Stem Cell-Related Prognostic Signature for Lung Adenocarcinoma

Zhanghao Huang
Affiliated Hospital of Nantong University
Muqi Shi
Nantong University
Hao Zhou
Affiliated Hospital of Nantong University
Jinjie Wang
Affiliated Hospital of Nantong University
You Lang Zhou
Affiliated Hospital of Nantong University
Hai-Jian Zhang
Affiliated Hospital of Nantong University
Jia-Hai Shi (sjh@ntu.edu.cn)
https://orcid.org/0000-0002-3223-8625

Research article

Keywords: stem cell, prognosis, signature, TCGA, GEO

Posted Date: July 16th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-41534/v1

Abstract

Background: Adenocarcinoma is characterized by high infiltration and negative growth, which is easy to invade the walls of blood vessels and lymphatic vessels. The function of the stem cell population is to control and maintain cell regeneration. Therefore, it is necessary to study the prognostic value of stem cells in LUAD.

Methods: Stem cells signature construction relies on the univariate, least absolute shrinkage operator (LASSO) and multivariate Cox regression analysis. Risk plot, Kaplan-Meier analysis and the area under ROC (AUC) are verified the accuracy of the signature in GEO (GSE20319 and GSE42127) and TCGA cohort respectively. What's more, to further improve persuasiveness, we conducted nomogram, drug efficacy analysis, immune infiltration analysis, and compared the relationship between mRNA stem cell index (mRNAsi) and signature.

Result: The patients were divided into two groups via the cutoff of risk score; it can be seen that the low-risk group has better survival. The robust signature that contains ten stem cells is independent prognostic factors for LUAD (C6orf62, DNER, NELL2, LATS2, LGR5, PTPRO, LRIG1, PABPC1, NT5E and SET). The nomogram showed that 1-year (0.805), 3-year (0.773), and 5-year (0.765) survival rates of the signature we constructed were all greater than 0.7, indicating that our signature was very feasible.

Conclusion: The signature can be used as a reliable and convenient tool for lung adenocarcinoma.

1. Background

Lung adenocarcinoma refers to a malignant tumor originating from lung epithelial tissue, which is a type of non-small cell lung cancer. In recent years, the incidence rate has gradually increased. In addition, due to the limitations of diagnosis and treatment, the mortality rate of LUAD ranks first in malignant tumors [1]. Tumor stem cells refer to cells that have self-renewal ability and can produce heterogeneous tumor cells, which play a significant role in tumor survival, proliferation, metastasis and recurrence [2, 3]. The ability of tumor stem cells to move and migrate makes tumor cells migration possible, at the same time, cancer stem cells can stay dormant for a long time and have a variety of drug-resistant molecules, but are not sensitive to external physical and chemical factors that kill tumor cells, which leads to the result that tumors often relapse after conventional cancer treatment eliminates most common tumor cells [4, 5].

The development of LUAD treatment plan and survival period are affected by many factors, but the TNM stage of tissue cells may be one of significant factor in determining treatment plan and estimating prognosis. TNM stage is based on anatomy and is a description of the cumulative range of tumors. However, it should be emphasized that the TNM stage also has shortcomings including the uneven source of case data and the relatively complicated stage of N. With the gradual development of diagnosis and treatment technology, we found that molecular markers have a greater prognosis for patients. Studying the genetic functions and pathways of LUAD could contribute to establish prognostic markers and therapeutic targets, which could accurately and comprehensively predict the prognosis of LUAD [6].

Therefore, the ideal of using cancer stem cell therapy provides a new direction for the diagnosis and treatment of LUAD and the regulatory mechanism of stem cells at the molecule level requires further digging.

In this research, we constructed a stem cell signature of 10 genes as a prognostic target for lung adenocarcinoma. Meanwhile, we analyzed the types of immune cells in LUAD to understand the connection between stem cells and the immune microenvironment.

2. Materials And Methods

2.1 Data Acquisition and Selection

The RNA-sequencing and clinical traits information of LUAD were obtained from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov) and Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.gov/geo/) that were served as training cohort and validation cohort, respectively. After classification and regularization, there were 54 normal samples and 497 tumor samples. At the same time, when merging clinical information, missing and incomplete samples were deleted. Besides, 166 tumor stem cells were downloaded from the cancerSEA database [7] to prepare for further signature construction. GSE30219 [8] was conducted by GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array). GSE42127 [9] was conducted by GPL6884 (Illumina HumanWG-6 v3.0 expression bead chip).

2.2 Signature Construction and Verification

It was worth emphasizing that the R package was an indispensable key tool for us to construct and verify a signature next. The signature was established by a two-step method, the first step was LASSO COX regression, and the second step was multivariate COX regression. Patients were divided into low-risk and high-risk groups based on the cutoff of risk score, which was calculated by formula as follows: HR 1 × gene 1 expression + HR 2 × gene 2 expression … + HR n × gene n expression [10]. In the TCGA and GEO cohorts, the risk curve was drawn to describe further the relationship between the patients' risk value and survival states and protein expression, the Kaplan-Meier curve and ROC curve were used to verify the reliability of the signature [11]. What's more, to further clarify the relationship and comparison between clinical factors and signature, subgroup analysis and nomogram analysis were conducted.

2.3 Gene Set Enrichment Analysis

GSEA is a method used to evaluate the distribution trend of genes in the gene list sorted by phenotype correlation and to understand gene positioning, function and biological significance. The molecular tag database is constructed, which contains multiple functional gene sets. By analyzing the gene expression data, it is obtained whether the expression is significantly enriched in a certain function. We presented the GO term and the KEGG pathway of the stem cell signature to analyze further it's possible biological functions [12]. The number of permutations was set to 1000, and our selection criteria are closely related to a nominal P-value ($P < 0.05$).

2.4 Therapeutic Efficacy and Signature

Some patients from TCGA recorded the results of the evaluation of the efficacy after the first treatment of the radiotherapy and chemotherapy, which also provided a direction for us to verify the reliability of the signature in term of efficacy. According to Response Evaluation Criteria in Solid Tumors (RECIST) and risk score, this part of patients was classified to compare whether there were differences between different therapeutic effect [13].

2.5 mRNAsi and Signature

In recent years, literature has proposed the concept of stemness indices (mRNAsi), which was calculated by a predictive model with an OCLR algorithm based on pluripotent stem cell samples from the Progenitor Cell Biology Consortium dataset (https://bioinformaticfmrp.github.io/PanCanStem_Web/). Specifically, the Spearman correlation algorithm (RNA expression data) contributed to the stem index model to score LUAD samples in the TCGA dataset. The stem indices were then mapped to the [0, 1] range by using a linear transformation that subtracted the minimum and divide by the maximum. The index is closer to 1, which indicated that the cell differentiation was worse, and the characteristic of stem cells was stronger. We merged mRNAsi into our signature to compare whether there was a difference between low- and high- risk groups [14].

2.6 Immune Infiltration Analysis

TIMER database, providing six types of immune cell infiltration and using RNA-Seq expression profiling data to detect immune cell infiltration in tumor tissue, was used to appraise potential links between risk grouping and tumor-infiltrating immune cells (TIICs). Deconvolution is a newly released statistical method that allows TIMER to infer the incidence of TICC from gene expression profiles. CIBERSORT (http://cistrome.shinyappes), a deconvolution algorithm, can estimate the cell composition of complex tissues based on standardized gene expression data, and the method can be used to energize specific cell types. With CIBERSORT, we can visualize the composition of immune cells in tumor samples of LUAD, and standard annotation files established gene expression datasets. P-value ($P < 0.05$) was a significant criterion to determine the type of immune cells affected by grouping [15].

