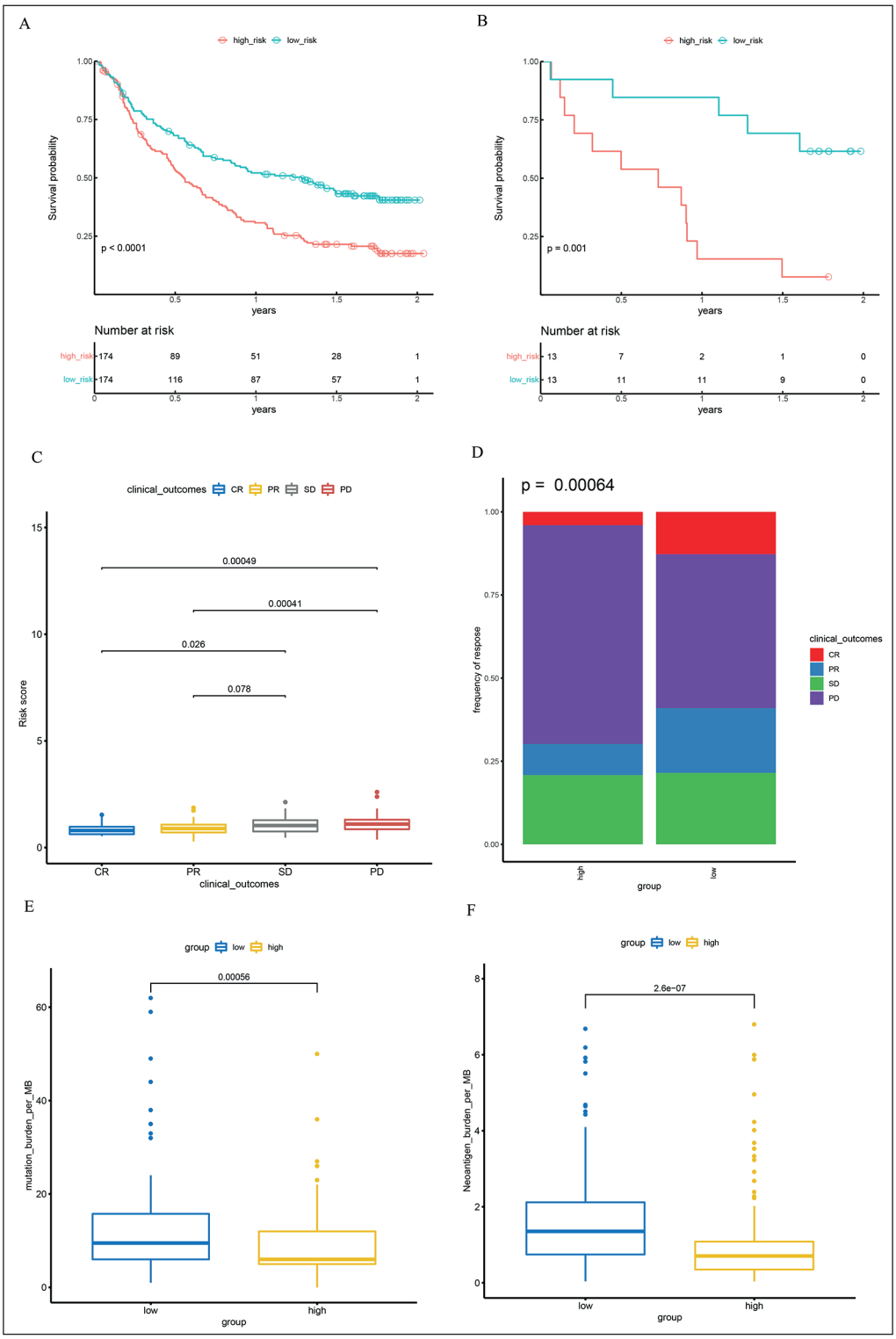

Figure 8. Prediction of risk scores for immunotherapy response in other cancer patients. **A-B,** Kaplan-Meier survival curve for patients in the IMvigor210 and anti-PD-L1 groups after treatment with PD-L1 blockade immunotherapy. **C,** Comparison of risk scores among different immunotherapy response groups. **D,** The proportion of patients in the IMvigor210 high- and low-risk groups with different immunotherapy responses. Fisher's test was used to determine the significance of the difference between the two groups. Boxes indicate the median ± 1 quartile, with the whiskers extending from the hinge to the smallest or largest value within 1.5× IQR from the box boundaries. **E,** Comparison of TMB scores between the high and low risk groups of patients in the IMvigor210 cohort. **F,** Comparison of neoantigen burden scores in the high and low risk groups of patients in the IMvigor210 cohort.

might be further studied in NSCLC patients in the high-risk group. These results suggest that the candidate risk model is associated with drug sensitivity, and therefore the risk score might be a promising biomarker for establishing appropriate treatment strategies.

Discussion

Increasing evidence shows that an altered TL in cancer cells might enable the cells to metastasize and cause recurrent disease, and as a consequence, could be a predictor of clinical outcome[39-41]. Thus, a comprehensive analysis of TL regulators is essential for understanding tumorigenesis and designing a telomere-based therapy. Although some evidence suggests that a single TL regulator is related to a patient's prognosis[42,43] and cancer immunity[19-21,34,43], the relationship between 118 TL regulators and OS and TME has not been thoroughly analyzed in a case of cancer. By mining the TCGA and GEO profiling data, we constructed and validated a robust 18-gene prognostic model on the basis of TRLG signatures via a set of bioinformatics. The result revealed that the model could not only predict the prognosis of patients with NSCLC, but also serve as an indicator of immunotherapeutic response.

