A Methylation-based nomogram for predicting survival in patients with lung adenocarcinoma

Xuelong Wang
Beijing Electric Power Hospital
Bin Zhou
Beijing Electric Power Hospital
Yuxin Xia
Beijing Electric Power Hospital
Jianxin Zuo
Beijing Electric Power Hospital
Yanchao Liu
Beijing Electric Power Hospital
Xin Bi
Beijing Electric Power Hospital
Xiong Luo
Beijing Nuclear industry hospital
Chengwei Zhang (bjdlyyxwk@163.com)
Beijing Electric Power Hospital https://orcid.org/0000-0003-1620-4510

Research article

Keywords: lung adenocarcinoma, DNA methylation, differentially methylated sites, prognosis, signature

Posted Date: September 23rd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-75403/v1

Version of Record: A version of this preprint was published at BMC Cancer on July 12th, 2021. See the published version at https://doi.org/10.1186/s12885-021-08539-4.

Abstract

Background

DNA methylation alteration is frequently observed in Lung adenocarcinoma (LUAD) and may play important roles in carcinogenesis, diagnosis, and prognosis. Thus, this study aimed to construct a reliable methylation-based nomogram, guiding prognostic classification screening and personalized medicine for LUAD patients.

Method:

The DNA methylation data, gene expression data and corresponding clinical information of lung adenocarcinoma samples were extracted from The Cancer Genome Atlas (TCGA) database. Differentially methylated sites (DMSs) and differentially expressed genes (DEGs) were obtained and then calculated expression correlation by pearson correlation coefficient. Functional enrichment analysis and Protein-protein interaction network were used to explore the biological roles of aberrant methylation genes. A prognostic risk score model was constructed using univariate Cox and LASSO analysis and was assessed in an independent cohort. A methylation-based nomogram that included the risk score and the clinical risk factors was developed, which was evaluated by concordance index and calibration curves.

Result

We identified a total of 1362 DMSs corresponding to 471 DEGs with significant negative correlation, including 752 hypermethylation sites and 610 hypomethylation sites. Univariate cox regression analysis showed that 59 DMSs were significantly associated with overall survival. Using LASSO method, we constructed a three-DMSs signature that was independent predictive of prognosis in the training cohort. Patients in high-risk group had a significant shorter overall survival than patients in low-risk group classified by three-DMSs signature (log-rank p = 1.9E-04). Multivariate cox regression analysis proved that the three-DMSs signature was an independent prognostic factor for LUAD in TCGA-LUAD cohort (HR = 2.29, 95%CI: 1.47–3.57, P = 2.36E-04) and GSE56044 cohort (HR = 2.16, 95%CI: 1.19–3.91, P = 0.011). Furthermore, a nomogram, combining the risk score with clinical risk factors, was developed with C-indexes of 0.71 and 0.70 in TCGA-LUAD and GSE56044 respectively.

Conclusions

The present study established a robust three-DMSs signature for the prediction of overall survival and further developed a nomogram that could be a clinically available guide for personalized treatment of LUAD patients.

Introduction

Lung cancer is the leading cause of cancer-related deaths worldwide [1], including two main types known as small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC). Lung adenocarcinoma (LUAD) is the most predominant subtype of NSCLC, with increased incidence over the past decades worldwide [2]. Despite recent advances in surgical techniques, radiotherapeutic interventions and combined chemotherapy strategies, the long-term survival rate of patients diagnosed with LUAD has not significantly improved [3]. Thus, it is indeed urgent to identify specific details regarding characteristic molecules in LUAD tissue to evaluate the prognosis of LUAD and develop strategies for personalized therapy.

DNA methylation, a primary epigenetic modification in the mammalian genome, often occurs at CpG islands. Increasing studies demonstrated that aberrant DNA methylation could play a key role in the progression and metastasis of LUAD, reflecting important biological features in the etiology [4–7]. Shen et al. demonstrates that the methylation status of homeobox A9 (*HOXA9*), keratin-associated protein 8 − 1 (*KRTAP8-1*), cyclin D1 (*CCND1*), and tubby-like protein 2 (*TULP2*) has great potential for the early recognition of LUAD in the undetermined lung nodules [8]. Seok et al. found that TGFBI promoter methylation is associated with poor prognosis in lung adenocarcinoma patients [9]. Furthermore, a prognostic DNA methylation signature was established by Sandoval et al. to distinguished patients with high- and low-risk early stage NSCLC, guiding the adjuvant chemotherapy [10]. Additionally, Zheng et al. suggested an internal CpG-based signature for survival prediction of lung adenocarcinoma patients. Such researches demonstrate that the methylation level is deemed a crucial molecular biomarker for the diagnosis and prognosis of LUAD patients [11–13]. Despite these remarkable findings, limited by either the current expertise on the association between the epigenetic modifications and clinical outcomes or lack of independent validation as small sample size, there is a lack of a robust prognostic DNA methylation signature for LUAD.

In the present study, we extracted the DNA methylation data, gene expression data and corresponding clinical information of lung adenocarcinoma samples from The Cancer Genome Atlas (TCGA) database to select the differentially methylated sites (DMSs) corresponding to dysregulated genes and further explore the biological processes in which the aberrant methylation genes might be involved. Moreover, performing univariate Cox and LASSO analysis, we constructed a robust DMSs-based prognostic signature and validated the prognostic performance in an independent cohort extracted from Gene Expression Omnibus (GEO). Furthermore, combing DMSs-based prognostic signature with clinical risk factors, we constructed a nomogram that could provide insight into regarding survival prediction and serve as a clinically available guide for personalized treatment of LUAD patients.

Materials And Methods

Data processing

All datasets and clinical information were described in Table 1 and Supplementary Table S1. The DNA methylation data (459 LUAD tissues and 30 normal tissues) and gene expression data (513 LUAD tissues and 59 normal tissues) of lung adenocarcinoma samples were extracted from TCGA (https://cancergenome.nih.gov/). Methylation beta-values derived from Illumina Infinium Human Methylation 450 BeadChip platform were extracted as site methylation measurements. The normalised count values of level 3 gene expression data derived from Illumina HiSeqV2 were extracted as gene expression measurements. Clinical information of 513 LUAD patients was obtained from TCGA. After corresponding patients with both methylation data and expression data, ninety-six LUAD patients were excluded because of unknown survival time, age, and tumor stage. Ultimately, 417 patients were retained in our study. An independent dataset (GSE56044 [14]) collected from GEO (https://www.ncbi.nlm.nih.gov/geo/) was used to test the prognostic ability, containing 82 LUAD patients with both methylation data and clinical information.

Table 1
Cohorts analyzed in present study

	Training cohort (TCGA-LUAD)			Validation cohort (GSE56044)	
	Methylation data	**Expression data**	**Clinical data**	**Methylation data**	**Clinical data**
Normal	30	59	-	-	-
Tumor	417	417	417	82	82
Platform	Illumina HM450	Illumina HiSeqV2	-	Illumina HM450	-

Identification of differentially methylated sites

The differentially expressed genes (DEGs) were firstly selected between tumor and normal tissues using edgeR package in R. The cutoff values were set at the FDR < 0.05 and |log2FC|>2. Then, we screened out the methylation sites corresponding to these DEGs and further test the differences in methylation levels between tumor and normal tissues to select differentially methylated sites (DMSs) by T-test with $p < 0.05$. Furthermore, pearson correlation analysis was performed to calculate the correlation between the methylation level of DMSs and expression level of corresponding DEGs. DMSs with significant negative correlation were retained.

Functional enrichment analysis

Functional annotations of DEGs containing DMSs were performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/), which enriched gene oncology and pathways. Three categories, including biological processes, molecular function and cellular components, were involved in Gene oncology (GO). Kyoto Encyclopedia of Genes and Genomes (KEGG, https://www.kegg.jp/) was used to carry out the pathway enrichment, which is an essential database resource for a deep understanding of functions and biological process from large-scale molecular

cohorts produced by high-throughput experimental technology. The criterion for significant enrichment was $p < 0.05$.