3. Result

3.1 Construction of Signature

All cancer stem cells were downloaded from CancerSEA (http://biocc.hrbmu.edu.cn/CancerSEA/home.jsp), which involved 14 functional states of 41900 single cancer cells from 25 cancer types. Twenty-eight CSCs associated with OS ($P < 0.05$) was measured as predictive CSCs for LASSO analysis (Fig. 1A-B). Through multivariate COX regression, we select ten CSCs to construct a robust signature for LUAD (Table 1). The calculation formula of the risk score is as follows: risk score = 0.578 × expression C6orf62 + 1.24 × expression DNER + 0.737 × expression NELL2 + 1.404 ×

expression LATS2 + 1.202 × expression LGR5 + 0.676 × expression PTPRO + 0.718 × expression LRIG1 + 1.306 × expression PABPC1 + 1.126 × expression NT5E + 1.458 × expression SET. According to the cutoff of risk scores, patients in TCGA were divided into low-risk group and high-risk group [16]. The area under the ROC curve for 1, 3, 5-year were 0.771, 0.734, 0.687 (Fig. 1C). The survival analysis suggested that the overall survival rate of the low-risk group was higher than that of the high-risk group ($P < 0.001$). The 5-year survival rate of the low-risk group was close to 50%, while the 5-year survival rate of the high-risk group was only 20% (Fig. 1D). The box plot displayed the stem cell contained in the signature in the normal tissues and tumor tissues. (Fig. 1E). The risk curve can clearly show the relationship between survival status, survival time and expression of CSCs and risk score (Fig. 1F) [17].

Table 1
Independent factors in the signature.

id	coef	HR	HR.95L	HR.95H	pvalue
C6orf62	-0.73221	0.480847	0.312891	0.738959	0.000838
DNER	0.195561	1.215993	1.073065	1.377959	0.002174
NELL2	-0.34583	0.707635	0.533484	0.938635	0.016425
LATS2	0.38125	1.464113	1.100657	1.947588	0.008826
LGR5	0.251941	1.28652	1.076618	1.537345	0.005566
PTPRO	-0.44697	0.639564	0.430824	0.94944	0.026599
LRIG1	-0.26479	0.767366	0.633265	0.929864	0.006893
PABPC1	0.246554	1.279608	0.983828	1.664311	0.066004
NT5E	0.132706	1.141914	1.006795	1.295166	0.038888
SET	0.414622	1.513798	1.03377	2.216727	0.033119

3.2 Validation of the Signature in TCGA

The univariate Cox regression showed factors related to prognosis like a stage, T, M, N and risk score ($P < 0.05$), while multivariate Cox regression showed that only stage and risk score were significant independent risk factor of LUAD. Compared with other clinical factors, the area under the ROC curve of the signature in each period was the largest, which implied that compared with other clinical factors, the predictive ability of the gene signature we constructed was optimal (Fig. 2A-B). The area under the ROC curve (AUC). The area under the ROC curve for 1-year, 3-year and 5-year OS were 0.771, 0.734 and 0.687, which implied that our signature had excellent predictive power (Fig. 2C-E) [18, 19]. We conducted a hierarchical analysis to clarify the link between subgroups and risk grouping. A further conclusion was drawn that all subgroups except N3 could identify high-risk and low-risk groups. In N3, there were only two patients, both of which belonged to the high-risk group. And in most subgroups, high- and low-risk groups had significant differences, such as age < = 65, age > 65, female, male, stage III, T2, T3, N0, N2 and

M0 (P < 0.05) (Fig. 3). What's more, the relationship between genes in the signature and various clinical factors was also clearly revealed through clinical correlation analysis. P-value < 0.05 was our criterion to judge whether it was meaningful (Fig. 4) [20]. We constructed nomogram to predict the score of the signature and clinical factors, the correction curve showed the predicted value, and the 45-degree line represented the actual survival result. The 45-degree lines of the calibration chart were close to 45 degrees in 1-, 3-, 5-year survival probability. The 1-year AUC was 0.805; the 3-year AUC was 0.773, the 5-year AUC was 0.765, which indicated that our signature was very reliable (Fig. 5A) [21].

3.3 Gene Set Enrichment Analysis

The biological characteristics of the signature were confirmed by the analysis of GO term and KEGG pathway. In GO term annotation, five categories were positively associated with the low-risk group, which were hexose catabolic process, NADH metabolic process, monosaccharide catabolic process, ATP generation from ADP and NAD metabolic process. At the same time, five categories were negatively related to the low-risk group, which were negative regulation of adaptive immune response, regulation of tumor necrosis factor biosynthetic process, bile acid metabolic process, positive regulation of tyrosine phosphorylation of STAT5 protein and regulation of type 2 immune response. In the KEGG pathway, there were five pathways were positively associated with the low-risk group, such as ECM receptor interaction, focal adhesion, glycosphingogolipid biosynthesis latco and neolatco series, pentose phosphate pathway and P53 signaling pathway. While there were five pathways were negatively related to the low-risk group, like JAK start signaling pathway, primary immunodeficiency, VEGF signaling pathway, B cell receptor signaling pathway and T cell receptor signaling pathway (Fig. 5B) [22].

3.4 Validation of the Signature in GEO

In order to further verify the feasibility of the gene signature, we verified through the GEO database. In GSE20319 and GSE42127, the relationship between survival status, survival time and the expression of the CSCs and risk score were consistent with the conclusion in TCGA. The survival analysis in GSE30219 and GSE42127 revealed that the overall survival rate of the low-risk group was significantly better than that of the high-risk group (P < 0.05). Besides, in GSE30219, the area under the ROC curve was 0.826, 0.638 and 0.599 in 1-year, 3-year and 5-year survival rates, respectively. One of the more worthwhile was that in GSE42127, the area under the ROC curve was 0.788, 0.657 and 0.582. This series of external verification fully demonstrated the feasibility and accuracy of our signature (Fig. 6) [23].

3.5 Therapeutic Efficacy Analysis

In the TCGA database, 126 patients recorded the results of the first treatment after radiotherapy and chemotherapy. At the same time, we tracked the evaluation of efficacy, 111 cases were complete response (CR), only one case was the partial response (PR), eight cases were stable disease (SD), and seven cases were progressive disease (PD). The three CSCs in the signature had significant differences in the efficacy of the different drug (P < 0.05). And it was worth noting that our signature had a certain meaning. P-value was 0.052, very close to 0.05 (Fig. 7A) [24, 25].

3.6 Relationship between Signature and mRNAsi

There were already clear articles that calculated the mRNAsi of 1174 genes. We matched the known mRNAsi with the samples and divided our patients into two groups by the median value of mRNAsi (high-mRNAsi group and low-mRNAsi group) [26]. It was found that mRNAsi could not effectively distinguish high- and low-mRNAsi in LUAD and the area under the ROC curve still had a certain gap compared with our signature. However, the mRNAsi of the high-risk group in our signature was also significantly higher than that of the low-risk group. This also verified that our signature was related to stem cells (Fig. 7B) [27, 28].

3.7 Tumor mutational burden and Immune Infiltration Analysis

According to the calculated mutation burden value, we found that it had significant differences in the low-risk and high-risk groups (Fig. 8A). TIMER database, which provides six types of immune cell infiltration, uses RNA-Seq expression profiling data to detect immune cell infiltration in tumor tissue. The signature showed a negative correlation with the levels of B cells, CD4 T cells, CD8 T cells, Dendritic cells, Macrophages and neutrophil cells ($P < 0.05$) (Fig. 8B). These situations revealed that our signature was indeed related to immune cells. In addition, we characterize the cellular composition of the tumor-infiltrating immune cells through CIBERSORT method. Compared with the high-risk group, CD8 T cells, monocytes, resting dendritic cells and resting mast cells had higher expressions ($P < 0.05$), while M0 macrophage had lower expression ($P < 0.001$) (Fig. 8C). CD4 memory activated T cells and CD8 T cells had the highest positive correlation ($r2 = 0.53$), which implied that there was a mutual effect between them. While plasma and M2 Macrophages had the highest negative correlation ($r2 = -0.37$) that suggested they were antagonistic to each other (Fig. 8D) [29].