In this project, we first identified 80 DE-TRLGs based on the GEO database, and then confirmed the prognostic ability of 49 genes. A number of tools has been introduced to narrow down the candidate genes[44]. In this study, LASSO with high specificity and widely application was applied to identify the candidates. Consequently, 18 cgenes were used to develop a TRLGs-related risk model to predict the OS of patients with NSCLC. Subsequently, all the patients were assigned to high- and low-risk groups based on their intermediate risk score, and the high-risk subgroup was closely associated with a worse prognosis (Figure 2A-D, **Supplementary Table V**). To further assess the prognostic value of the candidate model, a multivariate Cox regression analysis was performed. Those results indicated that the 18-gene signature was an autocephalous risk factor for OS (Figure 3A-B). An ROC analysis demonstrated that the risk score of this model was superior to conventional clinical features in predicting the prognosis of NSCLC. Importantly, the information derived from an established a nomogram showed excellent agreement between the observed and predicted rates of 1-year, 3-year, and 5-year OS (Figure 4A-D). Last but not least, the prognostic power of the 18-gene signature was externally validated in the TCGA lung cohorts. All the outcomes showed that the candidate risk model independently associated with OS was fairly accurate and could serve as a novel biomarker in further investigations.

In our functional analysis, 18 DE-TRLGs were enriched in cell cycle- and proliferation-related pathways (Figure 5A-B). Genes associated with E2F targets, G2M checkpoint, and Myc targets were positively correlated with the DE-TRLGs.

