DNA methylation-based lung adenocarcinoma subtypes can predict prognosis, recurrence, and immunotherapeutic implications

Feng Xu[1], Lulu He[2], Xueqin Zhan[3], Jiexin Chen[4], Huan Xu[5], Xiaoling Huang[1], Yangyi Li[1], Xiaohe Zheng[1], Ling Lin[5], Yongsong Chen[4]

[1]Department of Respiratory Medicine, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China
[2]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, China
[3]Department of Pulmonology, Children's Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang, China
[4]Department of Endocrinology, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China
[5]Department of Rheumatology, The First Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

Correspondence to: Yongsong Chen, Ling Lin; **email:** yongsongchen@126.com, https://orcid.org/0000-0002-8208-4951; llinc@163.net, https://orcid.org/0000-0001-5603-0688

ABSTRACT

The marked heterogeneity of lung adenocarcinoma (LUAD) makes its diagnosis and treatment difficult. In addition, the aberrant DNA methylation profile contributes to tumor heterogeneity and alters the immune response. We used DNA methylation array data from publicly available databases to establish a predictive model for LUAD prognosis. Thirty-three methylation sites were identified as specific prognostic biomarkers, independent of patients' clinical characteristics. These methylation profiles were used to identify potential drug candidates and study the immune microenvironment of LUAD and response to immunotherapy. When compared with the high-risk group, the low-risk group had a lower recurrence rate and favorable prognosis. The tumor microenvironment differed between the two groups as reflected by the higher number of resting dendritic cells and a lower number of monocytes and resting mast cells in the low-risk group. Moreover, low-risk patients reported higher immune and stromal scores, lower tumor purity, and higher expression of *HLA* genes. Low-risk patients responded well to immunotherapy due to higher expression of immune checkpoint molecules and lower stemness index. Thus, our model predicted a favorable prognosis and increased overall survival for patients in the low-risk methylation group. Further, this model could provide potential drug targets to develop effective immunotherapies for LUAD.

INTRODUCTION

Lung adenocarcinoma (LUAD), a common and aggressive subtype of non-small cell lung cancer (NSCLC), is the primary cause of cancer-related deaths worldwide [1–3]. The overall 5-year survival rate of patients with LUAD has remained low despite rapid advances in diagnostic techniques and molecular therapeutics [4]. Because molecular alterations in tumors occur earlier than clinical variations, novel and effective molecular biomarkers can accurately predict patient prognosis and cancer recurrence. Moreover,

these biomarkers could be used to develop individualized treatment plans.

Epigenetic changes, such as DNA methylation, are inherited modifications that regulate gene expression, without any alteration in the underlying nucleotide sequence [5]. Aberrant DNA methylations are known to occur early during tumorigenesis in several cancers including LUAD [6, 7], and keep accumulating as cancer progresses [8]. Because different cancer subtypes display distinct methylation profiles [9–13], a DNA methylation-based model could provide an effective means to predict and identify potential cancer therapeutics.

DNA methylation is a process during which methyl groups are selectively added to CpG sites to form 5-methylcytosine [14]. We used the DNA methylation array data of LUAD from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases to establish a robust prognosis and recurrence prediction model.

The tumor microenvironment of LUAD gets infiltrated by different immune cell types that contribute to malignancy [15–17]. Cancer immunotherapy involves the application of immune checkpoint blockers to stimulate the immune system against cancer cells. However, its beneficial effects have been reported in less than 20% of patients [18], and more reliable predictors of immune checkpoint blockade response are required. DNA methylation regulates the expression of several genes in the tumor microenvironment and could function as a reliable biomarker for these immune checkpoint blocks [19–22]. Publicly available drug sensitivity databases, including the Connectivity Map (CMap) at the Broad Institute and Genomics of Drug Sensitivity in Cancer (GDSC), could be used to identify candidate drugs against LUAD-specific DNA methylation signatures and develop individualized immunotherapy for patients with LUAD.

RESULTS

DNA methylation sites correlated with patients' survival

The HumanMethylation 450K (HM450K) bead array data of 503 samples (471 LUAD and 32 normal lung tissue) were screened and 21,120 methylation sites were identified. After the exclusion of patients with no survival data, we studied the correlation between DNA methylation sites and patient prognosis with a univariate Cox regression analysis to assess the overall survival of patients with LUAD. Next, the penalized Cox analysis with the Least Absolute Shrinkage and Selection

Operator (LASSO) was performed to narrow down the number of DNA methylation sites, which were selected 900 times over 1,000 repetitions. Finally, a stepwise multivariate Cox regression analysis was performed, and 16 methylation sites were identified as potential prognostic methylation sites. These were used to perform further analyses.

Consensus clustering revealed distinct DNA methylation-based prognostic subgroups

To determine DNA methylation-based clusters of LUAD, we performed an unsupervised hierarchical cluster analysis of patients with LUAD. Depending on the category number, the average cluster consensus and the coefficient of variation among clusters were calculated. We found that most samples in clusters 6 and 7 were stable (Figure 1A). Finally, the optimal cluster number assessed by the cumulative distribution function (CDF) delta area curve was 7 (Figure 1B). Therefore, we divided the samples into seven molecular subtypes. As shown in Figure 1C, a consensus matrix was used to identify the optimal number of clusters. The seven distinct clusters showed different DNA methylation profiles (Figure 1D). The Kaplan-Meier survival analysis revealed that DNA methylation affected the prognosis of patients with LUAD (Figure 1E).

Generation and validation of the prognostic methylation model for patients with LUAD

Because cluster 7 had the poorest prognosis and numerous CpG sites, we selected it as the seed cluster. DNA methylation profiles based on these 33 specific sites were measured for all samples, and we subsequently used them to calculate the risk score of each patient with LUAD. The optimal cut-off for dividing patients into high- or low-risk methylation group was set at 0.254 using the time-dependent receiver operating characteristic (ROC) curve analysis (Supplementary Figure 1). Figure 2A–2C display methylation profiles and risk score distribution. The Kaplan–Meier analysis showed that patients in the high-risk group had worse overall survival than those in the low-risk group ($p < 0.001$; Figure 2D). We next performed a ROC analysis to examine the specificity and sensitivity of the prognostic model. The time-dependent area under the curves (AUCs) for 1-, 3-, and 5-year overall survival rates of patients with LUAD using the prognostic model were 0.901, 0.868, and 0.850, respectively (Figure 2E). A higher AUC indicated better performance for LUAD-specific survival; thus, our data suggested excellent performance for survival prediction. To determine the reliability of the methylation model as a prognostic biomarker,

patients with LUAD were stratified into different subgroups based on the following clinical characteristics: (1) age (age < 60 and age ≥ 60 years), (2) sex (male and female), (3) stage (Stage I + Stage II and Stage III + Stage IV), and (4) T stage (T1 + T2 and T3 + T4). Kaplan–Meier overall survival curves also showed that high-risk patients had shorter overall survival than low-risk patients in different subtyes, further indicating the excellent predictive ability of the methylation model (Figure 3).

Next, we validated the reliability and stability of the model using HM27K bead array data, with survival data in the external validation dataset. The risk scores of patients were calculated using the above-mentioned formula based on the optimal cut-off value. Patients with LUAD were subsequently classified into low- and high-risk groups (Supplementary Figure 2A–2C). Consistent with the above findings, patients in the high-risk group in the validation set had shorter overall survival than those in the low-risk group ($p = 0.01$; Supplementary Figure 2D). The AUC of the ROC analysis for the prognosis model was 0.824, implying

high predictive accuracy and stability for survival prediction (Supplementary Figure 2E).

To study whether the model had similar prognostic values in different patients, we applied the same model to two other cohorts (GSE63384 and GSE83845) as external validation sets. The risk score for each patient with LUAD was calculated using these 33 CpG methylation sites. The patients were assigned to low- and high-risk groups (Supplementary Figure 3A–3C). Patients in the high-risk group had poorer overall survival in the meta-GEO dataset than those in the low-risk group (Supplementary Figure 3D). The AUC of the ROC analysis was 0.650 (Supplementary Figure 3E).

For further internal validation of the methylation model, patients with LUAD from TCGA HM450K were randomly divided into training and testing sets. Supplementary Table 1 shows the baseline characteristics of these two sets. No significant differences in clinical properties were observed between the two datasets ($p > 0.05$). We used the same risk score formula and computed the risk score for all patients in the training
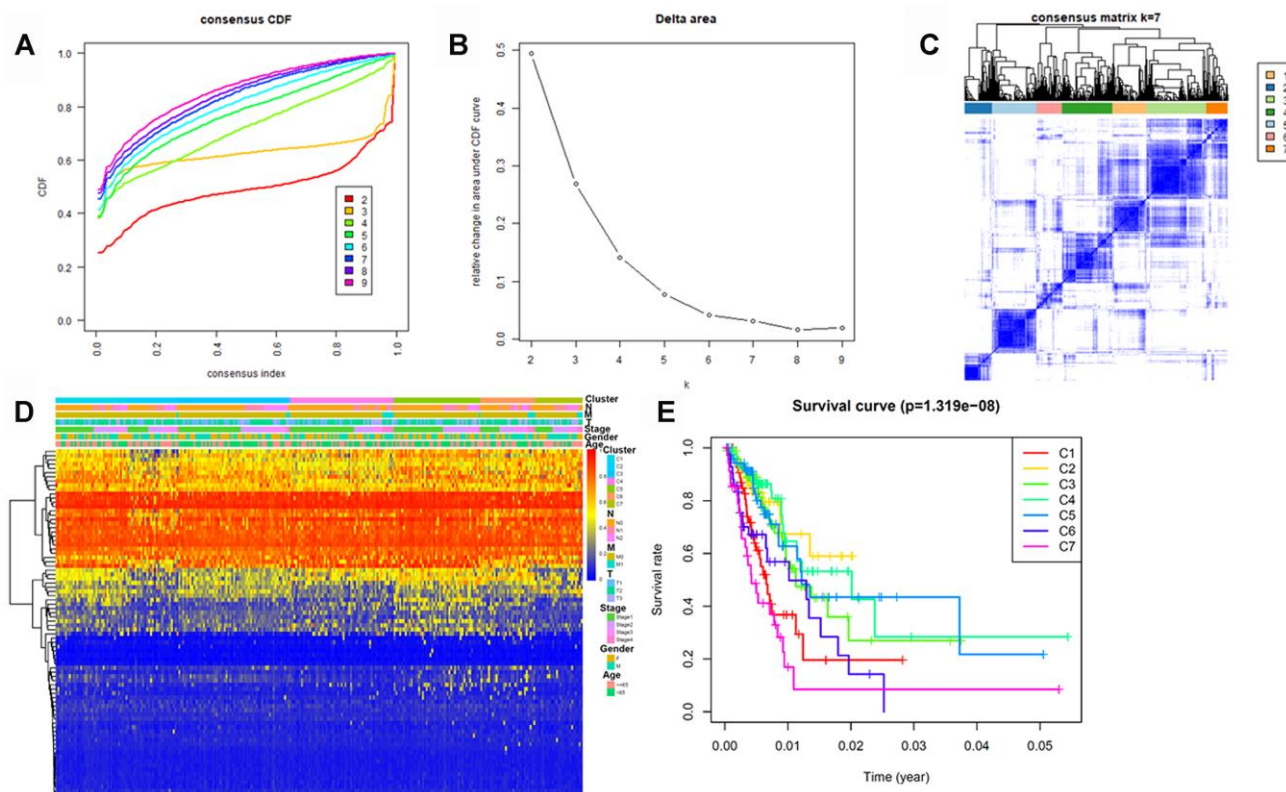