4. Discussion

Despite the dramatic progress in diagnosis and treatment, the prognosis of advanced lung adenocarcinoma is still unsatisfactory. With the development of clinical management of lung cancer, some prognostic factors are well characterized, such as age, grade and TNM grade. Cancer stem cells refer to cells that have self-renewal capacity and can produce heterogeneous tumor cells, which play a significant role in tumor survival, proliferation, metastasis and recurrence. Therefore, the use of CSCs to establish a prognostic model is conducive to the prediction and precise treatment of LUAD.

We established a signature containing ten genes(C6orf62, DNER, NELL2, LATS2, LGR5, PTPRO, LRIG1, PABPC1, NT5E, SET)and divided patients into high and low-risk groups based on the median risk value. The research on the mechanism of stem cells have been pervasive, but there is no experiment to build these ten stem cells into a signature.

Cancer stem cells or tumor initiating cells are considered to be the main drivers of disease progression and treatment resistance across various cancer types. DNER is a neuron-specific transmembrane protein

with extracellular EGF-like repeat sequences, which promotes the metastasis and proliferation of cancer cells by activating Girdin/PI3K/ATK signal transduction and progression of prostate cancer and the growth of PC-3 cells by regulating the main genes of cancer stem cells [30–32]. NELL2s is a rich glycoprotein that contains EGF-like domains in nerve tissues, interact with protein kinase C and has multiple physiological functions. Hypermethylation of promoter silences NELL2 and affects the progression of renal cell carcinoma [33–35]. LATS2, as a potential tumor suppressor, is a significant mediator of the apoptosis response pathway. LATS2-Wnt / β-catenin / DRP1 / mitochondrial division is identified as a signaling pathway that promotes cancer cells death [36, 37]. LGR5 is a promising marker of intestinal stem cells and cancer stem cells. Intestinal stem cell marker LGR5 is a receptor for R-spongin, and its role is to enhance Wnt signaling in hyperplastic crypts. Wnt pathway plays a significant key in ISC self-renewal by inducing RSPO receptor LGR5 expression. An abnormal increase in LGR5 expression may represent one of the most common molecular changes in some human cancers, resulting in long-term enhancement of canonical Wnt / β-catenin signaling [38–40]. PTPRO is a tumor suppressor and is abnormally expressed in various malignant tumors. PTPRO causes ulcerative colitis through TLR4 / NF-KB signaling pathway and plays a role in liver fibrosis by affecting PDGF signaling in HSC activation. It is noteworthy that PTPRO is a new candidate gene for emphysema with severe obstruction [41, 42]. LRIG1, a transmembrane protein, has a tumor suppressive effect, and its expression is down-regulated in a variety of cancers. It can antagonize epidermal growth factor receptor signaling in epithelial tissues and inhibit cell invasion, migration, VM (angiogenesis simulation) by regulating EGFR / ERK-mediated EMT (epithelial-mesenchymal transition) [43, 44]. PABPC1 can combine with adenylate-rich sequences in mRNA under the action of high affinity, which plays an important role in post-transcriptional regulation of genes and is also involved in many metabolic pathways of mRNA, including adenylate polymerization/adenylation, mRNA transport, mRNA translation, microRNA degradation related regulation [45]. NT5E is a ubiquitously expressed glycosylphatidylinositol-fixed glycoprotein, which can convert extracellular adenosine 5'-monophosphate to adenosine, and promote tumor development by inhibiting the anti-tumor immune response and promoting angiogenesis [46, 47].

GSEA proves that the constructed signature does involve related cancer pathway. P53 is a tumor suppressor protein that regulates the expression of various genes, including apoptosis, growth inhibition, differentiation, inhibition of cell cycle progression and accelerated DNA repair, genotoxicity and senescence after cellular stress. Like all other tumor suppressors, the P53 gene normally slows or monitors cell division. The JAK / STAT signaling pathway is involved in numerous significant biological processes such as cell differentiation, proliferation, migration, apoptosis, survival, and immune regulation. In addition, the JAK / STAT signaling pathway also participates in drug treatment of anemia, thrombocytopenia, neutropenia, and antiviral.

With immune infiltration analysis, we found that the signature regulates the immunity of lung adenocarcinoma through CD4 T cell, which can interfere the immune response of the immune system to the tumor, participate in the immune escape of the tumor, induce the immune tolerance of the tumor, and promote the occurrence and development of the tumor.

5. Conclusion

In conclusion, compared with TMN stage, our robust signature of ten stem cells is a very reliable and accurate prognostic target for LUAD. Further research should be devoted to the functional analysis of our research results and verification in clinical trials.

Abbreviations

LUAD
lung adenocarcinoma
CSC
cancer stem cell
LASSO
least absolute shrinkage operator
ROC
receiver operating characteristic
AUC
area under the curve
TIIC
tumor-infiltrating immune cell
TCGA
The Cancer Genome Atlas
GEO
Gene Expression Omnibus
RECIST
response evaluation criteria in solid tumors
CR
complete response
PR
partial response
SD
stable disease
PD
progressive disease
mRNAsi
mRNA stemness indice
OS
overall survival
GO
gene ontology

KEGG
kyoto encyclopedia of genes and genomes
GSEA
gene set enrichment analysis

Declarations

Ethics approval and consent to participate

Not approval.

Consent for publication

Not approval.

Available of data and materials

All data were from TCGA and GEO, which are publicly available.

Competing interests

The authors declare that they have no competing interests.

Funding

All figures were drawn by R Studio software. This article was supported by Nantong Key Laboratory of Translational Medicine in Cardiothoracic Diseases, Affiliated Hospital of Nantong University and Department of Cardiothoracic Surgery, Affiliated Hospital of Nantong University.

Authors' contributions

*These authors contributed equally to this work.

Acknowledgements

Not applicable.

Authors' information

First author: Zhanghao Huang

Corresponding author: Jiahai Shi, Youlang zhou.

References

1. Mezquita L, et al., *High Prevalence of Somatic Oncogenic Driver Alterations in Patients With NSCLC and Li-Fraumeni Syndrome.* J Thorac Oncol, 2020.
2. Jiang J, et al. *G* Protein-Coupled Receptor GPR87 Promotes the Expansion of PDA Stem Cells through Activating JAK2/STAT3. Mol Ther Oncolytics. 2020;17:384–93.
3. Zhang C, et al., *YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation*. Oncogene, 2020.
4. Mansour FA, et al. PD-L1 is overexpressed on breast cancer stem cells through notch3/mTOR axis. Oncoimmunology. 2020;9(1):1729299.
5. Satirapod C, et al. Estrogen regulation of germline stem cell differentiation as a mechanism contributing to female reproductive aging. Aging. 2020;12(8):7313–33.
6. Kohsaka S, et al., *Identification of Novel CD74-NRG2alpha Fusion From Comprehensive Profiling of Lung Adenocarcinoma in Japanese Never or Light Smokers.* J Thorac Oncol, 2020.
7. Yuan H, et al. CancerSEA: a cancer single-cell state atlas. Nucleic Acids Res. 2019;47(D1):D900–8.
8. Rousseaux S, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci Transl Med. 2013;5(186):186ra66.
9. Tang H, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. Clin Cancer Res. 2013;19(6):1577–86.
10. Kong FE, et al., *Identification of prognostic claudins signature in hepatocellular carcinoma from a hepatocyte differentiation model.* Hepatol Int, 2020.
11. Zeng F, et al. Comprehensive profiling identifies a novel signature with robust predictive value and reveals the potential drug resistance mechanism in glioma. Cell Commun Signal. 2020;18(1):2.
12. Yang L, et al. Long non-coding RNA HOTAIR promotes exosome secretion by regulating RAB35 and SNAP23 in hepatocellular carcinoma. Mol Cancer. 2019;18(1):78.
13. D'Angelo SP, et al., *Avelumab in patients with previously treated metastatic Merkel cell carcinoma: long-term data and biomarker analyses from the single-arm phase 2 JAVELIN Merkel 200 trial*. J Immunother Cancer, 2020. 8(1).
14. Qin S, et al. Co-Expression Network Analysis Identified Genes Associated with Cancer Stem Cell Characteristics in Lung Squamous Cell Carcinoma. Cancer Invest. 2020;38(1):13–22.
15. Shen H, et al. Prognostic biomarker MITD1 and its correlation with immune infiltrates in hepatocellular carcinoma (HCC). Int Immunopharmacol. 2020;81:106222.
16. Pece S, et al. Identification and clinical validation of a multigene assay that interrogates the biology of cancer stem cells and predicts metastasis in breast cancer: A retrospective consecutive study. EBioMedicine. 2019;42:352–62.
17. Santoro A, et al. p53 Loss in Breast Cancer Leads to Myc Activation, Increased Cell Plasticity, and Expression of a Mitotic Signature with Prognostic Value. Cell Rep. 2019;26(3):624–38 e8.
18. Di Z, et al. Integrated Analysis Identifies a Nine-microRNA Signature Biomarker for Diagnosis and Prognosis in Colorectal Cancer. Front Genet. 2020;11:192.