Protein-Protein Interaction (PPI) Network

To further explore the interaction among the DEGs, the Search Tool for the Retrieval of Interacting Genes (STRING, http://string-db.org/), a database containing all known and predicted protein interactions, was used to identify a PPI network of DEGs. Each interaction was evaluated by combined score ranged from 0 to 1. The higher the combined score, the more reliable the interaction. In present study, we used a strict combined score > 0.7 as the cut-off criterion to identify reliable interactions among the DEGs. The PPI network was visualized by Cytoscape software (version 3.7.0; www.cytoscape.org). Furthermore, the hub genes in PPI network were extracted using the cytoHubba application.

Construction of DMSs-based prognostic signature

The univariate Cox regression analysis was firstly performed to calculate the association between the methylation level of each DMS and patient's overall survival (OS) in training cohort. Those sites with P-values less than 0.05 were identified as prognosis-related DMSs. Then, using LASSO method to screen the prognosis-related DMSs and obtain an optimal model subsequently, the prognosis-related DMSs with coefficient not equal to 0 were retained as significant variables and a risk scoring model was established using the combination of weighted sites methylation values. The risk scores were calculated as shown in the following equation: Risk score = methylation of site 1 * β1 + methylation of site 2 * β2 +…methylation of site n * βn. βi is the regression coefficient of site i, which represents the contribution of site i to the prognostic risk score. Risk scores were assigned to LUAD patients in each cohort. Using the median risk score as the cutoff point, patients were divided into low-risk (risk score below the median value) or high-risk (risk score above the median value) group correspondingly.

Development of DMSs-based Nomogram

To translate the prognostic value of DMSs-based signature into clinical application, a nomogram, including the risk score and the clinical risk factors of LUAD patients evaluated by multivariate Cox proportional-hazards regression, was developed for predicting the 3- and 5-years OS in TCGA-LUAD cohort. The discriminatory ability of the nomogram was evaluated by calculating the concordance index (C-index), which is a measure of discrimination. Calibration plots were plotted to compare the observed and predicted probabilities for the nomogram.

Statistical Analysis

The multivariate Cox proportional-hazards regression model was used to evaluate the independent prognostic value of the signature after adjusting for age, gender and stage. Hazard ratios (HRs) and 95% confidence intervals (CIs) were computed based on the Cox regression analysis. Survival curves were estimated using the Kaplan–Meier method and were compared using the log-rank test. Fisher's exact test was used to observe the differences in mortality rate and lymph node metastasis rate between different

risk groups. Values of $p < 0.05$ were considered significant. All statistical analysis were performed using the R3.4.0.

Results

Identification of differentially methylated sites in LUAD

We initially performed differential expression analysis to select DEGs between LUAD and normal lung tissues in TCGA-LUAD dataset. With cut-off criteria of FDR < 0.05 and |log2FC| >2.0, a total of 960 DEGs were identified, including 653 up-regulated DEGs and 307 down-regulated DEGs (Fig. 1A). We then selected the methylation sites which were differentially methylated between LUAD and normal lung tissues and significantly negative correlated with the expression of corresponding DEGs. We thought that such methylation sites could influence the gene expression and further participate in tumor progression. The results showed that a total of 1362 DMSs corresponding to 471 DEGs were identified, including 752 hypermethylation sites and 610 hypomethylation sites (Fig. 1B).

Functional enrichment of DMGs

To further investigate the biological processes which the DMSs might be involved in, we performed GO annotation and KEGG pathway enrichment using DAVID database for the 471 corresponding DEGs. The DEGs were significantly enriched in many cancer-related pathways. The top significant terms emerging from the gene oncology enrichment analysis were shown in Fig. 2A. For instance, the most significant GO term, cell division, has been reported in multiple articles related to the progression and metastasis of cancer [15–17]. We also found that DEGs were significantly enriched in angiogenesis, which is a core hallmark of advanced cancers, especially in LUAD [18–20]. Besides, other significant GO terms, such as regulation of cell cycle and regulation of small GTPase mediated signal transduction were also related to cancer progression and chemoresistance reported in many studies [21, 22]. As shown in Fig. 2B, KEGG pathway enrichment analysis found twelve significantly enriched pathways related to cancer progression, such as PI3K-Akt signaling pathway [23, 24], ECM-receptor interaction [25, 26] and p53 signaling pathway [27, 28]. The results indicated that these DEGs played key roles in multiple cancer-related pathways, and further indicated that the DMSs might be involved in LUAD progression by regulating the corresponding gene expression.

Construction of PPI Network

Using STRING database, a PPI network was constructed to further explore the interactions between the 471 DEmRNAs. After removing unconnected nodes, the PPI network of DEGs is consisted of 188 nodes and 888 edges when combined score > 0.7 was set as the cutoff criterion (Fig. 3A). Furthermore, the top 10 hub genes, including cyclin dependent kinase 1 (CDK1), cyclin A2 (CCNA2), cyclin B1 (CCNB1), cell division cycle 20 (CDC20), cell division cycle associated 8 (CDCA8), aurora kinase B (AURKB), assembly factor for spindle microtubules (ASPM), PDZ binding kinase (PBK), ribonucleotide reductase regulatory

subunit M2 (RRM2) and centromere protein F (CENPF), were identified using the cytoHubba plugin for Cytoscape, with a higher degree of connectivity (Fig. 3B). Most of ten genes had been reported to be closely related to tumorigenesis and progression of LUAD.

Establishment of the DMSs-based prognostic signature

Performing the univariate Cox regression analysis, we identified DMSs with potential prognostic value in TCGA-LUAD cohort. Details of the clinical characteristics are presented in Supplementary Table S1. We found that 59 DMSs were significantly associated with overall survival, including 47 hypermethylation sites and 12 hypomethylation sites. The list of 59 DMSs is showed in Supplementary Table S2. Thus, these methylation sites were defined as prognosis-related DMSs to construct the prognostic signature. We used the glmnet package in R to perform LASSO regression analysis in TCGA-LUAD cohort. We obtained the optimal value of the parameter λ, which controlled the degree of LASSO regression complexity, and selected the significant variables through multiple cross-validation. We found that the parameter λ reached the optimal value, when the number of variables was three. Therefore, combining the regression coefficients of three DMSs under the optimal λ value, we constructed a three-DMSs risk score model to guide the prognosis of LUAD patients. The general information of the three DMSs is displayed in Table 2. The risk score formula was created as follows: Risk score = (1.0003*methylation level of cg21339084) + (0.1484*methylation level of cg07400091) + (-0.2536*methylation level of cg23843180). Calculating the risk score for each patient in TCGA-LUAD cohort, patients were classified into a high-risk or a low-risk group based on the median risk score. We found that the three-DMSs signature significantly stratified patients in terms of overall survival (log-rank p = 1.9E-04; Fig. 4A). Patients with high risk score had significantly shorter OS than those with low risk score. The mortality rate was 34.0% (71/209) in the high-risk group, significantly higher than 14.4% (30/208) in the low-risk group ($p < 0.001$, Fisher exact test; Fig. 4B). The risk score distribution, survival status, and methylation profile of the three prognostic DMSs are shown in Fig. 4C. As shown in Table 3, multivariate Cox regression analysis suggested that the three-DMSs signature was an independent prognostic factor, after adjusting for age, gender and stage (HR = 2.29, 95%CI: 1.47–3.57, P = 2.36E-04). Furthermore, noticing the patients with lymph node metastasis status, we found that patients in the high-risk group had a higher lymph node metastasis rate than those in the low-risk group (26.2% vs. 15.8%, p = 0.018, Fisher exact test; Fig. 4B). From the three DMSs, two were associated with high risk (cg21339084 and cg07400091; HR > 1) and one appeared to be protective (cg23843180; HR < 1). The methylation level of the three prognostic DMSs was detected and the differences between high- and low-risk groups were compared. We found that patients with high-risk scores tended to hypermethylation at risky sites, whereas patients in the low-risk group tended to hypomethylation at protective sites (Fig. 5A-5C).

