IVYSPRING
INTERNATIONAL PUBLISHER

International Journal of Biological Sciences

Research Paper

Diagnostic, progressive and prognostic performance of m6A methylation RNA regulators in lung adenocarcinoma

Zhizhi Zhuang[1*], Liping Chen[2*], Yuting Mao[3*], Qun Zheng[1], Huiying Li[4], Yueyue Huang[5], Zijing Hu[6], Yi Jin[1✉]

1. Department of Rheumatology and Immunology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China.
2. Department of Pharmacy, Sir Run Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310000, China.
3. Second clinical college of medicine, Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China.
4. Department of Respiratory medicine, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China.
5. Department of Hematology and Medical Oncology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China.
6. College of Pharmaceutical Sciences, Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China.

*These authors contributed equally to this work.

✉ Corresponding author: Yi Jin. Department of Rheumatology and Immunology, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang Province, China. Tel: +86-0577-88002611. E-mail: jinyi1990@wzhealth.com.

Abstract

Background: N6-methyladenosine (m6A) RNA methylation is dynamically and reversibly regulated by methyl-transferases ("writers"), binding proteins ("readers"), and demethylases ("erasers"). The m6A is restored to adenosine and thus to achieve demethylation modification. The abnormality of m6A epigenetic modification in cancer has been increasingly attended. However, we are rarely aware of its diagnostic, progressive and prognostic performance in lung adenocarcinoma (LUAD).

Methods and Results: The expression of 13 widely reported m6A RNA regulators in LUAD and normal samples were systematically analyzed. There were 12 m6A RNA methylation genes displaying aberrant expressions, and an 11-gene diagnostic score model was finally built (Diagnostic score =0.033*KIAA1429+0.116*HNRNPC+0.115*RBM15-0.067* METTL3-0.048*ZC3H13-0.221*WTAP+ 0.213*YTHDF1-0.132*YTHDC1-0.135* FTO+0.078*YTHDF2+0.014*ALKBH5). Receiver operating characteristic (ROC) analysis was performed to demonstrate superiority of the diagnostic score model (Area under the curve (AUC) was 0.996 of training cohort, P<0.0001; AUC was 0.971 of one validation cohort-GSE75037, P<0.0001; AUC was 0.878 of another validation cohort-GSE63459, P<0.0001). In both training and validation cohorts, YTHDC2 was associated with tumor stage (P<0.01), while HNRNPC was up expressed in progressed tumor (P<0.05). Besides, WTAP, RBM15, KIAA1429, YTHDF1, and YTHDF2 were all up expressed for TP53 mutation. Furthermore, using least absolute shrinkage and selection operator (lasso) regression analysis, a ten-gene risk score model was built. Risk score=0.169*ALKBH5-0.159*FTO+0.581*HNRNPC-0.348* YTHDF2-0.265*YTHDF1-0.123*YTHDC2 +0.434*RBM15+0.143*KIAA1429-0.200*WTAP-0.310*METTL3. There existed correlation between the risk score and TNM stage (P<0.01), lymph node stage (P<0.05), gender (P<0.05), living status (P<0.001). Univariate and multivariate Cox regression analyses of relevant clinicopathological characters and the risk score revealed risk score was an independent risk factor of lung adenocarcinoma (HR: 2.181, 95%CI (1.594-2.984), P<0.001). Finally, a nomogram was built to facilitate clinicians to predict outcome.

Conclusions: m6A epigenetic modification took part in the progression, and provided auxiliary diagnosis and prognosis of LUAD.

Key words: N6-methyladenosine, diagnostic score model, risk score, clinicopathological characters, lung adenocarcinoma

Introduction

According to Global Cancer Statistics 2018, there will have been approximately 18.1 million new cancer cases and 9.6 million death cases worldwide [1]. Researchers around the world are constantly striving to improve the medical technology for providing more sensitive method for tumor diagnosis and effective approach for tumor therapy. However, due to the complexity of tumor formation mechanism, it is far from enough to understand the nature of cancer from the genetic layer in the traditional sense [2]. It is recognized that the expression of oncogene depends not only on the gene itself, but also on epigenetic modification without changing gene sequence [2,3].