19. Cheng Y, et al. Molecular characterization of lung cancer: A two-miRNA prognostic signature based on cancer stem-like cells related genes. J Cell Biochem. 2020;121(4):2889–900.
20. Kariri YA, et al. Prognostic significance of KN motif and ankyrin repeat domains 1 (KANK1) in invasive breast cancer. Breast Cancer Res Treat. 2020;179(2):349–57.
21. Li W, Liu J, Zhao H. Identification of a nomogram based on long non-coding RNA to improve prognosis prediction of esophageal squamous cell carcinoma. Aging. 2020;12(2):1512–26.
22. Takashima Y, et al. GSEA-assisted gene signatures valid for combinations of prognostic markers in PCNSL. Sci Rep. 2020;10(1):8435.
23. Liao Y, et al. Weighted Gene Coexpression Network Analysis of Features That Control Cancer Stem Cells Reveals Prognostic Biomarkers in Lung Adenocarcinoma. Front Genet. 2020;11:311.
24. Ma F, et al., *Neoadjuvant chemotherapy improves the survival of patients with neuroendocrine carcinoma and mixed adenoneuroendocrine carcinoma of the stomach*. J Cancer Res Clin Oncol, 2020.
25. Castello A, et al., *Metabolic Switch in Hepatocellular Carcinoma Patients Treated with Sorafenib: a Proof-of-Concept Trial.* Mol Imaging Biol, 2020.
26. Malta TM, et al. Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell. 2018;173(2):338–54 e15.
27. Lian H, et al. Integrative analysis of gene expression and DNA methylation through one-class logistic regression machine learning identifies stemness features in medulloblastoma. Mol Oncol. 2019;13(10):2227–45.
28. Pan S, et al. Identification of Biomarkers for Controlling Cancer Stem Cell Characteristics in Bladder Cancer by Network Analysis of Transcriptome Data Stemness Indices. Front Oncol. 2019;9:613.
29. Wang P, et al. Comprehensive Analysis of the Tumor Microenvironment in Cutaneous Melanoma associated with Immune Infiltration. J Cancer. 2020;11(13):3858–70.
30. Wang L, et al. Delta/notch-like epidermal growth factor-related receptor promotes stemness to facilitate breast cancer progression. Cell Signal. 2019;63:109389.
31. Wang L, et al. Delta/notch-like epidermal growth factor-related receptor (DNER) orchestrates stemness and cancer progression in prostate cancer. Am J Transl Res. 2017;9(11):5031–9.
32. Ballester-Lopez C, et al. The Notch ligand DNER regulates macrophage IFNgamma release in chronic obstructive pulmonary disease. EBioMedicine. 2019;43:562–75.
33. Jeong JK, et al. A Role of Central NELL2 in the Regulation of Feeding Behavior in Rats. Mol Cells. 2017;40(3):186–94.
34. Yang Y, et al. iTRAQ-based quantitative proteomic analysis of cerebrospinal fluid reveals NELL2 as a potential diagnostic biomarker of tuberculous meningitis. Int J Mol Med. 2015;35(5):1323–32.
35. Lee da Y, et al. The cytosolic splicing variant of NELL2 inhibits PKCbeta1 in glial cells. Biochem Biophys Res Commun. 2014;454(3):459–64.

36. Zhang L, et al. Anti-tumor effect of LATS2 on liver cancer death: Role of DRP1-mediated mitochondrial division and the Wnt/beta-catenin pathway. Biomed Pharmacother. 2019;114:108825.
37. Guo C, et al. LATS2 inhibits cell proliferation and metastasis through the Hippo signaling pathway in glioma. Oncol Rep. 2019;41(5):2753–61.
38. Cao W, et al., *LGR5 marks targetable tumor-initiating cells in mouse liver cancer.* Nat Commun, 2020. **11**(1): p. 1961.
39. Nakajima T, et al. Characterization of LGR5 expression in poorly differentiated colorectal carcinoma with mismatch repair protein deficiency. BMC Cancer. 2020;20(1):319.
40. Xu L, et al. Lgr5 in cancer biology: functional identification of Lgr5 in cancer progression and potential opportunities for novel therapy. Stem Cell Res Ther. 2019;10(1):219.
41. Yan S, et al. MiR-6803-5p Promotes Cancer Cell Proliferation and Invasion via PTPRO/NF-kappaB Axis in Colorectal Cancer. Mediators Inflamm. 2019;2019:8128501.
42. Radder JE, et al. Extreme Trait Whole-Genome Sequencing Identifies PTPRO as a Novel Candidate Gene in Emphysema with Severe Airflow Obstruction. Am J Respir Crit Care Med. 2017;196(2):159–71.
43. Li W, Zhou Y. *LRIG1 acts as a critical regulator of melanoma cell invasion, migration, and vasculogenic mimicry upon hypoxia by regulating EGFR/ERK-triggered epithelial-mesenchymal transition*. Biosci Rep, 2019. 39(1).
44. Yu S, et al. Expression of LRIG1, a Negative Regulator of EGFR, Is Dynamically Altered during Different Stages of Gastric Carcinogenesis. Am J Pathol. 2018;188(12):2912–23.
45. Su R, et al. PABPC1-induced stabilization of BDNF-AS inhibits malignant progression of glioblastoma cells through STAU1-mediated decay. Cell Death Dis. 2020;11(2):81.
46. Minor M, et al. Cell type- and tissue-specific functions of ecto-5'-nucleotidase (CD73). Am J Physiol Cell Physiol. 2019;317(6):C1079–92.
47. Chen S, et al. CD73: an emerging checkpoint for cancer immunotherapy. Immunotherapy. 2019;11(11):983–97.