Figure 1. Identification of DNA methylation-based clusters in LUAD samples. (A) Consensus among DNA methylation-based clusters for each category number k. (B) Delta area curve of consensus clustering. (C) Consensus clustering of LUAD samples with k = 6. (D) Heatmap of LUAD methylation differences between each DNA methylation subtype. (E) Kaplan–Meier survival curves of LUAD in each DNA methylation subtype. LUAD, lung adenocarcinoma.

and testing cohorts (Supplementary Figure 4A–4C and 4F-4H). In line with the findings of the TCGA and meta-GEO cohorts, high-risk patients had shorter overall survival than low-risk patients in both cohorts (Supplementary Figure 4D and 4I). Time-dependent ROC analysis indicated that the AUCs for 1-, 3-, and 5-year overall survival rates in the training cohort were 0.895, 0.874, and 0.895, respectively (Supplementary Figure 4E). Moreover, the risk score-based classification of the testing TCGA cohort yielded

similar results. The AUCs for 1-, 3-, and 5-year overall survival rates were 0.906, 0.865, and 0.819, respectively (Supplementary Figure 4J).

Prognostic methylation profiles function as a recurrence model for patients with LUAD

We next constructed a recurrence model using specific methylation sites and disease-free survival time and recurrence status in the dataset from HM450K bead
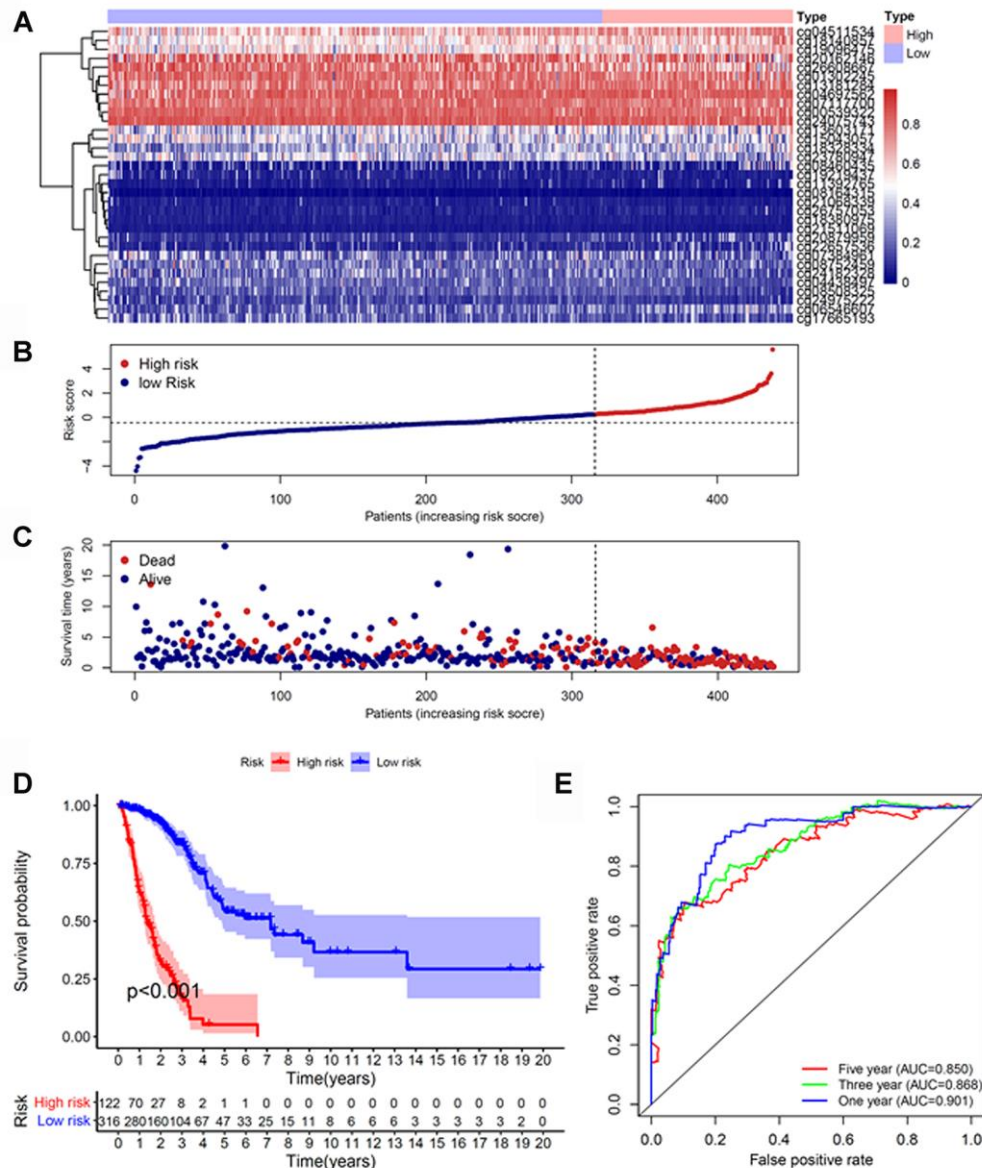

Figure 2. Construction of the prognostic methylation model for patients with LUAD. (**A**) The clustering analysis heatmap of methylation profile in DNA methylation signature sites. (**B**) The distribution of DNA methylation-based risk score. (**C**) Vital status of patients in the high- and low-risk groups. (**D**) Kaplan–Meier survival curves of the relative overall survival of patients in the high- and low-risk groups. (**E**) Accuracy of the prognostic model in predicting survival time by time-dependent ROC curve analysis. LUAD, lung adenocarcinoma; ROC, receiver operating characteristic.

array. When compared with patients in the low-risk group, those in the high-risk group had an elevated recurrence rate ($p < 0.001$) (Figure 4A). The recurrence model resulted in an AUC of 0.682, 0.752, and 0.745 for 1-, 3-, and 5-year disease-free survival time, respectively (Figure 4B). These results verified the predictive accuracy of the recurrence model. Further, these results validated the moderate sensitivity and specificity of the prognostic model. Next, we used this recurrence model to stratify patients with LUAD based on their clinical characteristics such as age, sex, stage, and T stage subgroups (Figure 5).

Tumor immune microenvironment of patients with high- and low-risk LUAD

Next, differences in immune cell infiltration of tumor microenvironment were studied in patients with high-
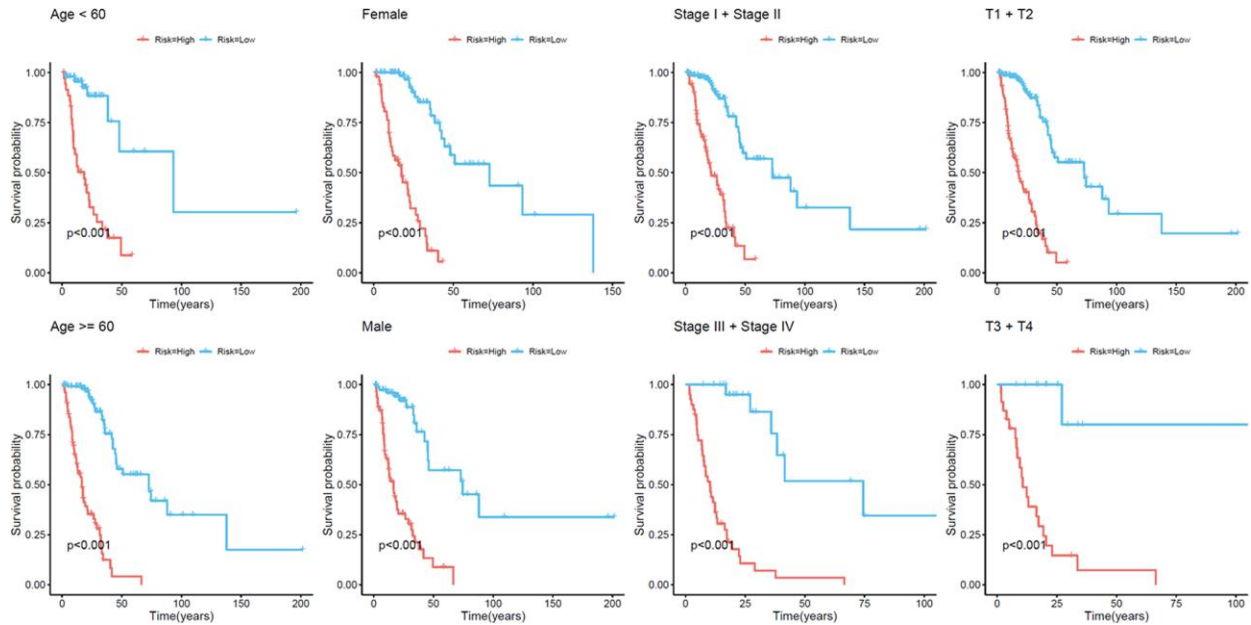