Table 2
General information of the three DMSs

ProbeID	Gene	chrom	chromStart	chromEnd	coefficient
cg21339084	LIMS2	chr2	128422432	128422434	1.0003
cg07400091	S1PR1	chr1	101704472	101704474	0.1484
cg23843180	NGEF	chr2	233852838	233852840	-0.2536

Table 3
Univariate and multivariate Cox regression analysis in TCGA-LUAD and GSE56044 cohorts

Variables	Univariate analysis		Multivariate analysis	
	HR (95% CI)	*P*	HR (95% CI)	*P*
TCGA-LUAD cohort				
Age				
< = 60/>60	0.96(0.63–1.47)	0.852	1.19(0.77–1.85)	0.437
Gender				
Male/Female	0.94(0.64–1.40)	0.774	0.99(0.66–1.49)	0.964
Stage				
I + II/III + IV	2.74(1.82–4.13)	1.30e−06	2.79(1.85–4.21)	1.06e−06
Risk score				
Low/High	2.22(1.45–3.42)	2.77e−04	2.29(1.47–3.57)	2.36e−04
GSE56044 cohort				
Age				
< = 60/>60	2.83(1.26–6.35)	0.012	2.88(1.26–6.59)	0.012
Gender				
Male/Female	1.10(0.63–1.93)	0.737	0.79(0.44–1.43)	0.438
Risk score				
Low/High	2.15(1.20–3.85)	0.010	2.16(1.19–3.91)	0.011

Prognostic validation of the three-DMSs signature

An independent cohort (GSE56044), containing 82 LUAD patients with both methylation data and clinical information, was used to validate the prognosis performance of the three-DMSs signature. Similarly, we

calculated the risk score for each patient using the three-DMSs signature, after which patients were classified into a high-risk (n = 41) or a low-risk (n = 41) group based on the median risk score. We found that patients in high-risk group had a shorter survival time than those in low-risk group (HR = 2.15, 95% CI: 1.20–3.85, log-rank p = 0.008, Fig. 6A). Furthermore, we calculated the mortality rate in each risk group. The result showed that the mortality rate in high-risk group was 32% higher than that in low-risk group (p = 0.006, Fisher exact test; Fig. 6B). The risk score distribution, survival status, and expression profile of the three prognostic DMSs are shown in Fig. 6C. As biased stage information, the stage variable is excluded when performed multivariate Cox regression analysis. In accordance with the result of training set, the multivariate Cox regression analysis confirmed that the three-DMSs signature was significantly correlated with overall survival as an independent prognostic factor (HR = 2.16, 95% CI: 1.19–3.91, P = 0.011, Table 3).

Construction of three-DMSs signature-based nomogram

Multivariate Cox analysis indicated that three variables (age, stage, and three-DMSs risk score) were independent risk factors for OS. Thus, A nomogram predicting 3- and 5-years OS was constructed based on the multivariate analysis data. As shown in Fig. 7, the total points for a patient can be obtained by adding the points from each independent prognostic factor listed in the nomogram. C-indexes for the nomogram were 0.71 (95%CI: 0.58–0.85) and 0.70 (95%CI: 0.52–0.88) in TCGA-LUAD and GSE56044 cohorts, respectively. The calibration plots for the probabilities of 3 and 5-year OS indicated no apparent departure from the ideal line, showing good agreement between the nomogram-predicted OS and actual OS of LUAD patients in both the training and validation cohorts (Fig. 8). Such results indicated that the three-DMSs signature-based nomogram could provide insight into regarding survival prediction and serve as a clinically available guide for personalized treatment of LUAD patients.

Discussion

As the most predominant subtype of NSCLC, LUAD is pathological and molecular heterogeneous, making the prediction of patient outcome very urgent [29, 30]. Aberrant DNA methylation is of considerable importance in LUAD onset and progression [31, 32]. A special focus on DNA methylation alterations to develop the prognostic and predictive signatures for LUAD patients would be meaningful for survival prediction, guiding the personalized treatment decisions. Zheng et al. [11] constructed a CpG-based signature for survival prediction of lung adenocarcinoma patients based on TCGA database. However, such studies were limited by either small sample size or lack of validation of the signature as an independent prognostic factor. Therefore, in-depth studies on the LUAD progressive mechanisms, identification of specific methylation CpG sites and construction of the robust prognostic signatures are urgently required.

In the present study, we screened the DMSs that significantly correlated with corresponding gene expression, which may be involved in cancer progression by regulating the gene expression. Thus, a three-DMSs methylation signature significantly associated with the OS of LUAD patients was constructed based on genome-wide DNA methylation profiles using the Cox regression and LASSO analyses. The

three-DMSs signature performed well in classifying patients into a high-risk or a low-risk group with significant survival difference. Furthermore, a nomogram was developed by combing the DMSs-based prognostic signature with clinical risk factors, which could provide a clinically available and robust guide for survival prediction and personalized treatment of LUAD patients.

Our study showed that three DMSs within prognostic signature had a critical role in progression and metastasis of LUAD. The three DMSs, including cg21339084, cg07400091 and cg23843180, correspond to LIMS2, S1PR1 and NGEF respectively. LIMS2, also known as PINCH2, functions as a natural regulator of the LIMS1-ILK-parvin complex formation and is associated with cell spreading and migration via integrins at focal adhesions. Many researches had demonstrated that frequent inactivation of LIMS2 might be important in tumor progression events [33, 34]. Kim et al. found the frequent epigenetic silencing of LIMS2 in gastric tumors [35]. Loss of expression of LIMS2 was significantly correlated with an increase of the CpG island hypermethylation. Moreover, the transient transfection of LIMS2-siRNA significantly stimulated cell migration in gastric cancer cells but had no effects on cell growth, suggesting that epigenetic alteration of LIMS2 in gastric cancer may be involved in invasion and metastasis. In our study, we also found that LIMS2 was hypermethylation in LUAD patients. Patients in high-risk group classified by the three-DMSs signature had significantly higher methylation level than those in low-risk group. Such results indicated that aberrant methylation of LIMS2 might also play a vital role in LUAD progression and metastasis. S1PR1 is a member of the G-protein-coupled receptors for Sphingosine-1-phosphate and a biologically active metabolite of sphingolipid. In endothelial cells, S1PR1 played key roles in the morphogenic differentiation of vascular endothelial cells and in angiogenesis. In the immune system, S1PR1 can recruit immune cells to be reintroduced into the lymphatic circulation, especially by guiding activated T cells from nonlymphoid tissues into the lymphatics [36]. Recent studies demonstrated that the expression patterns of S1PR1 are altered in several tumors. Many researchers found that a positive regulatory interaction with signal transducer and activator of transcription 3 (STAT3) is involved in cancer tumorigenesis, metastasis and chemoresistance, such as gastric cancer [37], esophageal squamous cell carcinoma [38] and colorectal cancer [39]. In our study, we found that the expression of S1PR1 was significantly lower in LUAD patients, regulated by hypermethylation at cg07400091. Moreover, the methylation of S1PR1 in high-risk patients was significantly higher than that in low-risk patients. These results indicated the special role of S1PR1 hypermethylation in promoting LUAD progression, but the underlying mechanisms need further experimental verification. NGEF is a novel member of the family of Dbl genes and functions as a guanine nucleotide exchange factor for the Rho-type GTPases. Although preliminary experiments had showed that NGEF had transforming potential in cell culture and was able to induce tumors in nude mice [40], more studies need to be carried out in order to explore the role in carcinogenesis in the future. Interestingly, in our study, NEGF was significantly hypo-methylated in high-risk patients, suggesting that aberrant NEGF methylation might play more important role in LUAD progression and prognosis by elevating the NEGF expression.