Figures

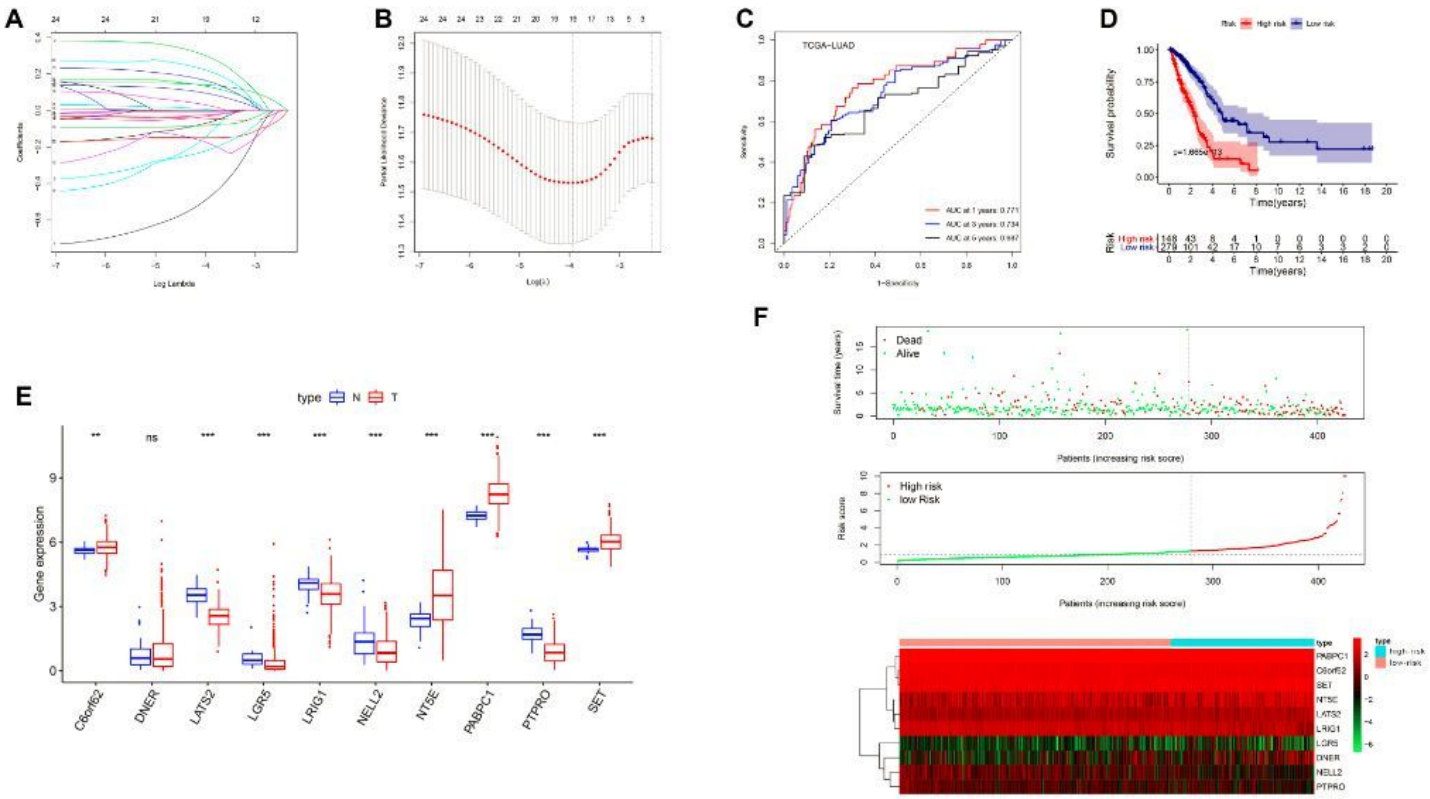

Figure 1

Construction of signature. (A-B) Least absolute shrinkage operator (LASSO) analysis is a method to construct signature. (C) Kaplan-Meier survival reveals the overall survival among different risk stratification groups. (D) ROC curve illustrates the risk prediction of the signature for 1, 3 and 5 years in the TCGA cohort. (E) Box plot shows the expressions and differences of ten stem cells in the signature in normal tissues and cancer tissues. ** represents $P \leq 0.01$, *** represents $P \leq 0.001$, ns represents meaningless. (F) The risk curve in the TCGA cohort displays the patients' risk score, survival time and status and stem cells expression.

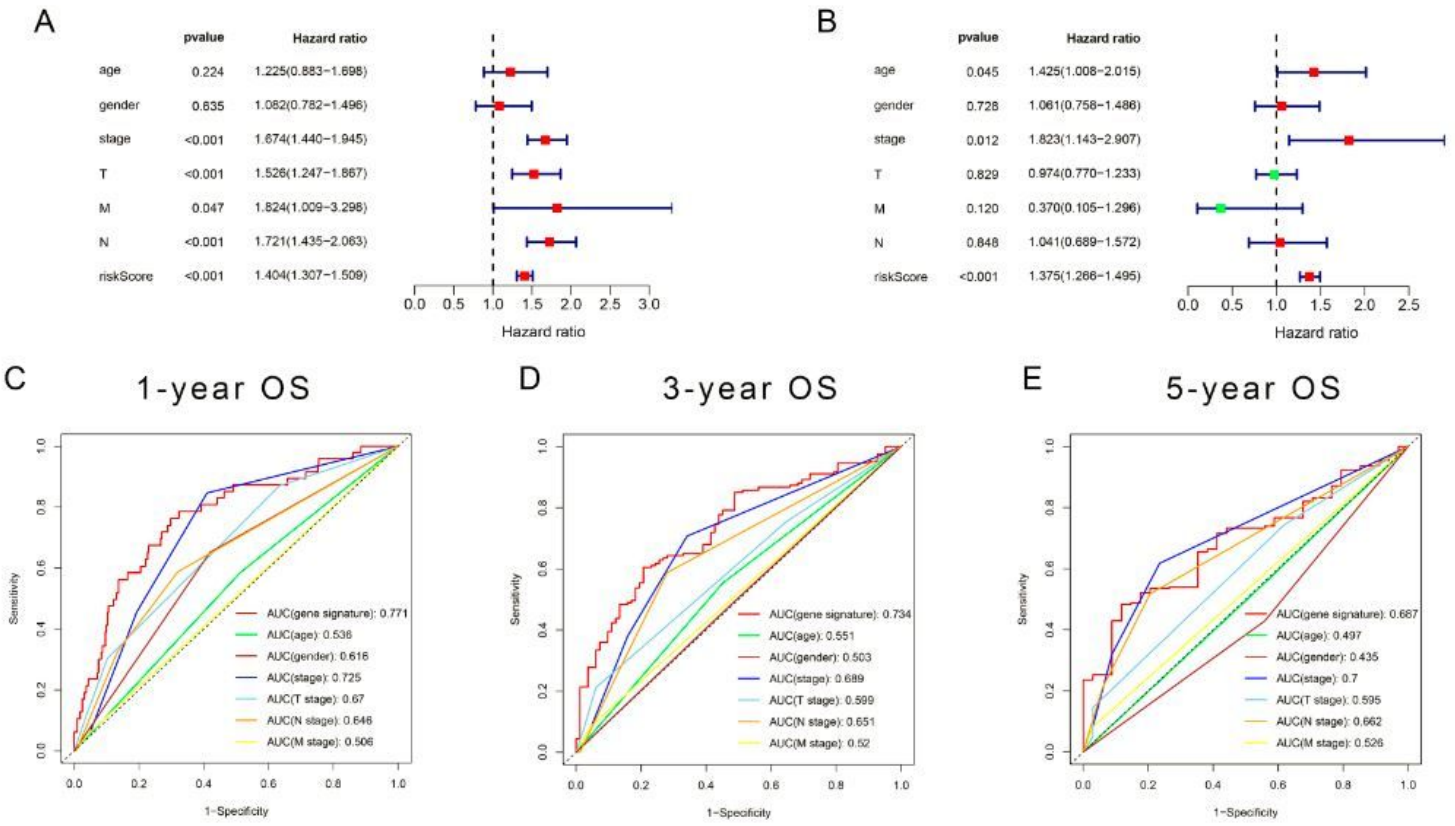

Figure 2

Validation of the signature in TCGA. (A-B) Univariate and Multivariate Cox regression analysis of clinical factors related to overall survival in the TCGA cohort. (C-E) ROC curve demonstrates the risk prediction compared with other clinical factors in the TCGA cohort.