Figure 3. Kaplan–Meier analysis of overall survival for patients with LUAD. Patients were classified according to age (age < 60 and age ≥ 60 years), sex (male and female), TNM stage (Stage I + Stage II and Stage III + Stage IV), and T stage (T1 + T2 and T3 + T4). LUAD, lung adenocarcinoma; TNM, tumor/node/metastasis.
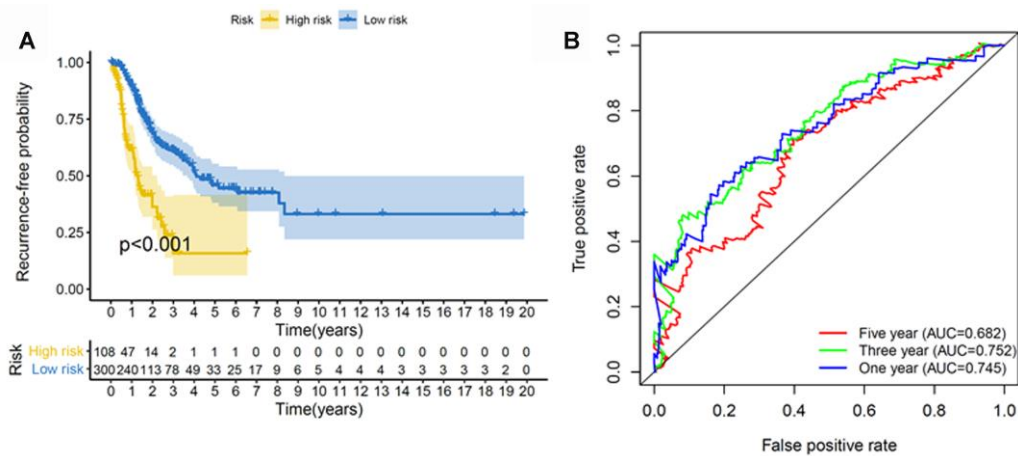

Figure 4. Construction of the recurrence methylation model for patients with LUAD. (**A**) Kaplan-Meier curves of the recurrence model of patients in the high- and low-risk groups. (**B**) Accuracy of the prognostic model in predicting recurrence rate by time-dependent ROC curve analysis. LUAD, lung adenocarcinoma; ROC, receiver operating characteristic.

and low-risk LUAD. The Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data (ESTIMATE) algorithm revealed that the immune and stromal scores were higher in patients in the low-risk group than those in the high-risk group (Figure 6A and 6B). Moreover, we compared the tumor purity of the three LUAD subtypes and found opposite trends—patients in the high-risk group ranked higher than those in the low-risk group (Figure 6C). Because of their clinical implications in immunotherapy, we investigated any potential correlation between the LUAD subtypes and the expression of human leukocyte antigen (*HLA*) genes. Interestingly, patients in the low-risk group reported higher expression of the majority of *HLA* genes than those in the low-risk group (Figure 6D). We next evaluated the differences in immune infiltration of 22 immune cell types between the two groups using the CIBERSORT method in association with the LM22 model matrix. Compared with the high-risk group, patients in the low-risk group had a higher number of resting dendritic cells and a lower number of monocytes and resting mast cells (Figure 6E). To study the biological characteristics of differentially expressed genes between high- and low-risk patients with LUAD, we conducted gene ontology (GO) enrichment analyses. We found differentially expressed genes to be clustered and mostly enriched in immune functions, such as antigen receptor-mediated signaling pathways, immune response-regulating cell surface receptor signaling

pathways, immune response-activating cell surface receptor signaling pathways, regulation of lymphocyte activation, and T cell activation (Figure 6F).

Immunotherapeutic response of LUAD subtypes

We subsequently determined the expression of several key immunomodulators, including TIGIT, ICOS, TIM-3 (HAVCR2), CTLA4, and PD-L1 (CD274) to study the immunotherapeutic response. As shown in Figure 7A–7E, low-risk patients had higher expression of immune checkpoint molecules than high-risk patients. Cancer stem cells are vital for cancer growth, metastasis, and recurrence, and contribute to the resistance of tumors to conventional radiation therapy and chemotherapy. We used a one-class logistic regression (OCLR) machine-learning algorithm to calculate the tumor mRNA expression-based stemness index (mRNAsi). We observed that patients with the high-risk LUAD subtype had elevated stemness indices compared with those in the low-risk group (Figure 7F). Next, we used the Tumor Immune Dysfunction and Exclusion (TIDE) algorithm to predict the likelihood of the response to immunotherapy. Interestingly, we found that patients in the low-risk group were more likely to respond to immunotherapy than those in the high-risk group (Figure 7G). These data further supported our finding that patients with low-risk LUAD subtype had better prognosis and might respond well to immunotherapies.
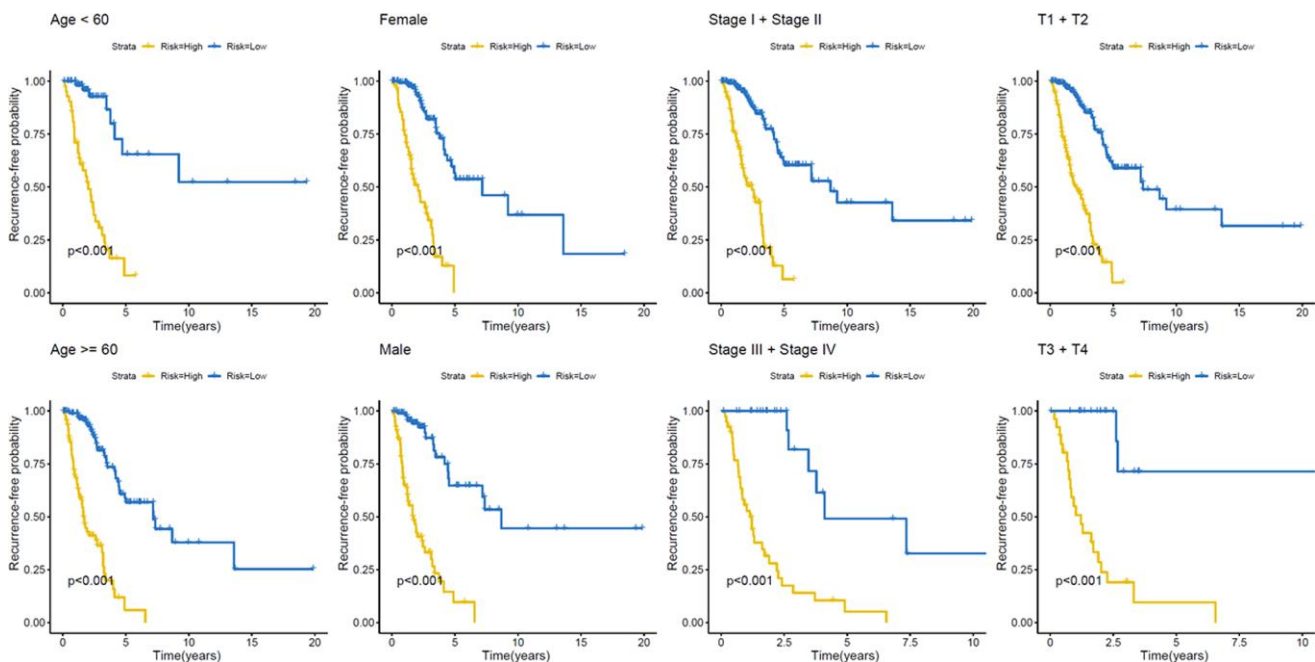

Figure 5. Kaplan–Meier analysis of recurrence-free survival of patients with LUAD. Patients were classified according to age (age < 60 and age ≥ 60 years), sex (male and female), TNM stage (Stage I + Stage II and Stage III + Stage IV), and T stage (T1 + T2 and T3 + T4). LUAD, lung adenocarcinoma; TNM, tumor/node/metastasis.

Identification of novel candidate drugs targeting the methylation signature

We next identified 82 compounds as potential drugs targeting the methylation signatures. The mode-of-action (MoA) analysis of these compounds revealed 59 shared mechanisms of action (Figure 8). The analysis using predictive databases, i.e., CMap and GDSC revealed that five drugs (phenoxybenzamine, terazosin, timolol, dihydroergocristine, and nadolol) shared the MoA of an adrenergic receptor antagonist, five drugs (levomepromazine, trifluoperazine, chlorpromazine, mesoridazine, and pimozide) shared the MoA of a dopamine receptor antagonist, and three drugs (lisuride, quinpirole, and bromocriptine) shared the MoA of a dopamine receptor agonist. Thus, we identified drugs that targeted different methylation profiles in patients with LUAD and could be used for further analysis.

Correlations between the methylation model and clinical properties

We next checked whether the prognostic model was independent of other conventional clinical properties. Univariate Cox regression analysis revealed that the tumor/node/metastasis (TNM), T, M, and N stages and

the risk score correlated with poor survival. Multivariate Cox regression analysis revealed risk score as a specific prognostic indicator for LUAD ($p < 0.001$; Figure 9A). Subsequently, a nomogram integrating the seven factors was constructed for predicting 1-, 3- and 5-year overall survival rates. Compared with the clinical properties, the risk score for the prognostic model displayed superior predictive performance in the nomogram (Figure 9B).

Univariate and multivariate Cox regression analyses were further performed to examine whether the recurrence model was independent of other clinical properties. Cox regression analyses revealed the recurrence model to be significant ($p < 0.001$; Figure 9C). Next, a nomogram that integrated the risk score and clinical risk factors was constructed, in which the risk score for the recurrence model demonstrated good accuracy for predicting 1-, 3- and 5-year disease-free survival rates of patients with LUAD (Figure 9D).