Conclusion

Analyzing methylation and expression data comprehensively, our study identified a robust three-DMSs prognostic signature, which was significantly associated with the OS of LUAD patients. Furthermore, a nomogram was developed by combing the three-DMSs prognostic signature with clinical risk factors, which could provide a clinically available and robust guide for survival prediction and personalized treatment of LUAD patients. Further studies on the functional mechanism of the three DMSs could be carried out, which might provide helpful guidance for LUAD therapy as promising therapeutic targets in the near future.

Abbreviations

LUAD: Lung adenocarcinoma; GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; DMSs: Differentially methylated sites; DEGs: Differentially expressed genes; GO: gene oncology; KEGG: Kyoto Encyclopaedia of Genes and Genomes; LASSO: The least absolute shrinkage and selection operator; OS: overall survival; HR: Hazard ratios; CI: Confidence interval.

Declarations

Acknowledgments

None.

Funding

Not applicable.

Consent for publication

Not applicable.

Authors' contributions

Xuelong Wang, Bin Zhou, Yuxin Xia and Chengwei Zhang conceived and designed the study. Yanchao Liu, Xin Bi and Xiong Luo downloaded and screened the data. Chengwei Zhang and Jianxin Zuo edited the manuscript. All authors read and approved the final manuscript.

Data availability statement

The data used and analyzed during the current study are available from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) and Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/).

Ethics approval and consent to participate

Approval from the ethical board for this study was not required because of the public nature of all the data.

Competing interests

The authors declare that they have no competing interests.

References

1. Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, Paz-Ares L. Lung cancer: current therapies and new targeted treatments. Lancet. 2017;389(10066):299–311.
2. Zhou C. Lung cancer molecular epidemiology in China: recent trends. Transl Lung Cancer Res. 2014;3(5):270–9.
3. Chansky K, Sculier JP, Crowley JJ, Giroux D, Van Meerbeeck J, Goldstraw P, International Staging C, Participating I. [The International Association for the Study of Lung Cancer Staging Project. Prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer]. Zhongguo Fei Ai Za Zhi. 2010;13(1):9–18.
4. Gao C, Zhuang J, Li H, Liu C, Zhou C, Liu L, Sun C. Exploration of methylation-driven genes for monitoring and prognosis of patients with lung adenocarcinoma. Cancer Cell Int. 2018;18:194.
5. Song X, Zhao C, Jiang L, Lin S, Bi J, Wei Q, Yu L, Zhao L, Wei M. High PITX1 expression in lung adenocarcinoma patients is associated with DNA methylation and poor prognosis. Pathol Res Pract. 2018;214(12):2046–53.
6. Su C, Shi K, Cheng X, Han Y, Li Y, Yu D, Liu Z. Methylation of CLEC14A is associated with its expression and lung adenocarcinoma progression. J Cell Physiol. 2019;234(3):2954–62.
7. Zhang R, Lai L, Dong X, He J, You D, Chen C, Lin L, Zhu Y, Huang H, Shen S, et al. SIPA1L3 methylation modifies the benefit of smoking cessation on lung adenocarcinoma survival: an epigenomic-smoking interaction analysis. Mol Oncol. 2019;13(5):1235–48.
8. 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.
9. Seok Y, Lee WK, Park JY, Kim DS. TGFBI Promoter Methylation is Associated with Poor Prognosis in Lung Adenocarcinoma Patients. Mol Cells. 2019;42(2):161–5.
10. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, Moran S, Heyn H, Vizoso M, Gomez A, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. J Clin Oncol. 2013;31(32):4140–7.
11. Zheng R, Xu H, Mao W, Du Z, Wang M, Hu M, Gu X. A novel CpG-based signature for survival prediction of lung adenocarcinoma patients. Exp Ther Med. 2020;19(1):280–6.
12. Wang R, Zhu H, Yang M, Zhu C. DNA methylation profiling analysis identifies a DNA methylation signature for predicting prognosis and recurrence of lung adenocarcinoma. Oncol Lett. 2019;18(6):5831–42.

13. Wang Y, Wang Y, Wang Y, Zhang Y. Identification of prognostic signature of non-small cell lung cancer based on TCGA methylation data. Sci Rep. 2020;10(1):8575.
14. Karlsson A, Jonsson M, Lauss M, Brunnstrom H, Jonsson P, Borg A, Jonsson G, Ringner M, Planck M, Staaf J. Genome-wide DNA methylation analysis of lung carcinoma reveals one neuroendocrine and four adenocarcinoma epitypes associated with patient outcome. Clin Cancer Res. 2014;20(23):6127–40.
15. Lopez-Lazaro M. The stem cell division theory of cancer. Crit Rev Oncol Hematol. 2018;123:95–113.
16. Shostak A. **Circadian Clock, Cell Division, and Cancer: From Molecules to Organism**. *Int J Mol Sci* 2017, 18(4).
17. Xia Z, Ou-Yang W, Hu T, Du K. Prognostic significance of CDC25C in lung adenocarcinoma: An analysis of TCGA data. Cancer Genet. 2019;233–234:67–74.
18. Cong Z, Diao Y, Li X, Jiang Z, Xu Y, Zhou H, Qiang Y, Wu H, Shen Y. Long non-coding RNA linc00665 interacts with YB-1 and promotes angiogenesis in lung adenocarcinoma. Biochem Biophys Res Commun. 2020;527(2):545–52.
19. Frezzetti D, Gallo M, Maiello MR, D'Alessio A, Esposito C, Chicchinelli N, Normanno N, De Luca A. VEGF as a potential target in lung cancer. Expert Opin Ther Targets. 2017;21(10):959–66.
20. Popper HH. Progression and metastasis of lung cancer. Cancer Metastasis Rev. 2016;35(1):75–91.
21. Chen B, Huang S, Pisanic Ii TR, Stark A, Tao Y, Cheng B, Li Y, Wei Y, Zhao W, Wang TH, et al. Rab8 GTPase regulates Klotho-mediated inhibition of cell growth and progression by directly modulating its surface expression in human non-small cell lung cancer. EBioMedicine. 2019;49:118–32.
22. Liu M, Zhang H, Li Y, Wang R, Li Y, Zhang H, Ren D, Liu H, Kang C, Chen J. HOTAIR, a long noncoding RNA, is a marker of abnormal cell cycle regulation in lung cancer. Cancer Sci. 2018;109(9):2717–33.
23. Fumarola C, Bonelli MA, Petronini PG, Alfieri RR. Targeting PI3K/AKT/mTOR pathway in non small cell lung cancer. Biochem Pharmacol. 2014;90(3):197–207.
24. Zhu HE, Yin JY, Chen DX, He S, Chen H. Agmatinase promotes the lung adenocarcinoma tumorigenesis by activating the NO-MAPKs-PI3K/Akt pathway. Cell Death Dis. 2019;10(11):854.
25. Bao Y, Wang L, Shi L, Yun F, Liu X, Chen Y, Chen C, Ren Y, Jia Y. Transcriptome profiling revealed multiple genes and ECM-receptor interaction pathways that may be associated with breast cancer. Cell Mol Biol Lett. 2019;24:38.
26. Yeh MH, Tzeng YJ, Fu TY, You JJ, Chang HT, Ger LP, Tsai KW. Extracellular Matrix-receptor Interaction Signaling Genes Associated with Inferior Breast Cancer Survival. Anticancer Res. 2018;38(8):4593–605.
27. Zeng SG, Lin X, Liu JC, Zhou J. Hypoxiainduced internalization of connexin 26 and connexin 43 in pulmonary epithelial cells is involved in the occurrence of nonsmall cell lung cancer via the P53/MDM2 signaling pathway. Int J Oncol. 2019;55(4):845–59.
28. Zhong G, Chen X, Fang X, Wang D, Xie M, Chen Q. Fra-1 is upregulated in lung cancer tissues and inhibits the apoptosis of lung cancer cells by the P53 signaling pathway. Oncol Rep.