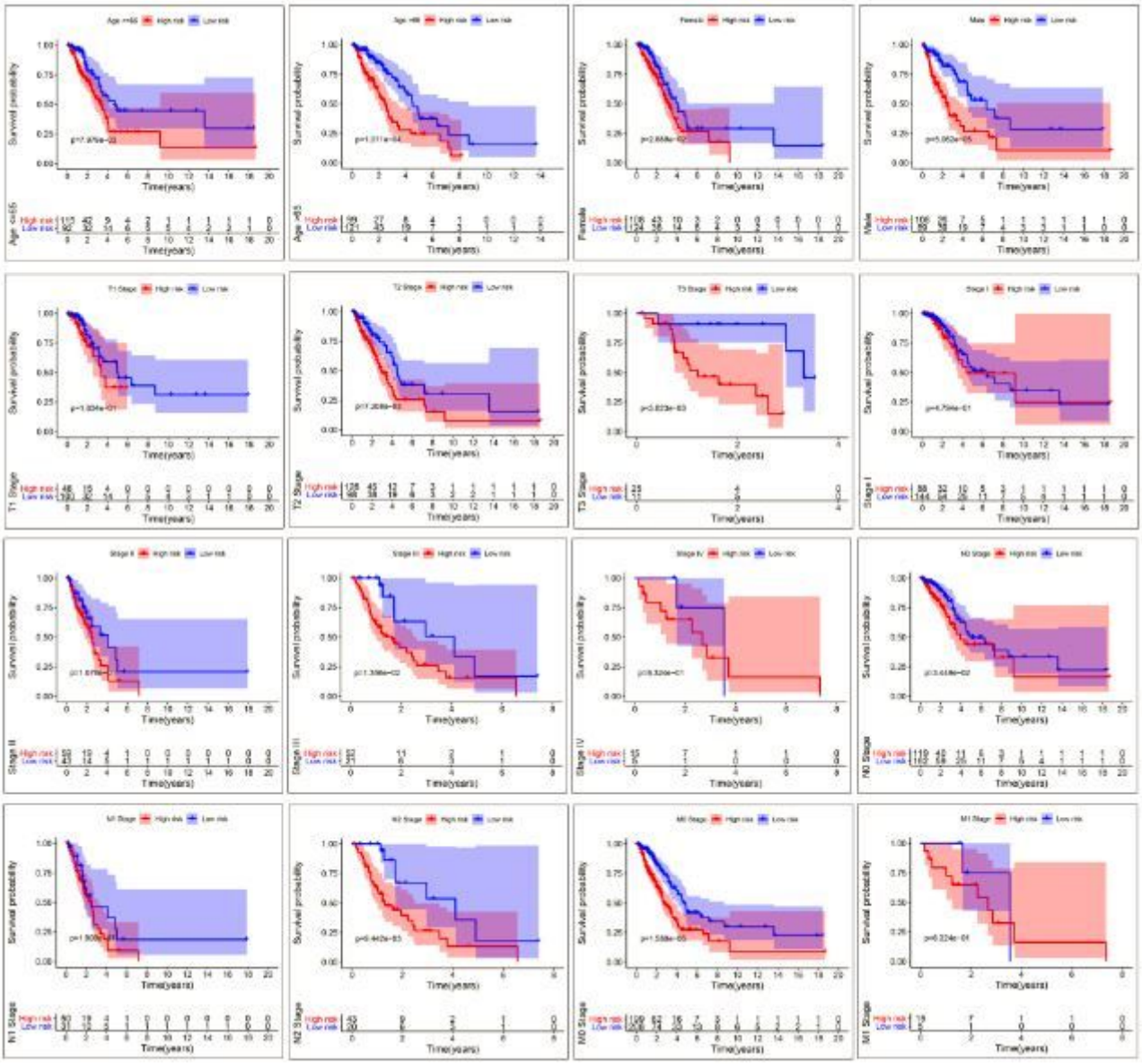

Figure 3

Survival curve analysis. Kaplan-Meier survival illustrates the overall survival of subgroups, which was stratified by age ≤ 65, age > 65, gender and TNM stage.

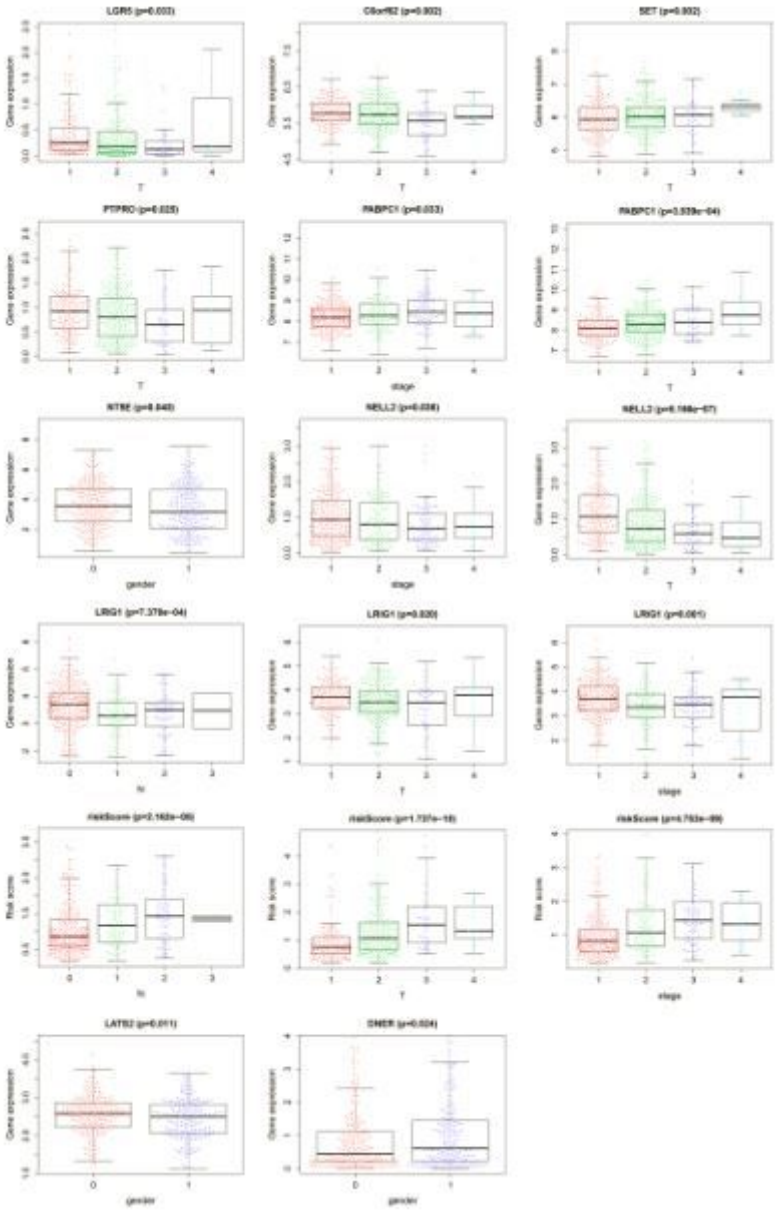

Figure 4

Subgroup analysis. Box plot shows the relationship between stem cells in the signature and each clinical subgroup.