DISCUSSION

Over the past several decades, LUAD has become a major public health concern due to its highly malignant nature [23]. The tumor microenvironment is known to
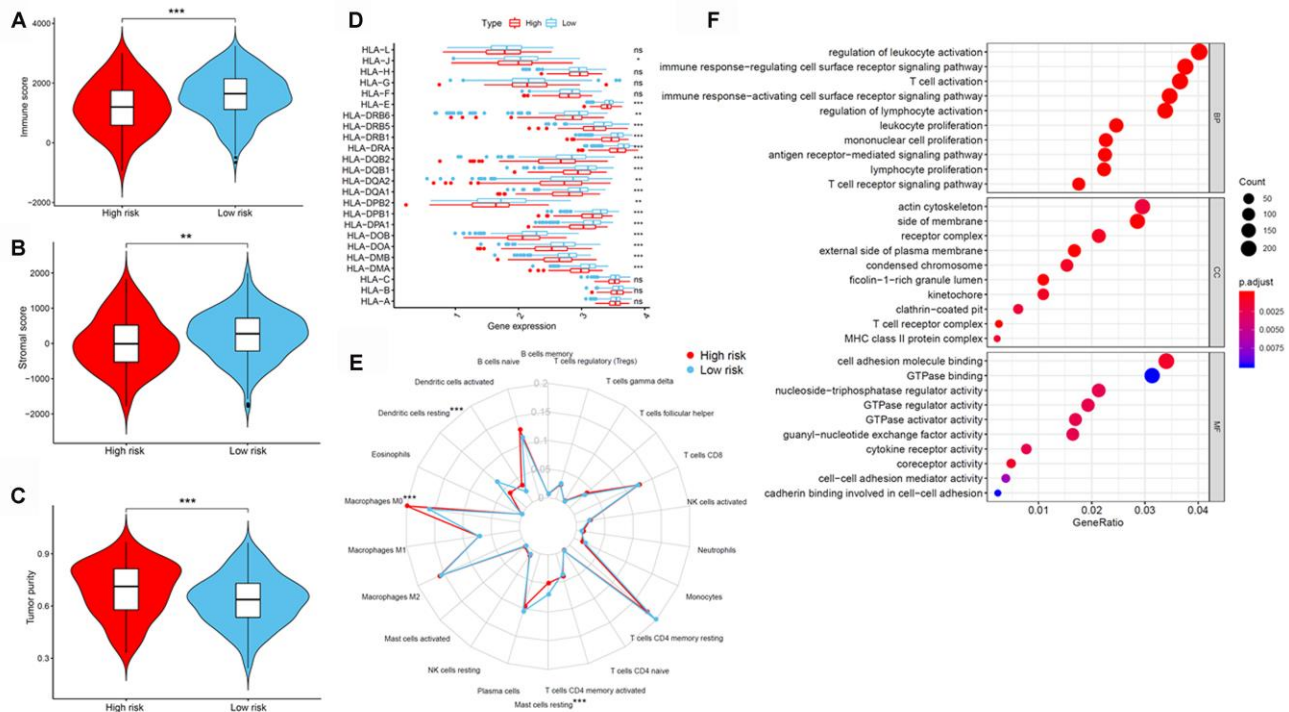

Figure 6. Tumor immune microenvironment of patients in high- and low-risk groups with LUAD. (**A**) Immune scores. (**B**) Stromal scores. (**C**) Tumor purity between patients with high and low risk. (**D**) The expression of *HLA* genes between patients with high and low risk. (**E**) The difference in immune cell infiltration in different LUAD subtypes. (**F**) GO enrichment analyses. GO, gene ontology; HLA, human leukocyte antigen; LUAD, lung adenocarcinoma.

contribute to tumorigenesis and malignant phenotypes [15, 16]. Furthermore, DNA methylation contributes to tumorigenesis by altering the tumor microenvironment of several cancers including LUAD. For example, aberrant DNA methylation can alter immune functions such as T cell differentiation and T cell exhaustion, and the expression of inhibitory immune checkpoint genes, such as *PD-L1*, *PD-L2*, and *CTLA4* [24].
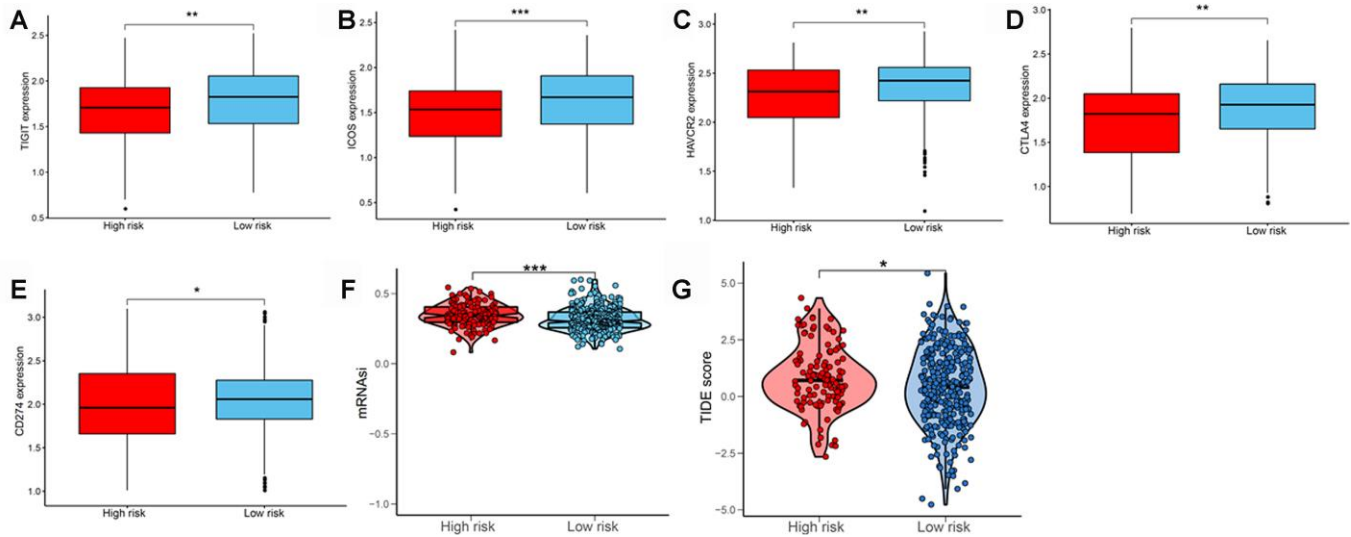

Figure 7. Evaluating the immunotherapeutic response of LUAD subtypes. The expression of immune checkpoint molecules including (**A**) TIGIT, (**B**) ICOS, (**C**) TIM-3 (HAVCR2), (**D**) CTLA4, and (**E**) PD-L1 (CD274) between patients with high and low risk. (**F**) Stemness index values of patients in high- and low-risk groups. (**G**) Immunotherapeutic responses of patients with high and low risk. LUAD, lung adenocarcinoma; PD-L1, programmed cell death-ligand 1.
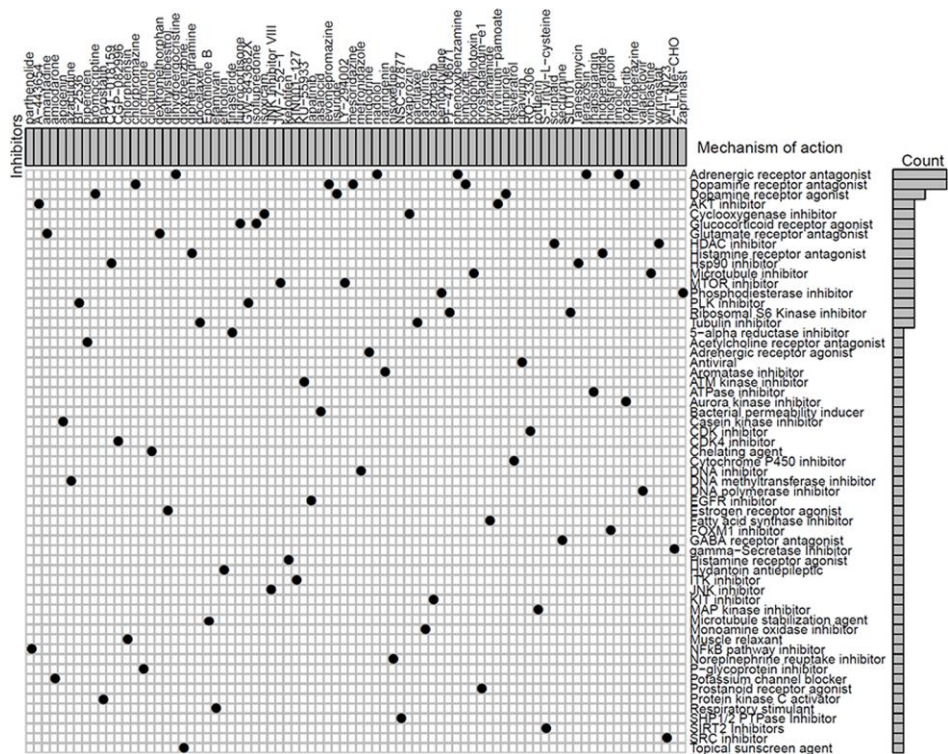

Figure 8. Identification of novel candidate drugs targeting methylation signatures.

Immune checkpoint blockade or immunotherapy is a promising strategy for treating various cancers. For example, antibodies against programmed cell death-1/programmed cell death-ligand 1 (PD-1/PD-L1) immune checkpoint pathway rescued the tumoricidal function of effector T cells [25]. Similarly, anti-PD-1 antibodies are effective in treating several cancers including LUAD, and improving the overall survival [26–28]. However, not all patients with lung cancer respond well to these inhibitors, which could be attributed to checkpoint inhibitor complexity and patients' limited tumor immunity [29, 30]. An improved immune signature-based classification of LUAD could identify subsets of patients who may benefit the most from current therapies [31]. For instance, Xue et al. verified that DNA methylation signatures could reliably predict the immunotherapy response and function as effective biomarkers [19].

We identified 33 DNA methylation sites as novel prognostic and recurrence biomarkers and therapeutic targets for LUAD. Based on clinical characteristics, such as age, sex, TNM stage, and T stage, patients were divided into subgroups to validate the independent predictive value of methylation signatures and study the difference in the overall survival and recurrence rate between the high- and low-risk methylation groups.

Cancer stemness is associated with worse outcomes and suppressed immune responses, such as reduced expression of *PD-L1* [32]. We found that patients with low-risk LUAD subtype reported higher immune and stromal scores, infiltration of resting dendritic cells, and elevated expression of *HLA* and immune checkpoint genes. Moreover, the mRNAsi negatively correlated with the LUAD methylation level, suggesting that DNA methylation negatively affected the transcriptome of LUAD stem cells.