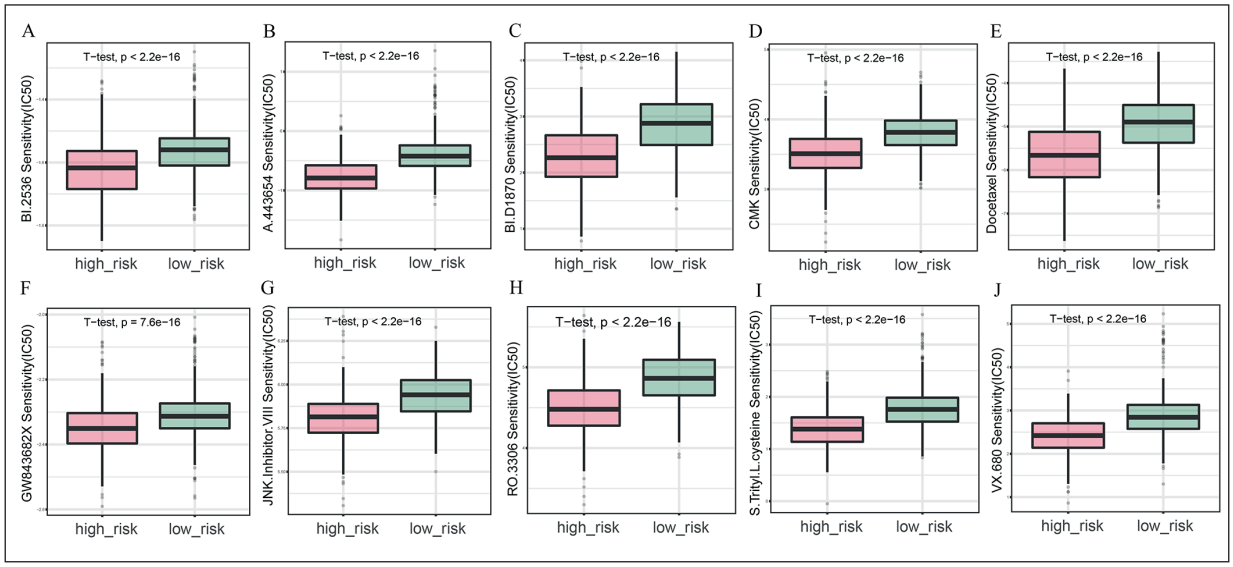

Figure 9. Identification of novel candidate compounds targeting the risk model.

We found that 18 DE-TRLGs worked synergistically with these proliferation-related genes to ensure proper cell growth, suggesting that 18 DE-TRLGs were needed for the rapid division of cancer cells. Besides cell-cycle-related pathways, negative correlations were found between the 18 DE-TRLGs and immunological signatures, suggesting that 18 DE-TRLGs played crucial roles in tumor immunity. Lastly, we also found that several metabolic-related pathways including fatty acid metabolism and histidine metabolism were mostly concentrated in the low-risk subgroup, which was in agreement with previous research showing that RAP1 could function as a metabolic regulator by suppressing fat accumulation[35].

A recent study reported that the level of telomere-binding protein RAP1 was predictive of the success of chemotherapy in breast and colon cancer[45]. Our study showed that the risk model based on 18 DE-TRLGs was highly correlated with the TME in terms of immunomodulators and immune cells (Figure 6A-C). Therefore, we speculated that the 18-DE-TRLG signature could be regarded as an adequate marker for predicting a patient's response to cancer therapy, including chemotherapy, targeted therapies, and immunotherapy. To verify this hypothesis, a comprehensive analysis of potential therapeutic effects on patients in the high- and low-risk groups was performed. We found that the 18-DE-TRLG signature was associated with sensitivity to drugs. Additionally, we also found that the risk model could predict the response of patients to anti-PD-L1 immunotherapy (Figure 8C). By identifying different degrees of response to a PD-L1 blocker, we concluded that the patients in the low-risk subgroup would display a favorable response in the IMvigor210 cohort. These data indicated that the model based on an 18-DE-TRLG signature could be utilized to predict a patient's response to treatment with a PD-L1 blocker.

Conclusions

We systematically analyzed an 18-DE-TRLG signature which could affect the TME and predict the prognosis of patients with NSCLC. In addition, the newly developed risk model also predicted the therapeutic utility of telomere-targeted therapy and immunotherapy. This project verified the clinical significance of the telomere system and may facilitate the development of personalized immune-based therapeutic strategies for NSCLC patients.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Availability of data and materials
All data collected and analyzed in this study are available from the corresponding author on reasonable request.

Conflicts of Interest
The authors declare no conflicts of interest.