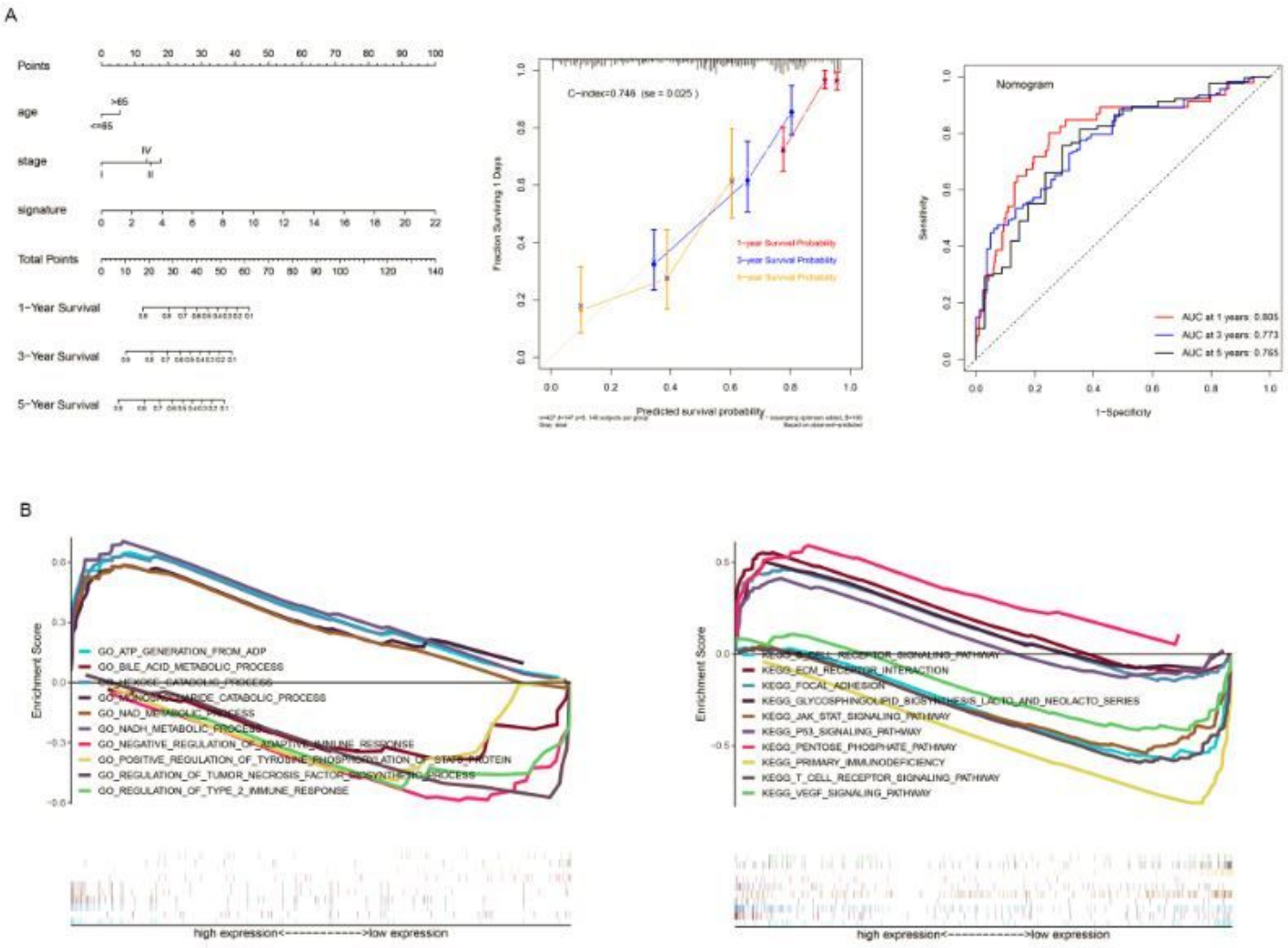

Figure 5

Construction of nomogram and Gene Set Enrichment Analysis. (A) The nomogram contains age, stage, signature containing ten stem cells. The x-axis of the calibration chart is the predicted recurrence probability result, and the y-axis is the actual recurrence probability. ROC analysis detects the accuracy of prediction and inspection. (B) GO term and KEGG pathway show five positive correlation groups and five negative correlation groups, respectively.

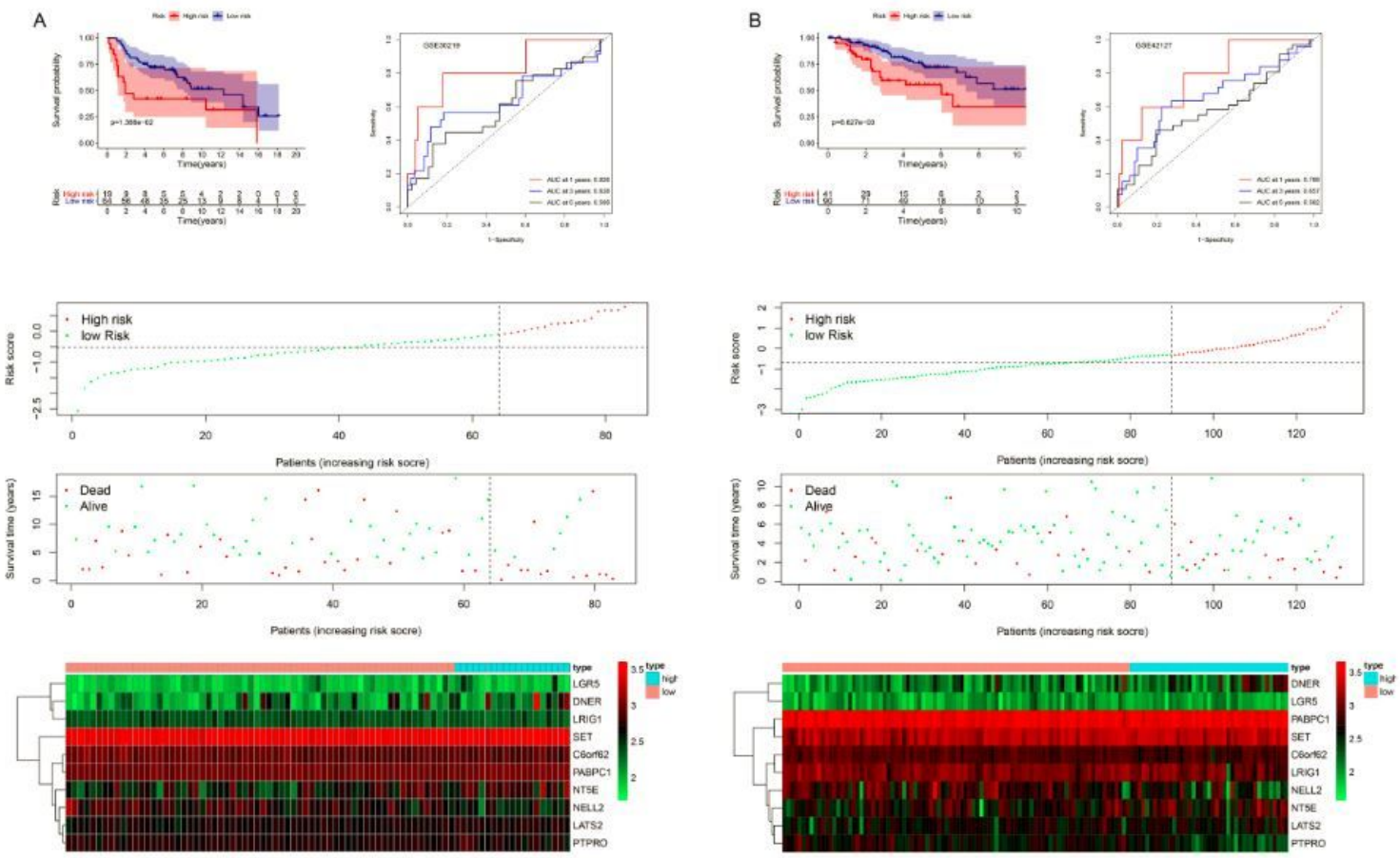

Figure 6

Validation of the signature in GEO. (A) Kaplan-Meier survival, ROC curve and risk plot were used to verify the signature in the GSE30219. (B) Kaplan-Meier survival, ROC curve and risk plot were used to validate the signature in the GSE42127.