2016;35(1):447–53.
29. Chen Z, Fillmore CM, Hammerman PS, Kim CF, Wong KK. Non-small-cell lung cancers: a heterogeneous set of diseases. Nat Rev Cancer. 2014;14(8):535–46.
30. Greulich H. The genomics of lung adenocarcinoma: opportunities for targeted therapies. Genes Cancer. 2010;1(12):1200–10.
31. Duruisseaux M, Esteller M. Lung cancer epigenetics: From knowledge to applications. Semin Cancer Biol. 2018;51:116–28.
32. Mehta A, Dobersch S, Romero-Olmedo AJ, Barreto G. Epigenetics in lung cancer diagnosis and therapy. Cancer Metastasis Rev. 2015;34(2):229–41.
33. Park CH, Rha SY, Ahn JB, Shin SJ, Kwon WS, Kim TS, An S, Kim NK, Yang WI, Chung HC. PINCH-2 presents functional copy number variation and suppresses migration of colon cancer cells by paracrine activity. Int J Cancer. 2015;136(10):2273–83.
34. Xu H, Cao H, Xiao G. Signaling via PINCH: Functions, binding partners and implications in human diseases. Gene. 2016;594(1):10–5.
35. Kim SK, Jang HR, Kim JH, Noh SM, Song KS, Kim MR, Kim SY, Yeom YI, Kim NS, Yoo HS, et al. The epigenetic silencing of LIMS2 in gastric cancer and its inhibitory effect on cell migration. Biochem Biophys Res Commun. 2006;349(3):1032–40.
36. Baeyens A, Fang V, Chen C, Schwab SR. Exit Strategies: S1P Signaling and T Cell Migration. Trends Immunol. 2015;36(12):778–87.
37. Song S, Min H, Niu M, Wang L, Wu Y, Zhang B, Chen X, Liang Q, Wen Y, Wang Y, et al. S1PR1 predicts patient survival and promotes chemotherapy drug resistance in gastric cancer cells through STAT3 constitutive activation. EBioMedicine. 2018;37:168–76.
38. Liu Y, Zhi Y, Song H, Zong M, Yi J, Mao G, Chen L, Huang G. S1PR1 promotes proliferation and inhibits apoptosis of esophageal squamous cell carcinoma through activating STAT3 pathway. J Exp Clin Cancer Res. 2019;38(1):369.
39. Lin Q, Ren L, Jian M, Xu P, Li J, Zheng P, Feng Q, Yang L, Ji M, Wei Y, et al. The mechanism of the premetastatic niche facilitating colorectal cancer liver metastasis generated from myeloid-derived suppressor cells induced by the S1PR1-STAT3 signaling pathway. Cell Death Dis. 2019;10(10):693.
40. Rodrigues NR, Theodosiou AM, Nesbit MA, Campbell L, Tandle AT, Saranath D, Davies KE. Characterization of Ngef, a novel member of the Dbl family of genes expressed predominantly in the caudate nucleus. Genomics. 2000;65(1):53–61.

Figures

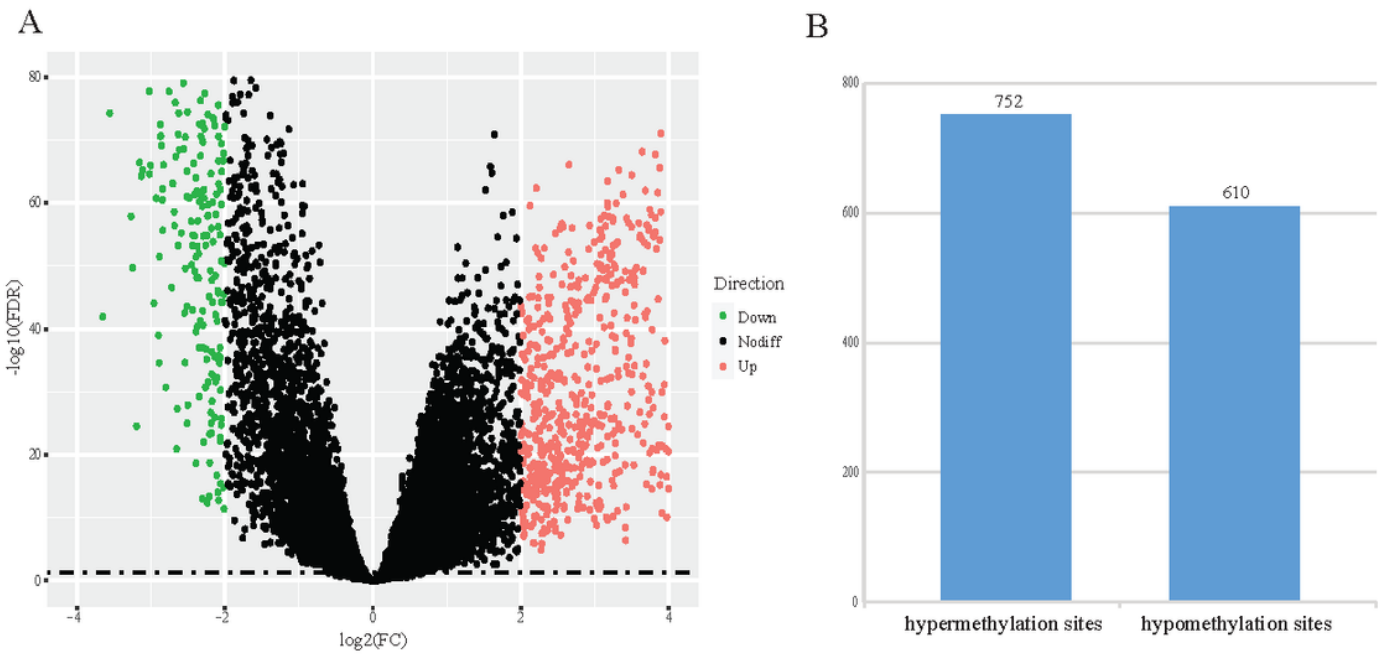

Figure 1

Identification of differentially expressed genes and differentially methylated sites. (A) Volcano plot of differentially expressed genes. (B) Histogram of differentially methylated sites.