Funding
Not applicable.

Author contributions
Research design: Ruijun Cai. Data acquisition: Shijie Mai. Data analysis: Xiguang Liu and Mei Li. Manuscript writing: Xiguang Liu and Mei Li. Manuscript revising: Ruijun Cai and Shijie Mai. All authors read and approved the final submitted version.

References

1) Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65: 87-108.
2) Molina JR, Yang P, Cassivi SD, Schild SE, Adjei AA. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. Mayo Clin Proc 2008; 83: 584-594.
3) Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin 2018; 68: 7-30.
4) Marijon H, Bouyon A, Vignot S, Besse B. Prognostic and predictive factors in lung cancer. Bull Cancer 2009; 96: 391-404.
5) Cuyún Carter G, Barrett AM, Kaye JA, Liepa AM, Winfree KB, John WJ. A comprehensive review of nongenetic prognostic and predictive factors influencing the heterogeneity of outcomes in advanced non-small-cell lung cancer. Cancer Manag Res 2014; 6: 437-449.
6) Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: a decade of progress. Chest 2002; 122: 1037-1057.
7) Gansner JM, Rosas IO. Telomeres in lung disease. Transl Res 2013; 162: 343-352.

8) Hou L, Zhang X, Gawron AJ, Liu J. Surrogate tissue telomere length and cancer risk: shorter or longer? Cancer Lett 2012; 319: 130-135.
9) Svenson U, Roos G. Telomere length as a biological marker in malignancy. Biochim Biophys Acta 2009; 1792: 317-323.
10) Heaphy CM, Meeker AK. The potential utility of telomere-related markers for cancer diagnosis. J Cell Mol Med 2011; 15: 1227-1238.
11) Wu X, Amos CI, Zhu Y, Zhao H, Grossman BH, Shay JW, Luo S, Hong WK, Spitz MR. Telomere dysfunction: a potential cancer predisposition factor. J Natl Cancer Inst 2003; 95: 1211-1218.
12) Jang JS, Choi YY, Lee WK, Choi JE, Cha SI, Kim YJ, Kim CH, Kam S, Jung TH, Park JY. Telomere length and the risk of lung cancer. Cancer Sci 2008; 99: 1385-1389.
13) Seow WJ, Cawthon RM, Purdue MP, Hu W, Gao YT, Huang WY, Weinstein SJ, Ji BT, Virtamo J, Hosgood HD, 3rd, Bassig BA, Shu XO, Cai Q, Xiang YB, Min S, Chow WH, Berndt SI, Kim C, Lim U, Albanes D, Caporaso NE, Chanock S, Zheng W, Rothman N, Lan Q. Telomere length in white blood cell DNA and lung cancer: a pooled analysis of three prospective cohorts. Cancer Res 2014; 74: 4090-4098.
14) Shen M, Cawthon R, Rothman N, Weinstein SJ, Virtamo J, Hosgood HD, 3rd, Hu W, Lim U, Albanes D, Lan Q. A prospective study of telomere length measured by monochrome multiplex quantitative PCR and risk of lung cancer. Lung Cancer 2011; 73: 133-137.
15) Lan Q, Cawthon R, Gao Y, Hu W, Hosgood HD, 3rd, Barone-Adesi F, Ji BT, Bassig B, Chow WH, Shu X, Cai Q, Xiang Y, Berndt S, Kim C, Chanock S, Zheng W, Rothman N. Longer telomere length in peripheral white blood cells is associated with risk of lung cancer and the rs2736100 (CLPTM1L-TERT) polymorphism in a prospective cohort study among women in China. PLoS One 2013; 8: e59230.
16) Fernández-Marcelo T, Gómez A, Pascua I, de Juan C, Head J, Hernando F, Jarabo JR, Calatayud J, Torres-García AJ, Iniesta P. Telomere length and telomerase activity in non-small cell lung cancer prognosis: clinical usefulness of a specific telomere status. J Exp Clin Cancer Res 2015; 34: 78-83.