The DNA methylation model could effectively predict the overall survival and recurrence rates, independent of patients' clinical properties. Genes predicted by this model were specifically enriched in immune response. Prediction using the TIDE algorithm indicated that patients with low-risk subtype responded well to immunotherapy. Based on these results, we speculate that this prediction model could provide reliable immune biomarkers for cancer therapy. Further, using CMap and GDSC databases, we identified 82 potential compounds with 59 MoAs and higher immune response in the low-risk group; these could be used to target
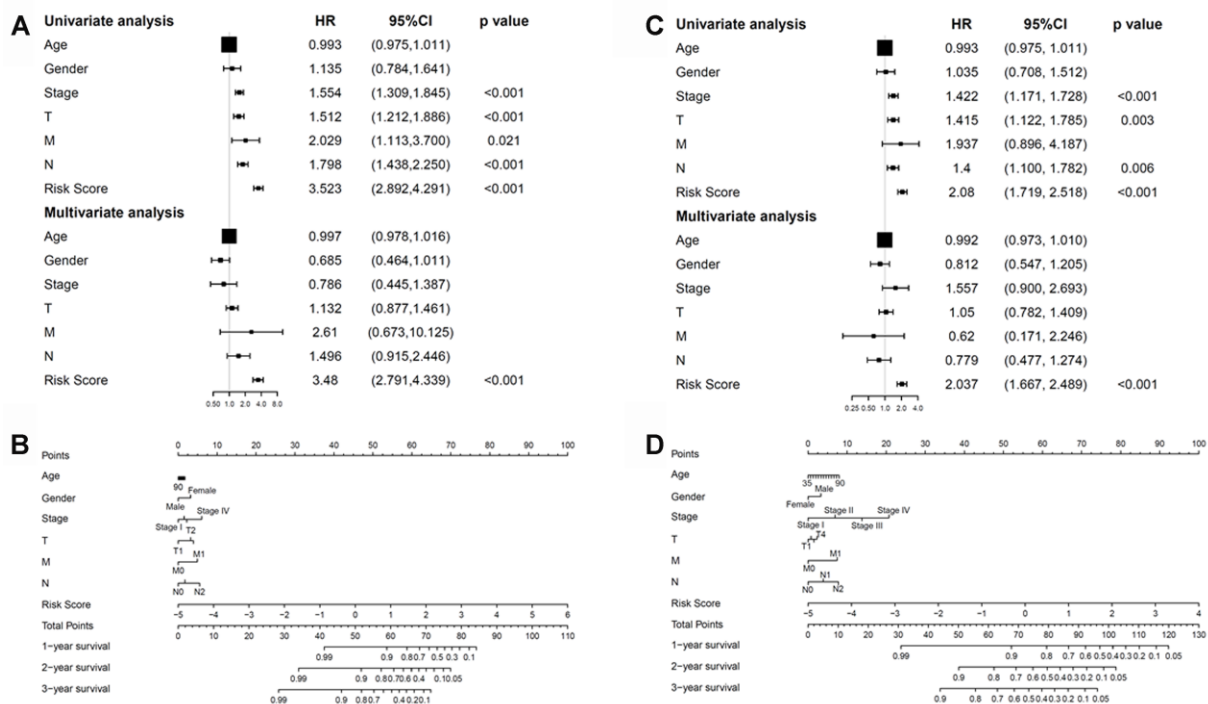

Figure 9. Correlations between the methylation model and clinical characteristics. (**A**) The prognostic model. (**B**) Nomogram for predicting the probability of 1-, 3-, and 5-year overall survival of patients with LUAD. (**C**) The recurrent model. (**D**) Nomogram for predicting the probability of 1-, 3-, and 5-year disease-free survival of patients with LUAD. LUAD, lung adenocarcinoma.

DNA methylation signatures to treat patients with LUAD.

Dendritic cells express adrenergic receptors on their surfaces, stimulation of which by β-agonists modifies the cytokine secretion profiles of these cells [33]. Dendritic cells express dopamine receptors in addition to the machinery necessary to synthesize, store, and degrade dopamine [34]. Glutamate, released by Dendritic cells, is a novel and highly effective regulator in the initiation of T cell-mediated immune responses during T cell–DC interaction [35]. The role of histone deacetylases (HDACs) in the epigenetic regulation of innate and adaptive immunity is of significant interest. HDAC inhibition acetylated and activated signal transducer and activator of transcription-3 (STAT-3), which was critical for the induction of indoleamine 2,3-dioxygenase (IDO) and regulation of Dendritic cells [36]. The phosphatidylinositol-3 kinase/protein kinase B/mammalian target of rapamycin (PI3K-Akt-mTOR) pathway is an important upstream regulator of glycolytic metabolism and plays a central role in Dendritic cells activation and immune responses [37]. Heat shock protein 90 (Hsp90) plays a critical role in protein folding, transport, and cellular activity. Hsp90 was shown to inhibition significantly inhibit Dendritic cell function [38].

During routine clinical work, the pathological stage functions as a prognostic determinant for lung cancer. However, patients with LUAD in the same stage report different clinical outcomes, revealing the inaccuracy of current staging systems in making reliable predictions and revealing LUAD heterogeneity. Therefore, it is necessary to obtain potential predictive and therapeutic biomarkers. The established DNA methylation model offers a novel method for identifying patients with LUAD in addition to predicting the prognosis and recurrence and taking therapeutic decisions.

MATERIALS AND METHODS

Data source and pre-processing

The DNA methylation data of patients with LUAD were generated by Illumina Infinium HumanMethylation 450K and 27K BeadChips (HM450K and HM27K) using the UCSC genome browser. Methylation levels of each CpG site were expressed using the β-value that ranged from unmethylated to fully methylated. First, we excluded the CpG sites with a missing ratio of more than 70% of all samples. Next, we excluded cross-reactive genome CpG sites based on the identification of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation microarray [39]. In addition, the CpG sites in the sex chromosomes were excluded, and those in

the promotor regions were further examined [40]. In total, 503 samples and 21,120 methylation sites for HM450K, and 150 samples and 21,120 methylation sites for HM27K were included in subsequent analyses. RNA sequencing and its clinical LUAD patient data were obtained from the TCGA website. A univariate Cox regression analysis was performed to determine the association between the methylation level of each CpG site and the overall survival of patients with LUAD for HM450K and to identify CpG sites related to overall survival (p-value < 0.05). After primary filtration, a LASSO Cox regression analysis was performed to reduce the number of CpG sites using the R package "glmnet". A multivariate Cox regression analysis was ultimately performed to evaluate the contribution of CpG sites as an independent predictive indicator for patients with LUAD, as previously described [41, 42].

Consensus cluster analysis to identify methylation-based subtypes

To perform the consensus classification of LUAD for HM450K, we used the R package "Consensus ClusterPlus", which provides stable quantitative and visual evidence for estimating the number of unsupervised clusters in a dataset [43]. In each cluster, 80% of the tumors were sampled 100 times, and a k-means algorithm with the Euclidean metric was used. The clustering number was assessed according to the area under the CDF curve [39].

Construction and validation of a prediction methylated risk model

The above-mentioned specific methylation sites were used to construct a prediction model. To validate the methylated signature, the risk score was calculated according to the prognostic signatures in two GEO datasets (GSE63384 and GSE83845) using the R software package "GEOquery". Next, patients with LUAD from TCGA were randomly divided into training and testing cohorts as additional validation datasets.

Evaluation of immune microenvironment

Immune scores were evaluated by applying the ESTIMATE algorithm to the gene expression data from TCGA [44, 45]. Tumor purity was obtained based on the ESTIMATE score using a fitted formula as previously described [45].

Estimation of tumor-infiltrating immune cells

The normalized gene expression data with standard annotation files were uploaded to the CIBERSORT web portal. Next, the algorithm was determined by 1,000

permutations and LM22 gene signature, as described in previous studies [42, 46]. The R "Genefilter" package was used to screen each LUAD sample, and a *p*-value < 0.05 was used to set the threshold.

Functional enrichment analysis

The GO analysis was performed on differentially expressed genes using the R "clusterProfiler" package [47]. The thresholds for analyses were set using a *p*-value < 0.05 that indicated enriched functional annotations.

Calculation of stemness index

Stemness indices were calculated using an innovative OCLR machine-learning algorithm, as previously described [44, 48]. Next, we calculated Spearman's correlations between the stemness index model and the lung cancer sample expression profiles based on the data obtained from TCGA. The stemness index was mapped to the [0,1] range using a linear transformation that subtracted the minimum and divided the maximum value.

Immunotherapeutic response prediction

Several immune checkpoint pathways are involved in tumor immune evasion. Therefore, immune checkpoint inhibitors would enhance anticancer immunity [49]. We used the TIDE algorithm to predict clinical responses of immune checkpoint inhibitors as previously described [42, 50].

Compounds therapeutic response prediction

CMap was used to predict the target therapeutic compounds using the top 1,000 differentially expressed genes [48]. Further, we used the R package "pRRophetic" to predict the half maximal inhibitory concentration (IC_{50}) of chemotherapeutics obtained from the GDSC website in patients with LUAD [51].

Independence of methylation-based model from patients' clinical characteristics

To examine whether prognostic and recurrence models were independent variables as compared with other conventional clinical characteristics (age, gender, and TNM, T, N, and M stages) in patients with LUAD, univariate and multivariate Cox regression analyses were performed.