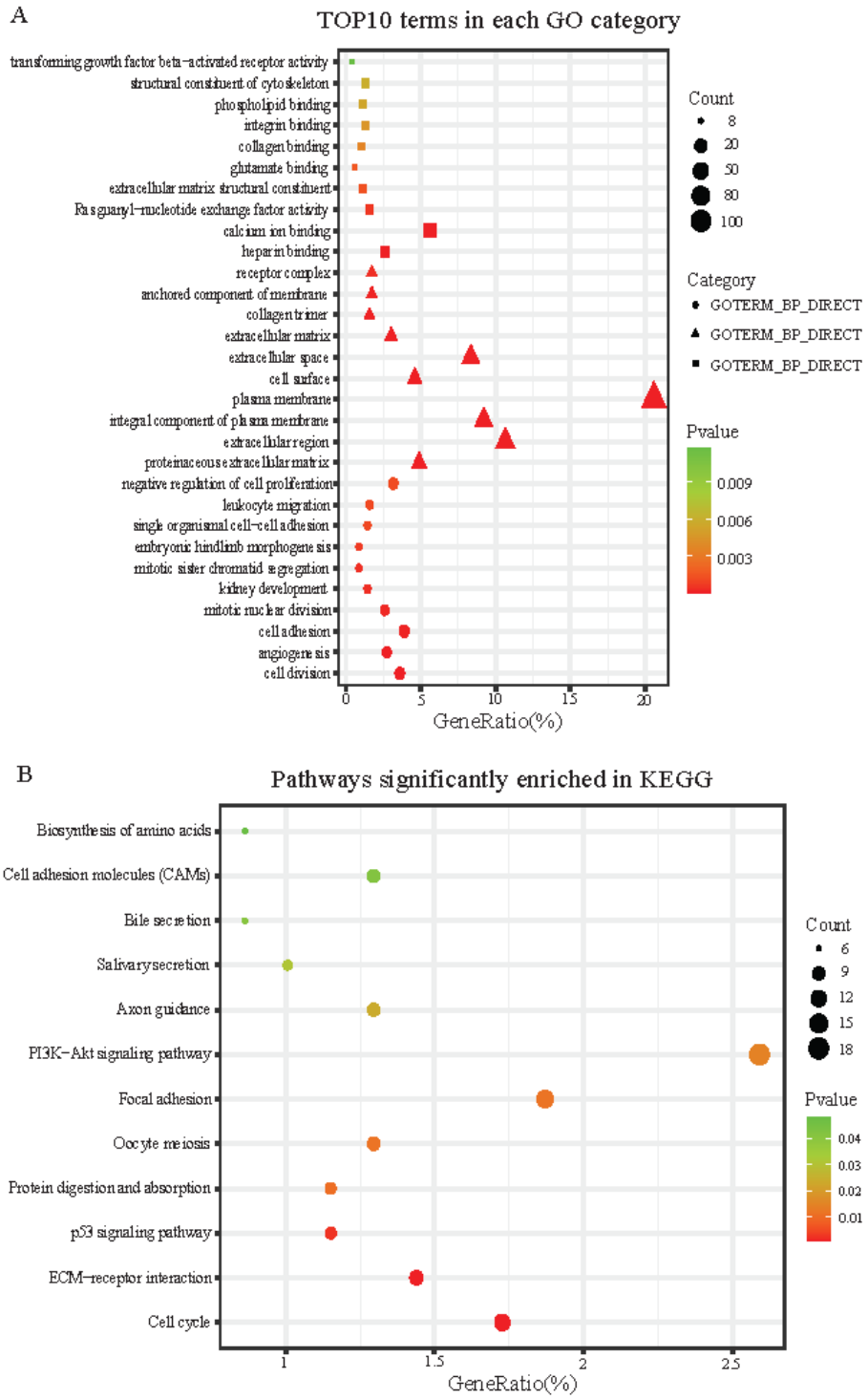

Figure 2

Functional enrichment of differentially expressed genes corresponding to differentially methylated sites. (A) Top ten terms significantly enriched in each Gene Ontology (GO) category. (B) Pathways significantly enriched in Kyoto Encyclopedia of Genes and Genomes (KEGG).

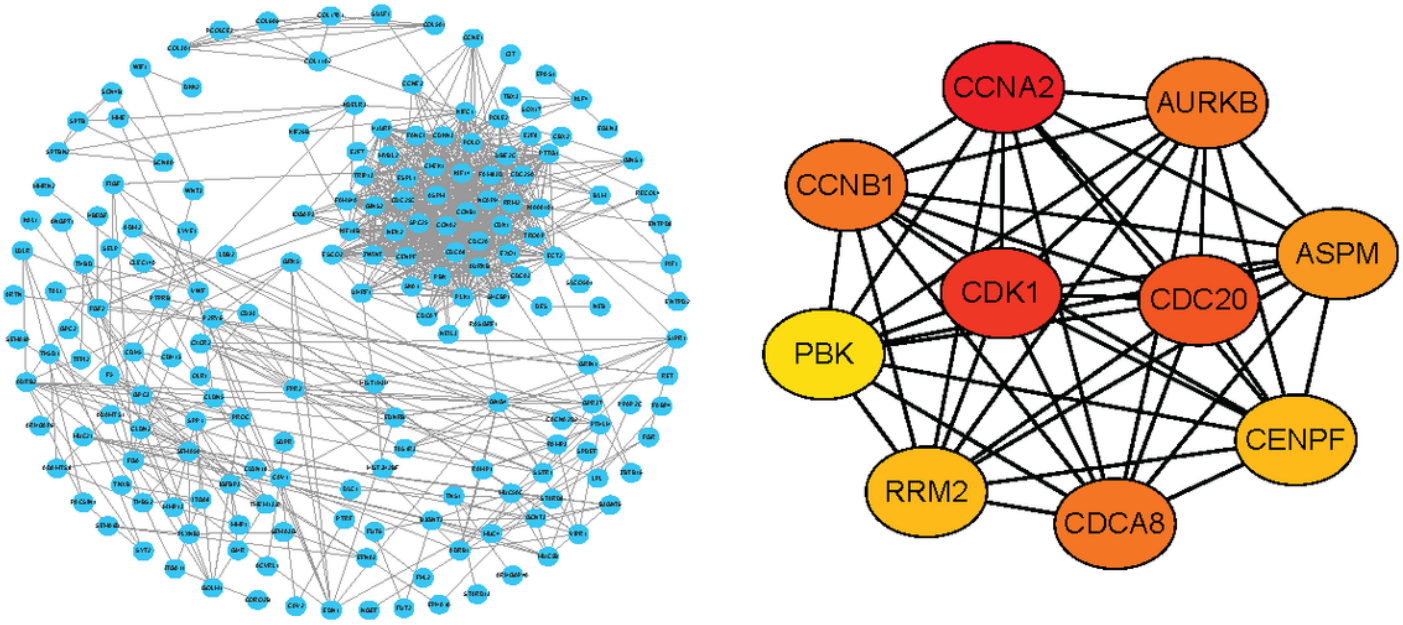

Figure 3

Construction of protein-protein interaction (PPI) network and Identification of hub genes. (A) PPI network. (B) Ten hub genes extracted from PPI network.