17) Frías C, García-Aranda C, De Juan C, Morán A, Ortega P, Gómez A, Hernando F, López-Asenjo JA, Torres AJ, Benito M, Iniesta P. Telomere shortening is associated with poor prognosis and telomerase activity correlates with DNA repair impairment in non-small cell lung cancer. Lung Cancer 2008; 60: 416-425.
18) Hug N, Lingner J. Telomere length homeostasis. Chromosoma 2006; 115: 413-425.
19) Effros RB. Telomere/telomerase dynamics within the human immune system: effect of chronic infection and stress. Exp Gerontol 2011; 46: 135-140.
20) Chen Y, Qu F, He X, Bao G, Liu X, Wan S, Xing J. Short leukocyte telomere length predicts poor prognosis and indicates altered immune functions in colorectal cancer patients. Ann Oncol 2014; 25: 869-876.
21) Qu F, Li R, He X, Li Q, Xie S, Gong L, Ji G, Lu J, Bao G. Short telomere length in peripheral blood leukocyte predicts poor prognosis and indicates an immunosuppressive phenotype in gastric cancer patients. Mol Oncol 2015; 9: 727-739.
22) Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. Bioinformatics 2012; 28: 882-883.
23) Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, Kadel EE, III, Koeppen H, Astarita JL, Cubas R, Jhunjhunwala S, Banchereau R, Yang Y, Guan Y, Chalouni C, Ziai J, Şenbabaoğlu Y, Santoro S, Sheinson D, Hung J, Giltnane JM, Pierce AA, Mesh K, Lianoglou S, Riegler J, Carano RAD, Eriksson P, Höglund M, Somarriba L, Halligan DL, van der Heijden MS, Loriot Y, Rosenberg JE, Fong L, Mellman I, Chen DS, Green M, Derleth C, Fine GD, Hegde PS, Bourgon R, Powles T. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature 2018; 554: 544-548.
24) Snyder A, Nathanson T. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. 2017; 14: e1002309.
25) Sun H, Kim P, Jia P, Park AK, Liang H, Zhao Z. Distinct telomere length and molecular signatures in seminoma and non-seminoma of testicular germ cell tumor. Brief Bioinform 2019; 20: 1502-1512.
26) Lin JZ, Lin N, Zhao WJ. Identification and validation of a six-lncRNA prognostic signature with its ceRNA networks and candidate drugs in lower-grade gliomas. Genomics 2020; 112: 2990-3002.
27) Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, Porta-Pardo E, Gao GF, Plaisier CL, Eddy JA, Ziv E, Culhane AC, Paull EO, Sivakumar IKA, Gentles AJ, Malhotra R, Farshidfar F, Colaprico A, Parker JS, Mose LE, Vo NS, Liu J, Liu Y, Rader J, Dhankani V, Reynolds SM, Bowlby R, Califano A, Cherniack AD, Anastassiou D, Bedognetti D, Mokrab Y, Newman AM, Rao A, Chen K, Krasnitz A, Hu H, Malta TM, Noushmehr H, Pedamallu CS, Bullman S, Ojesina AI, Lamb A, Zhou W, Shen H, Choueiri TK, Weinstein JN, Guinney J, Saltz J, Holt RA, Rabkin CS, Lazar AJ, Serody JS, Demicco EG, Disis ML, Vincent BG, Shmulevich I. The Immune Landscape of Cancer. Immunity 2018; 48: 812-830.e814.
28) Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. Cell Rep 2017; 18: 248-262.
29) Wu Z, Wang M, Liu Q, Liu Y, Zhu K, Chen L, Guo H, Li Y, Shi B. Identification of gene expression

profiles and immune cell infiltration signatures between low and high tumor mutation burden groups in bladder cancer. Int J Med Sci 2020; 17: 89-96.