AUTHOR CONTRIBUTIONS

Feng Xu, Lulu He, Xueqin Zhan, and Jiexin Chen designed the study, analyzed the data, and wrote the manuscript. Huan Xu, Xiaoling Huang, Yangyi Li, and Xiaohe Zheng analyzed the data and contributed to writing the manuscript. Ling Lin and Yongsong Chen supervised the research, analyzed the data, and wrote the manuscript. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

FUNDING

REFERENCES

1. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018; 553:446–54.
   https://doi.org/10.1038/nature25183 PMID:29364287

2. Denisenko TV, Budkevich IN, Zhivotovsky B. Cell death-based treatment of lung adenocarcinoma. Cell Death Dis. 2018; 9:117.
   https://doi.org/10.1038/s41419-017-0063-y
   PMID:29371589

3. Xu F, Chen JX, Yang XB, Hong XB, Li ZX, Lin L, Chen YS. Analysis of lung adenocarcinoma subtypes based on immune signatures identifies clinical implications for cancer therapy. Mol Ther Oncolytics. 2020; 17:241–49.
   https://doi.org/10.1016/j.omto.2020.03.021
   PMID:32346613

4. Zhang L, Zhang Z, Yu Z. Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. J Transl Med. 2019; 17:423.
   https://doi.org/10.1186/s12967-019-02173-2
   PMID:31847905

5. Hulshoff MS, Xu X, Krenning G, Zeisberg EM. Epigenetic regulation of endothelial-to-mesenchymal transition in chronic heart disease. Arterioscler Thromb Vasc Biol. 2018; 38:1986–96.
https://doi.org/10.1161/ATVBAHA.118.311276
PMID:30354260

6. Shi B, Thomas AJ, Benninghoff AD, Sessions BR, Meng Q, Parasar P, Rutigliano HM, White KL, Davies CJ. Genetic and epigenetic regulation of major histocompatibility complex class I gene expression in bovine trophoblast cells. Am J Reprod Immunol. 2018; 79:e12779.
https://doi.org/10.1111/aji.12779
PMID:29131441

7. Shen N, Du J, Zhou H, Chen N, Pan Y, Hoheisel JD, Jiang Z, Xiao L, Tao Y, Mo X. A diagnostic panel of DNA methylation biomarkers for lung adenocarcinoma. Front Oncol. 2019; 9:1281.
https://doi.org/10.3389/fonc.2019.01281
PMID:31850197

8. He W, Ju D, Jie Z, Zhang A, Xing X, Yang Q. Aberrant CpG-methylation affects genes expression predicting survival in lung adenocarcinoma. Cancer Med. 2018; 7:5716–26.
https://doi.org/10.1002/cam4.1834 PMID:30353687

9. Ding W, Feng G, Hu Y, Chen G, Shi T. Co-occurrence and mutual exclusivity analysis of DNA methylation reveals distinct subtypes in multiple cancers. Front Cell Dev Biol. 2020; 8:20.
https://doi.org/10.3389/fcell.2020.00020
PMID:32064261

10. Sharp GC, Ho K, Davies A, Stergiakouli E, Humphries K, McArdle W, Sandy J, Davey Smith G, Lewis SJ, Relton CL. Distinct DNA methylation profiles in subtypes of orofacial cleft. Clin Epigenetics. 2017; 9:63.
https://doi.org/10.1186/s13148-017-0362-2
PMID:28603561

11. Fleischer T, Tekpli X, Mathelier A, Wang S, Nebdal D, Dhakal HP, Sahlberg KK, Schlichting E, Børresen-Dale AL, Borgen E, Naume B, Eskeland R, Frigessi A, et al, and Oslo Breast Cancer Research Consortium (OSBREAC). DNA methylation at enhancers identifies distinct breast cancer lineages. Nat Commun. 2017; 8:1379.
https://doi.org/10.1038/s41467-017-00510-x
PMID:29123100

12. Ferraresso S, Aricò A, Sanavia T, Da Ros S, Milan M, Cascione L, Comazzi S, Martini V, Giantin M, Di Camillo B, Mazzariol S, Giannuzzi D, Marconato L, Aresu L. DNA methylation profiling reveals common signatures of tumorigenesis and defines epigenetic prognostic subtypes of canine diffuse large B-cell lymphoma. Sci Rep. 2017; 7:11591.

https://doi.org/10.1038/s41598-017-11724-w
PMID:28912427

13. Long J, Chen P, Lin J, Bai Y, Yang X, Bian J, Lin Y, Wang D, Yang X, Zheng Y, Sang X, Zhao H. DNA methylation-driven genes for constructing diagnostic, prognostic, and recurrence models for hepatocellular carcinoma. Theranostics. 2019; 9:7251–67.
https://doi.org/10.7150/thno.31155 PMID:31695766

14. Santos KF, Mazzola TN, Carvalho HF. The prima donna of epigenetics: the regulation of gene expression by DNA methylation. Braz J Med Biol Res. 2005; 38:1531–41.
https://doi.org/10.1590/s0100-879x2005001000010
PMID:16172746

15. Mony JT, Schuchert MJ. Prognostic implications of heterogeneity in intra-tumoral immune composition for recurrence in early stage lung cancer. Front Immunol. 2018; 9:2298.
https://doi.org/10.3389/fimmu.2018.02298
PMID:30374348

16. Liu X, Wu S, Yang Y, Zhao M, Zhu G, Hou Z. The prognostic landscape of tumor-infiltrating immune cell and immunomodulators in lung cancer. Biomed Pharmacother. 2017; 95:55–61.
https://doi.org/10.1016/j.biopha.2017.08.003
PMID:28826097

17. Johnson SK, Kerr KM, Chapman AD, Kennedy MM, King G, Cockburn JS, Jeffrey RR. Immune cell infiltrates and prognosis in primary carcinoma of the lung. Lung Cancer. 2000; 27:27–35.
https://doi.org/10.1016/s0169-5002(99)00095-1
PMID:10672781

18. He Y, Jiang Z, Chen C, Wang X. Classification of triple-negative breast cancers based on immunogenomic profiling. J Exp Clin Cancer Res. 2018; 37:327.
https://doi.org/10.1186/s13046-018-1002-1
PMID:30594216

19. Xue G, Cui ZJ, Zhou XH, Zhu YX, Chen Y, Liang FJ, Tang DN, Huang BY, Zhang HY, Hu ZH, Yuan XY, Xiong J. DNA methylation biomarkers predict objective responses to PD-1/PD-L1 inhibition blockade. Front Genet. 2019; 10:724.
https://doi.org/10.3389/fgene.2019.00724
PMID:31475034

20. Elashi AA, Sasidharan Nair V, Taha RZ, Shaath H, Elkord E. DNA methylation of immune checkpoints in the peripheral blood of breast and colorectal cancer patients. Oncoimmunology. 2018; 8:e1542918.
https://doi.org/10.1080/2162402X.2018.1542918
PMID:30713804

21. Sasidharan Nair V, Toor SM, Taha RZ, Shaath H, Elkord E. DNA methylation and repressive histones in the

promoters of PD-1, CTLA-4, TIM-3, LAG-3, TIGIT, PD-L1, and galectin-9 genes in human colorectal cancer. Clin Epigenetics. 2018; 10:104.
https://doi.org/10.1186/s13148-018-0539-3
PMID:30081950

22. Sasidharan Nair V, El Salhat H, Taha RZ, John A, Ali BR, Elkord E. DNA methylation and repressive H3K9 and H3K27 trimethylation in the promoter regions of PD-1, CTLA-4, TIM-3, LAG-3, TIGIT, and PD-L1 genes in human primary breast cancer. Clin Epigenetics. 2018; 10:78.
https://doi.org/10.1186/s13148-018-0512-1
PMID:29983831

23. Pérez-Ramírez C, Cañadas-Garre M, Robles AI, Molina MÁ, Faus-Dáder MJ, Calleja-Hernández MÁ. Liquid biopsy in early stage lung cancer. Transl Lung Cancer Res. 2016; 5:517–24.
https://doi.org/10.21037/tlcr.2016.10.15
PMID:27826533

24. Wang Y, Deng H, Xin S, Zhang K, Shi R, Bao X. Prognostic and predictive value of three DNA methylation signatures in lung adenocarcinoma. Front Genet. 2019; 10:349.
https://doi.org/10.3389/fgene.2019.00349
PMID:31105737

25. Nadal E, Massuti B, Dómine M, García-Campelo R, Cobo M, Felip E. Immunotherapy with checkpoint inhibitors in non-small cell lung cancer: insights from long-term survivors. Cancer Immunol Immunother. 2019; 68:341–52.
https://doi.org/10.1007/s00262-019-02310-2
PMID:30725206

26. Shukuya T, Carbone DP. Predictive markers for the efficacy of anti-PD-1/PD-L1 antibodies in lung cancer. J Thorac Oncol. 2016; 11:976–88.
https://doi.org/10.1016/j.jtho.2016.02.015
PMID:26944305

27. Chen L, Han X. Anti-PD-1/PD-L1 therapy of human cancer: past, present, and future. J Clin Invest. 2015; 125:3384–91.
https://doi.org/10.1172/JCI80011
PMID:26325035

28. Liu J, Nie S, Wu Z, Jiang Y, Wan Y, Li S, Meng H, Zhou S, Cheng W. Exploration of a novel prognostic risk signatures and immune checkpoint molecules in endometrial carcinoma microenvironment. Genomics. 2020; 112:3117–34.
https://doi.org/10.1016/j.ygeno.2020.05.022
PMID:32474122

29. Sasidharan Nair V, Elkord E. Immune checkpoint inhibitors in cancer therapy: a focus on T-regulatory cells. Immunol Cell Biol. 2018; 96:21–33.
https://doi.org/10.1111/imcb.1003
PMID:29359507

30. Long L, Zhao C, Ozarina M, Zhao X, Yang J, Chen H. Targeting immune checkpoints in lung cancer: current landscape and future prospects. Clin Drug Investig. 2019; 39:341–53.
https://doi.org/10.1007/s40261-018-00746-5
PMID:30778885

31. Xiao Q, Zhou D, Rucki AA, Williams J, Zhou J, Mo G, Murphy A, Fujiwara K, Kleponis J, Salman B, Wolfgang CL, Anders RA, Zheng S, et al. Cancer-associated fibroblasts in pancreatic cancer are reprogrammed by tumor-induced alterations in genomic DNA methylation. Cancer Res. 2016; 76:5395–404.
https://doi.org/10.1158/0008-5472.CAN-15-3264
PMID:27496707

32. Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyeuski A, Bruun J, Micke P, de Reynies A, Nelson BH. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc Natl Acad Sci USA. 2019; 116:9020–29.
https://doi.org/10.1073/pnas.1818210116
PMID:30996127