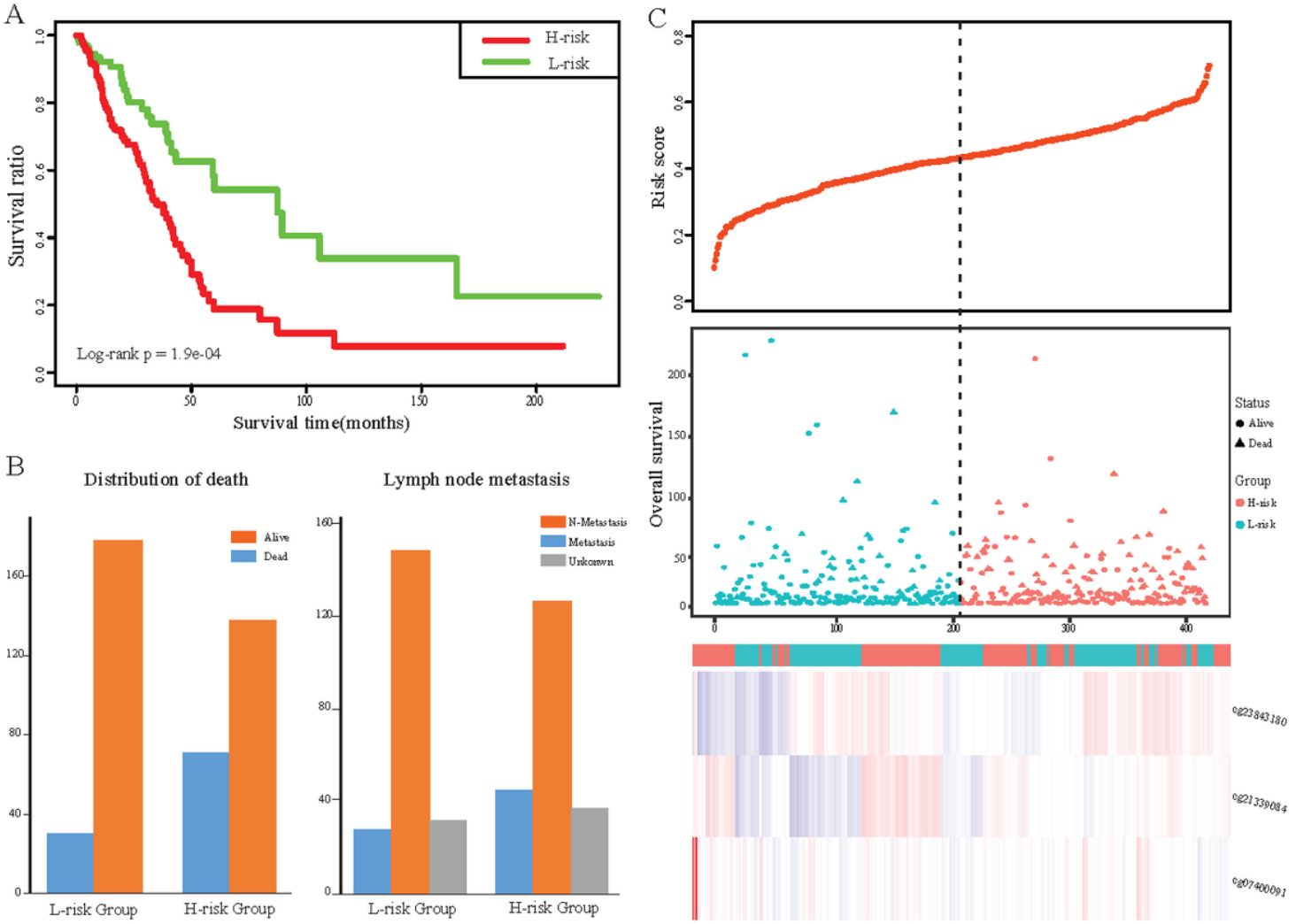

Figure 4

Construction of the three-DMSs prognostic signature in TCGA-LUAD cohort. (A) Kaplan-Meier curve of the overall survival for high-risk and low-risk scores ranking by the three-DMSs prognostic signature. (B) The distribution of death and lymphatic node metastasis in high-risk and low-risk groups respectively. (C) Risk score distribution, survival status and methylation heat map of three DMSs corresponding to each sample above.

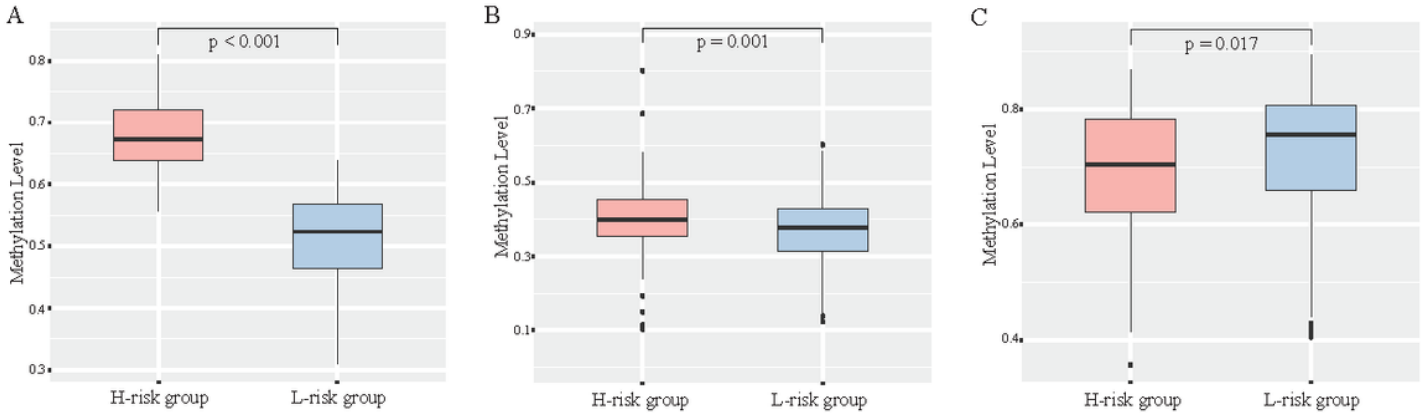

Figure 5

Methylation difference of three differentially methylated sites between high-risk and low-risk groups respectively. (A) cg21339084. (B) cg07400091. (C) cg23843180.

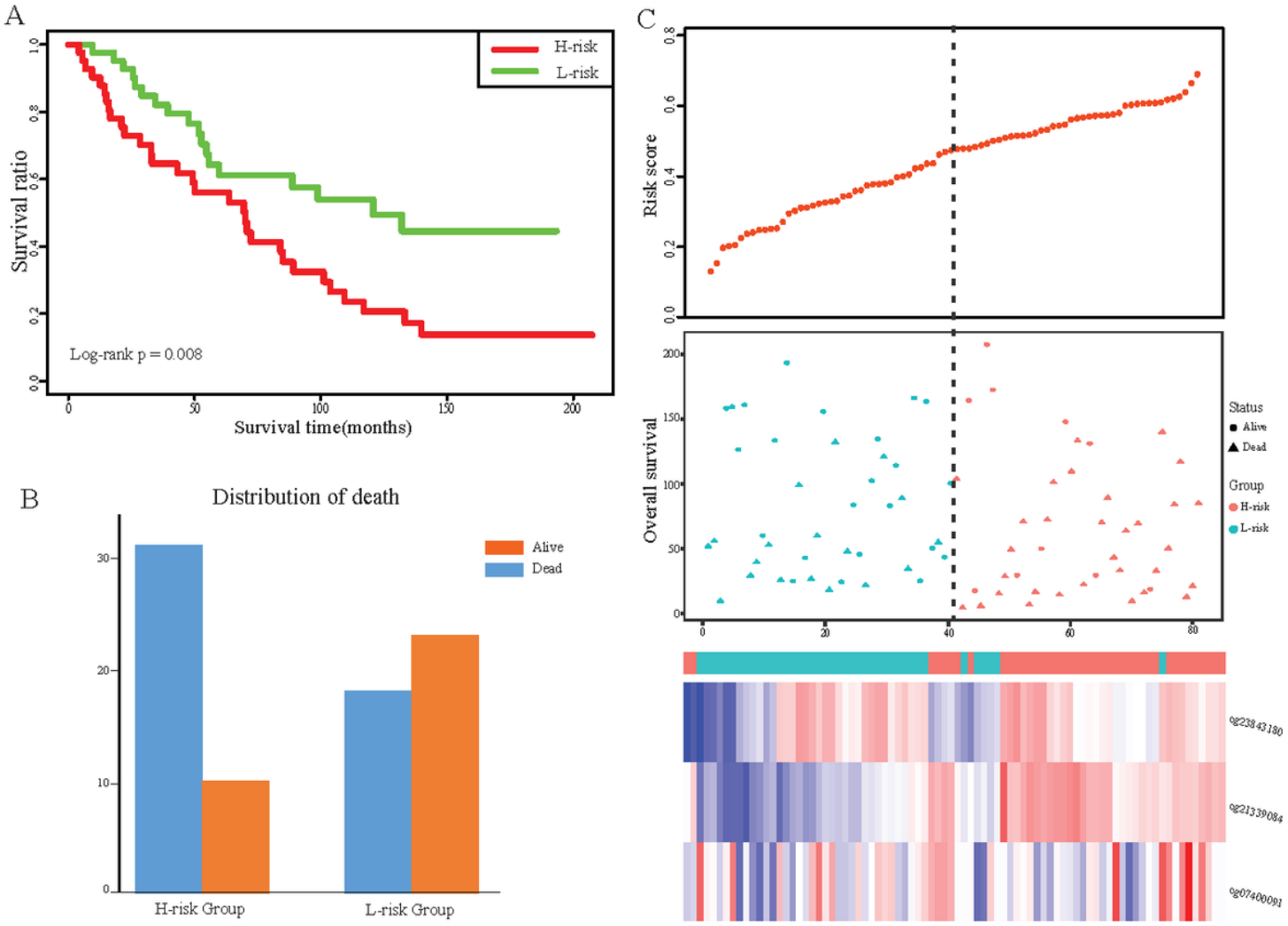

Figure 6

Validation of the three-DMSs prognostic signature in an independent cohort. (A) Kaplan-Meier curve of the overall survival for high-risk and low-risk scores ranking by the three-DMSs prognostic signature. (B) The distribution of death in high-risk and low-risk groups. (C) Risk score distribution, survival status and methylation heat map of three DMSs corresponding to each sample above.