30) Xu F, Zhan X, Zheng X, Xu H, Li Y, Huang X, Lin L, Chen Y. A signature of immune-related gene pairs predicts oncologic outcomes and response to immunotherapy in lung adenocarcinoma. Genomics 2020; 112: 4675-4683.

31) Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-d613.

32) Martínez P, Blasco MA. Telomeric and extra-telomeric roles for telomerase and the telomere-binding proteins. Nat Rev Cancer 2011; 11: 161-176.

33) El Maï M, Wagner KD, Michiels JF, Ambrosetti D, Borderie A, Destree S, Renault V, Djerbi N, Giraud-Panis MJ, Gilson E, Wagner N. The telomeric protein TRF2 regulates angiogenesis by binding and activating the PDGFRβ promoter. Cell Rep 2014; 9: 1047-1060.

34) Cherfils-Vicini J, Iltis C, Cervera L, Pisano S, Croce O, Sadouni N, Győrffy B, Collet R, Renault VM, Rey-Millet M, Leonetti C, Zizza P. Cancer cells induce immune escape via glycocalyx changes controlled by the telomeric protein TRF2. EMBO J 2019; 38: e100012.

35) Yeung F, Ramírez CM, Mateos-Gomez PA, Pinzaru A, Ceccarini G, Kabir S, Fernández-Hernando C, Sfeir A. Nontelomeric role for Rap1 in regulating metabolism and protecting against obesity. Cell Rep 2013; 3: 1847-1856.

36) Gibney GT, Weiner LM, Atkins MB. Predictive biomarkers for checkpoint inhibitor-based immunotherapy. Lancet Oncol 2016; 17: e542-e551.

37) Miao D, Margolis CA. Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. Science 2018; 359: 801-806.

38) Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, Sucker A, Hillen U, Foppen MHG, Goldinger SM, Utikal J, Hassel JC, Weide B, Kaehler KC, Loquai C, Mohr P, Gutzmer R, Dummer R, Gabriel S, Wu CJ, Schadendorf D, Garraway LA. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 2015; 350: 207-211.

39) Bisoffi M, Heaphy CM, Griffith JK. Telomeres: prognostic markers for solid tumors. Int J Cancer 2006; 119: 2255-2260.

40) La Torre D, Aguennouz M, Conti A, Giusa M, Raffa G, Abbritti RV, Germano A, Angileri FF. Potential clinical role of telomere length in human glioblastoma. Transl Med UniSa 2011; 1: 243-270.

41) Garcia-Aranda C, de Juan C, Diaz-Lopez A, Sanchez-Pernaute A, Torres AJ, Diaz-Rubio E, Balibrea JL, Benito M, Iniesta P. Correlations of telomere length, telomerase activity, and telomeric-repeat binding factor 1 expression in colorectal carcinoma. Cancer 2006; 106: 541-551.

42) Arimura-Omori M, Kiyohara C, Yanagihara T, Yamamoto Y, Ogata-Suetsugu S, Harada E, Hamada N, Tsuda T, Takata S, Shimabukuro I, Nagata N, Yatera K, Torii R, Okamoto M, Fujita M, Nakanishi Y. Association between Telomere-Related Polymorphisms and the Risk of IPF and COPD as a precursor lesion of lung cancer: findings from the Fukuoka Tobacco-Related Lung Disease (FOLD) registry. Asian Pac J Cancer Prev 2020; 21: 667-673.

43) Luo Z, Liu W, Sun P, Wang F, Feng X. Pan-cancer analyses reveal regulation and clinical outcome association of the shelterin complex in cancer. Brief Bioinform 2021; 22: bbaa441.

44) Zhang W, Wan YW, Allen GI, Pang K, Anderson ML, Liu Z. Molecular pathway identification using biological network-regularized logistic models. BMC Genomics 2013; 14 Suppl 8: S7.

45) Khattar E, Maung KZY, Chew CL, Ghosh A. Rap1 regulates hematopoietic stem cell survival and affects oncogenesis and response to chemotherapy. Nat Commun 2019; 10: 5349-5362.