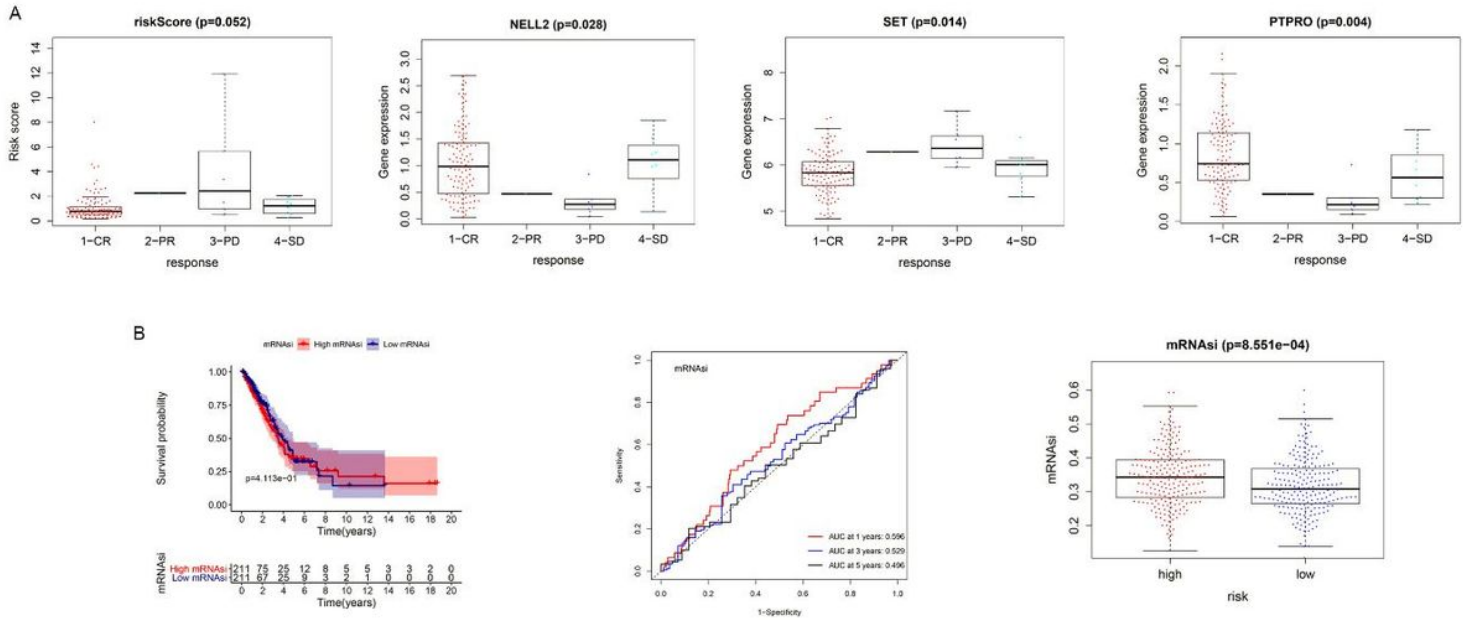

Figure 7

Relationship between the signature and therapeutic efficacy and stem cell index. (A) Box plot suggests the links between signature and stem cells in the signature and drug efficacy. (B) Kaplan-Meier survival, ROC curve and box plot were used demonstrate the risk prediction of signature based on stem cell index.

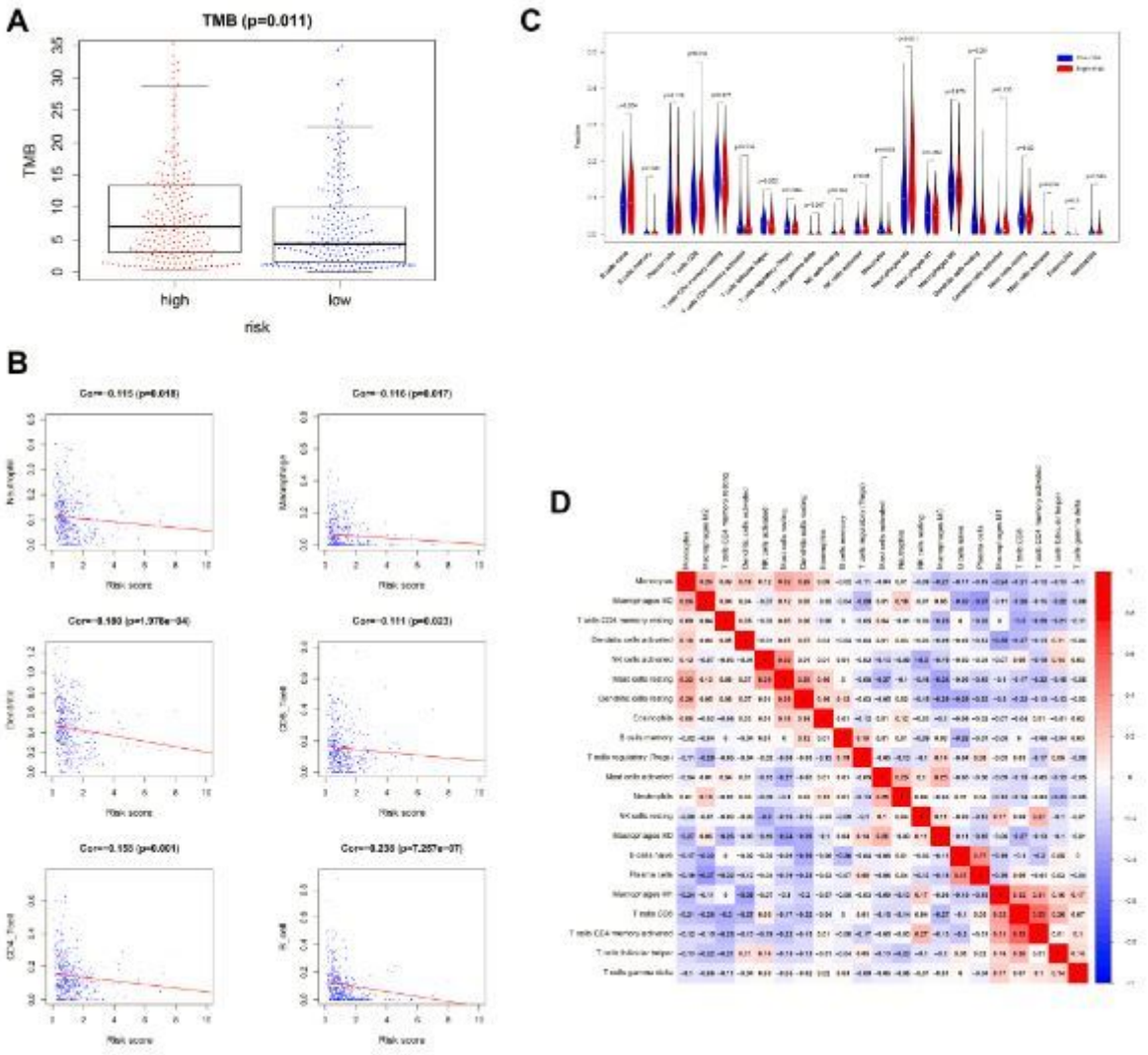

Figure 8

Immune infiltration analyais. (A) Difference analysis of TMB in high-risk and low-risk groups. (B) TIMER indicates the correlations between the six immune cells and signature. (C) Composition of 22 kinds of immune cells in high risk and low risk groups. (D) Correlation heat map of 22 immune cells in LUAD.