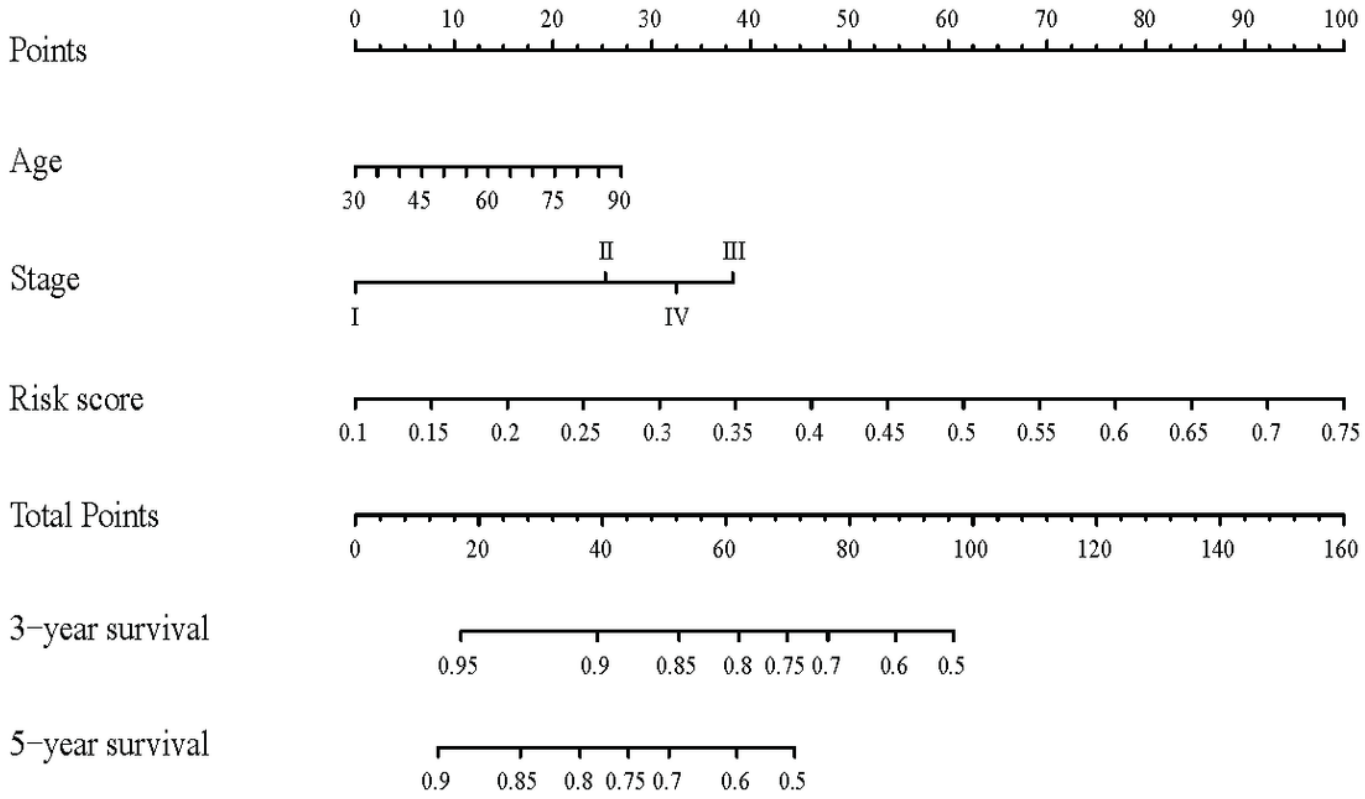

Figure 7

A nomogram for the prediction of 3- and 5-years overall survival in LUAD patients.

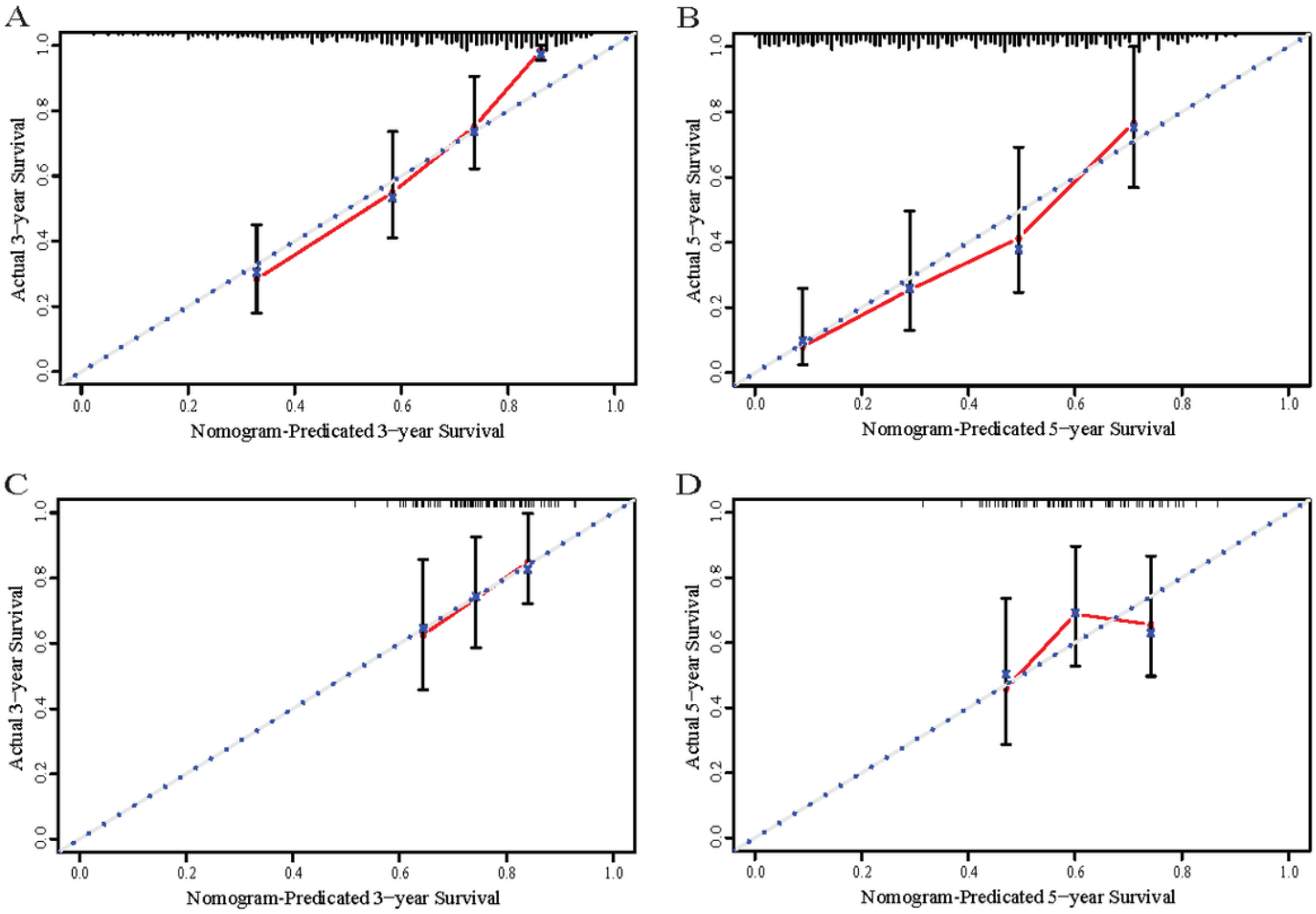

Figure 8

Calibration curves of the nomogram for 3- and 5-years overall survival predictions in TCGA-LUAD cohort (A,B) and GSE56044 cohort (C,D).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- tableS2.docx
- tableS1.docx