33. Wu H, Chen J, Song S, Yuan P, Liu L, Zhang Y, Zhou A, Chang Y, Zhang L, Wei W. B2-adrenoceptor signaling reduction in dendritic cells is involved in the inflammatory response in adjuvant-induced arthritic rats. Sci Rep. 2016; 6:24548.
https://doi.org/10.1038/srep24548
PMID:27079168

34. Prado C, Contreras F, González H, Díaz P, Elgueta D, Barrientos M, Herrada AA, Lladser Á, Bernales S, Pacheco R. Stimulation of dopamine receptor D5 expressed on dendritic cells potentiates Th17-mediated immunity. J Immunol. 2012; 188:3062–70.
https://doi.org/10.4049/jimmunol.1103096
PMID:22379034

35. Pacheco R, Oliva H, Martinez-Navío JM, Climent N, Ciruela F, Gatell JM, Gallart T, Mallol J, Lluis C, Franco R. Glutamate released by dendritic cells as a novel modulator of T cell activation. J Immunol. 2006; 177:6695–704.
https://doi.org/10.4049/jimmunol.177.10.6695
PMID:17082582

36. Sun Y, Chin YE, Weisiger E, Malter C, Tawara I, Toubai T, Gatza E, Mascagni P, Dinarello CA, Reddy P. Cutting edge: negative regulation of dendritic cells through acetylation of the nonhistone protein STAT-3. J Immunol. 2009; 182:5899–903.
https://doi.org/10.4049/jimmunol.0804388
PMID:19414739

37. Snyder JP, Amiel E. Regulation of dendritic cell immune function and metabolism by cellular nutrient sensor mammalian target of rapamycin (mTOR). Front Immunol. 2019; 9:3145.

https://doi.org/10.3389/fimmu.2018.03145
PMID:30692999

38. Bae J, Mitsiades C, Tai YT, Bertheau R, Shammas M, Batchu RB, Li C, Catley L, Prabhala R, Anderson KC, Munshi NC. Phenotypic and functional effects of heat shock protein 90 inhibition on dendritic cell. J Immunol. 2007; 178:7730–37.
https://doi.org/10.4049/jimmunol.178.12.7730
PMID:17548610

39. Yang C, Zhang Y, Xu X, Li W. Molecular subtypes based on DNA methylation predict prognosis in colon adenocarcinoma patients. Aging (Albany NY). 2019; 11:11880–92.
https://doi.org/10.18632/aging.102492
PMID:31852837

40. Li C, Ke J, Liu J, Su J. DNA methylation data-based molecular subtype classification related to the prognosis of patients with cervical cancer. J Cell Biochem. 2020; 121:2713–24.
https://doi.org/10.1002/jcb.29491 PMID:31680300

41. Xu F, Lin H, He P, He L, Chen J, Lin L, Chen Y. A TP53-associated gene signature for prediction of prognosis and therapeutic responses in lung squamous cell carcinoma. Oncoimmunology. 2020; 9:1731943.
https://doi.org/10.1080/2162402X.2020.1731943
PMID:32158625

42. Xu F, Zhang H, Chen J, Lin L, Chen Y. Immune signature of T follicular helper cells predicts clinical prognostic and therapeutic impact in lung squamous cell carcinoma. Int Immunopharmacol. 2020; 81:105932.
https://doi.org/10.1016/j.intimp.2019.105932
PMID:31836430

43. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26:1572–73.
https://doi.org/10.1093/bioinformatics/btq170
PMID:20427518

44. Daily K, Ho Sui SJ, Schriml LM, Dexheimer PJ, Salomonis N, Schroll R, Bush S, Keddache M, Mayhew C, Lotia S, Perumal TM, Dang K, Pantano L, et al. Molecular, phenotypic, and sample-associated data to describe pluripotent stem cell lines and derivatives. Sci Data. 2017; 4:170030.
https://doi.org/10.1038/sdata.2017.30
PMID:28350385

45. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW,

Levine DA, Carter SL, Getz G, Stemke-Hale K, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013; 4:2612.
https://doi.org/10.1038/ncomms3612
PMID:24113773

46. Zhou R, Zhang J, Zeng D, Sun H, Rong X, Shi M, Bin J, Liao Y, Liao W. Immune cell infiltration as a biomarker for the diagnosis and prognosis of stage I-III colon cancer. Cancer Immunol Immunother. 2019; 68:433–42.
https://doi.org/10.1007/s00262-018-2289-7
PMID:30564892

47. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16:284–87.
https://doi.org/10.1089/omi.2011.0118
PMID:22455463

48. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwińska P, Mazurek S, et al, and Cancer Genome Atlas Research Network. Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell. 2018; 173:338–54.e15.
https://doi.org/10.1016/j.cell.2018.03.034
PMID:29625051

49. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. J Clin Oncol. 2015; 33:1974–82.
https://doi.org/10.1200/JCO.2014.59.4358
PMID:25605845

50. Lu X, Jiang L, Zhang L, Zhu Y, Hu W, Wang J, Ruan X, Xu Z, Meng X, Gao J, Su X, Yan F. Immune signature-based subtypes of cervical squamous cell carcinoma tightly associated with human papillomavirus type 16 expression, molecular features, and clinical outcome. Neoplasia. 2019; 21:591–601.
https://doi.org/10.1016/j.neo.2019.04.003
PMID:31055200

51. Geeleher P, Cox NJ, Huang RS. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. Genome Biol. 2014; 15:R47.
https://doi.org/10.1186/gb-2014-15-3-r47
PMID:24580837
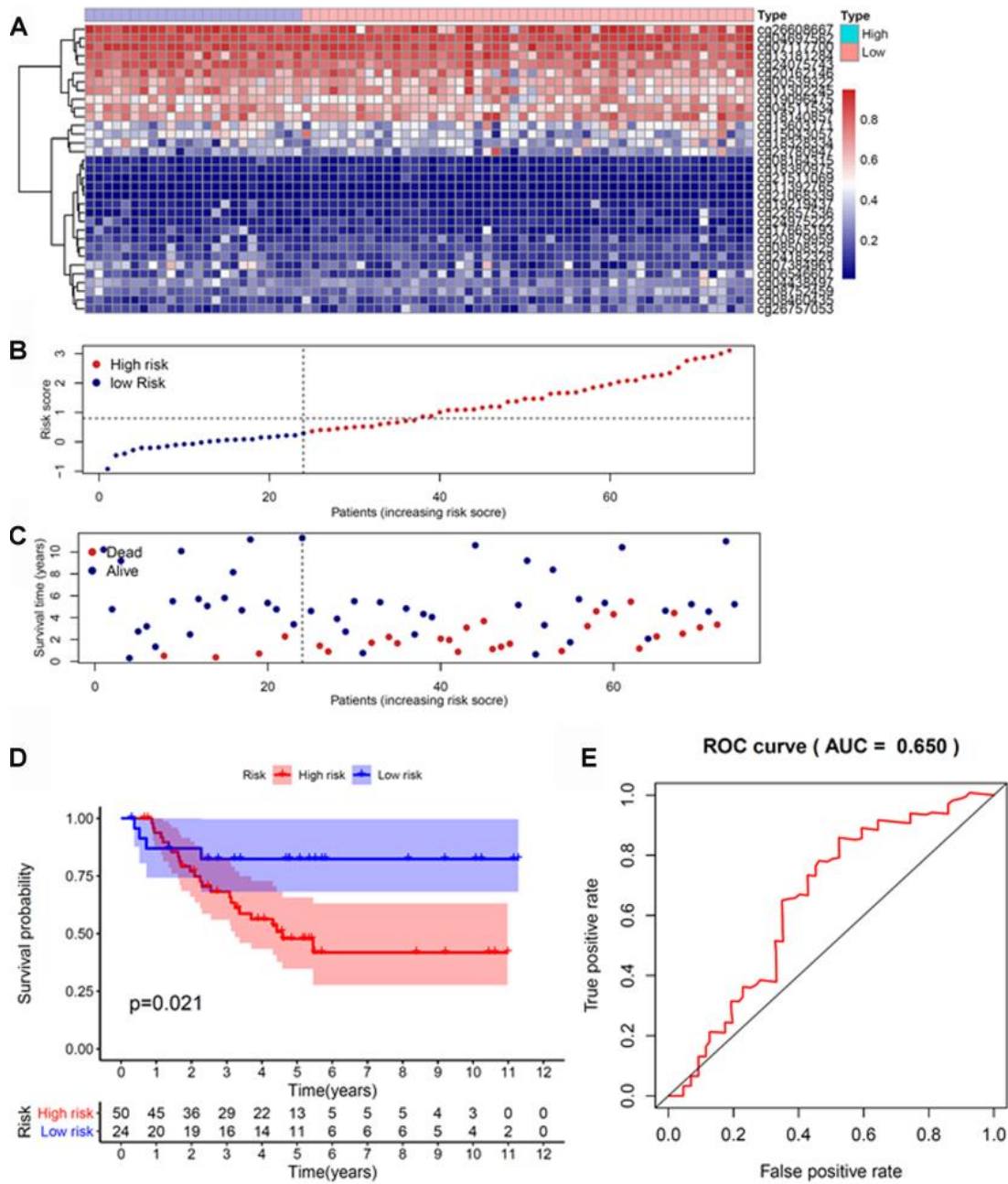
SUPPLEMENTARY MATERIALS

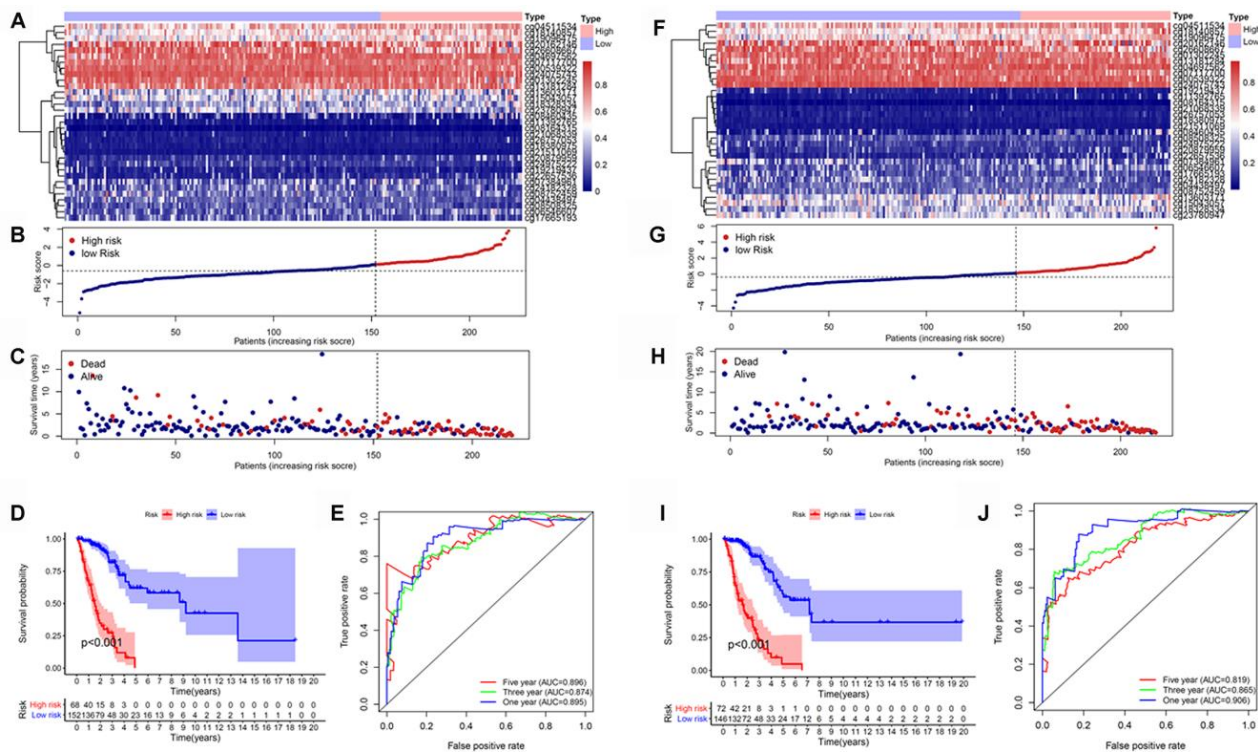## Supplementary Figures
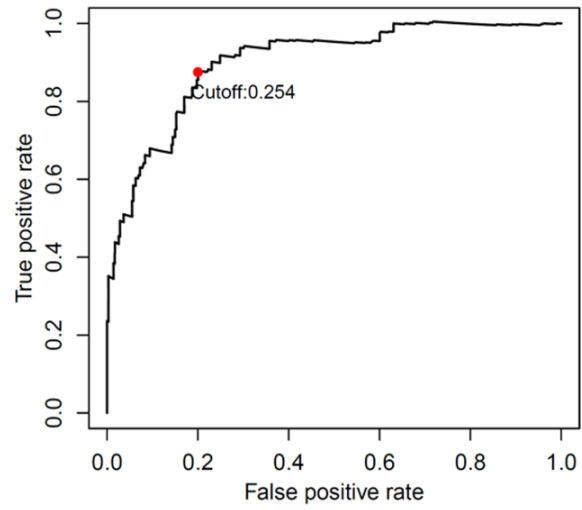
Supplementary Figure 1. Optimal cut–off value for the risk score.
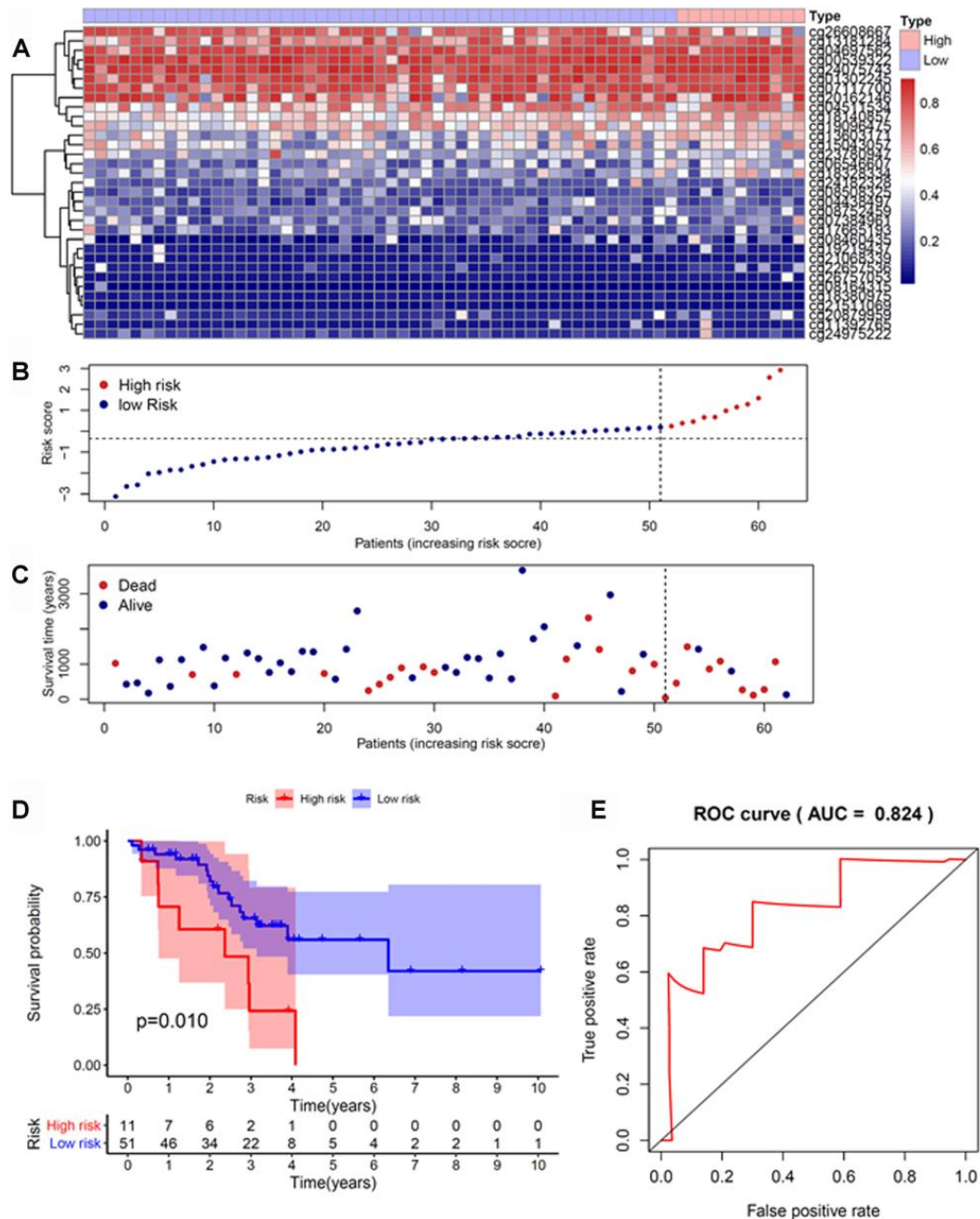
Supplementary Figure 2. Validation of the prognostic methylation model for patients with LUAD using the HM27K dataset.
(**A**) The clustering analysis heatmap of methylation profile in DNA methylation signature sites. (**B**) Distribution of the DNA methylation-based risk score. (**C**) Vital status of patients in the high- and low-risk groups. (**D**) Kaplan–Meier survival curves of the relative overall survival of patients in the high- and low-risk groups. (**E**) Accuracy of the prognostic model in predicting survival time by ROC curve analysis. LUAD, lung adenocarcinoma; HM27K, HumanMethylation 27K; ROC, receiver operating characteristic.

Supplementary Figure 3. Validation of the prognostic methylation model for patients with LUAD using the meta-GEO dataset. (**A**) The clustering analysis heatmap of methylation profile in DNA methylation signature sites. (**B**) The distribution of the DNA methylation-based risk score. (**C**) Vital status of patients in the high- and low-risk groups. (**D**) Kaplan–Meier survival curves of the relative overall survival of patients in high- and low-risk groups. (**E**) Accuracy of the prognostic model in predicting survival time by ROC curve analysis. LUAD, lung adenocarcinoma; GEO, Gene Expression Omnibus; ROC, receiver operating characteristic.

Supplementary Figure 4. Validation of the prognostic methylation model for patients with LUAD using TCGA internal validation. (**A**) The clustering analysis heatmap of methylation profile in the DNA methylation signature sites for the training set. (**B**) Distribution of the DNA methylation-based risk score for the training set. (**C**) Vital status of patients in the high- and low-risk groups for the training set. (**D**) Kaplan–Meier survival curves of the relative overall survival of patients in high- and low-risk groups for the training set. (**E**) Accuracy of the prognostic model in predicting survival time by ROC curve analysis for the training set. (**F**) The clustering analysis heatmap of methylation profile in DNA methylation signature sites for the testing set. (**G**) Distribution of the DNA methylation-based risk score for testing set. (**H**) Vital status of patients in the high- and low-risk groups for the testing set. (**I**) Kaplan–Meier survival curves of the relative overall survival of patients in the high- and low-risk groups for the testing set. (**J**) Accuracy of the prognostic model in predicting survival time by ROC curve analysis for the testing set. LUAD, lung adenocarcinoma; ROC, receiver operating characteristic; TCGA, The Cancer Genome Atlas.

Supplementary Table

Supplementary Table 1. Summary of patients' characteristics.

Characteristics	Training set		Testing set		*p*-Value
	Number	**Percentage**	**Number**	**Percentage**	
Age (years)					
<60	59	26.82	64	29.36	0.5823
≥60	157	71.36	148	67.89	
Not available	4	1.82	6	2.75	
Sex					
Female	116	52.73	114	52.29	1
Male	104	47.27	104	47.71	
Stage					
Stage I–II	171	77.73	173	79.36	0.9805
Stage III–IV	45	20.45	44	20.18	
Not available	4	1.82	1	0.46	
T					
T1–T2	190	86.36	192	88.07	0.6682
T3–T4	29	13.18	24	11.01	
Not available	1	0.45	2	0.92	
M					
M0	141	64.09	136	62.39	0.9909
M1	10	4.55	9	4.13	
Not available	69	31.36	73	33.49	
N					
N0	141	64.09	146	66.97	0.7056
N1–N3	73	33.18	69	31.65	
Not available	6	2.73	3	1.38	