Epigenetics is a research hotspot in recent years, which is defined as no change in DNA sequence, but with heritable changes in gene expression [3]. Previously, the epigenetic researchers are mainly focused on DNA and histone modifications. It was even believed that, for example, mRNA only played the role in information transmission. However, with the rapid development of high-throughput sequencing technology, it is found that mRNA also undergo various modifications such as N6-methyladenosine (m^6A), N1-methyladenosine (m^1A) and pseudouridine methylation during the process of exon splicing, 5'-capping and 3'-tailing [4-6]. These modifications will then affect mRNA splicing, nucleation, stabilization, translation and other mRNA metabolism processes, thereby regulating gene expression. Up to now, 171 RNA modifications have been discovered [7]. Researches on m^6A epigenetic modification have been increasingly attended. M^6A is a methylation modification that occurs on RNA adenine (A), and is one of the most abundant modifications in most eukaryotic mRNAs and long-chain non-coding RNAs (lncRNAs) [8-10]. In transfer RNA (tRNA), ribosome RNA (rRNA), microRNA, the m^6A methylation is likewise detected [11]. Similar to DNA and protein modification, the m^6A methylation is dynamically and reversibly regulated by methyl-transferases ("writers"), binding proteins ("readers"), and demethylases ("erasers") [12]. The RNAs undergo the process of methyl group modification under the action of "writers" that such as METTL3, METTL14, WTAP, RBM15, KIAA1429, and ZC3H13 [13-19]. Then the "readers" including YTHDF1, YTHDF2, YTHDC1, YTHDC2, HNRNPC recognize those m^6A-modified RNAs and function in RNA processing, nuclear export, translation and decay [12,17,20]. Relying on the "Erasers" (FTO, ALKBH5), the m^6A is restored to adenosine and thus to achieve demethylation modification [21]. Once the component involved in the regulation of m^6A modification has been lost, the physiological functions such as cell differentiation and embryo development would be affected and expression of genes would be abnormally regulated [4,22].

At present, some diseases, such as obesity, diabetes, neuronal disease, infertility, autoimmune diseases, have been reported associated with m^6A modification [23-28]. In addition, m^6A methylation is likewise closely related to the tumor development. M^6A-related protein is an important regulator of tumorigenesis, high or low expression level often directly determines the tumor pathological process. For example, in cervical squamous cell carcinoma (CSCC), FTO is significantly high expressed [29]. It would lower the m^6A methyl level and in turn activate the beta-catenin pathway and affect ERCC1 expression and then lead to poor prognosis. In colorectal cancer, YTHDF1 is highly expressed associated with the tumor diameter (P=0.009), lymph node metastasis (P=0.044), distant metastasis (P=0.036) and clinical stage (P=0.0226) [30]. As for lung cancer, report has indicated that FTO enhances the expression of myeloid zinc finger 1 (MZF1) by reducing m^6A levels, thereby promoting the development of lung squamous cell carcinoma (LUSC) [31]. Another evidence points that miR-33a inhibits the proliferation of non-small cell lung cancer (NSCLC) cells by targeting METTL3 mRNA 3'UTR binding sites [32]. Thus, the fully understanding of the pivotal m^6A RNA methylation regulators is vital for lung cancer treatment. In all categories of lung cancers, lung adenocarcinoma (LUAD) is one of the most common histological types [33]. However, it lacks a comprehensive analysis of the expression of m^6A RNA methylation regulators in lung cancer, especially in LUAD. The clinicopathological characteristics, diagnosis and prognostic value of such regulators remains to be explored.

In this study, the expression of 13 widely reported m^6A RNA regulators in LUAD and normal samples were systematically analyzed. The different clinicopathological characters which related to each m^6A modification regulator were then provided. It was found that the expression of m^6A RNA methylation regulators played a responsible role in the progression of LUAD, which even were identified as the effective diagnostic and prognostic factors.

Methods

Datasets

Genotype-Tissue Expression (GTEx) dataset for normal lung tissues was downloaded from https://www.gtexportal.org/ [34]. Data of lung adenocarcinoma and squamous cell lung carcinoma of

TCGA [35] was downloaded from UCSC Xena (https://xena.ucsc.edu/) [36]. For GTEx and TCGA dataset, RNA-sequencing data (FPKM values) were normalized into $\log_2$ (FPKM+1). Clinical data of TCGA was downloaded from cBioPortal (http://www.cbioportal.org/) [37,38]. Data of GSE75037 [39], GSE63459 [40], GSE29013 [41], GSE30219 [42], GSE37745 [43], and GSE50081 [44] were downloaded from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/).

After normalization by "LiMMA-normalize BetweenArrays" (http://www.bioconductor.org/), GSE75037 and GSE63459 were served as validation cohorts of diagnostic model. GSE29013, GSE30219, GSE37745, and GSE50081 were merged by Batch normalization to serve as validation cohorts of prognostic risk model. The details of expression data were shown in Table 1. GTEx and TCGA dataset were acted as training cohort, and GSE75037, GSE63459, GSE29013, GSE30219, GSE37745, and GSE50081 as validation cohort. For GTEx and TCGA dataset, we defined expression value as $\log_2$ (FPKM+1), while for gene chips, expression value was defined as $\log_2$(expression).

Selection of investigative genes

There were 13 m⁶A RNA methylation genes brought into the study, namely, METTL3, METTL14, WTAP, KIAA1429, RBM15, ZC3H13, YTHDC1, YTHDC2, YTHDF1, YTHDF2, HNRNPC, FTO, and ALKBH5 [45,46]. We investigated their role in diagnosis, progression, and prognosis of lung adenocarcinoma of both training and validation cohort.

Diagnostic role of 13 m⁶A RNA methylation genes in lung adenocarcinoma

After normalization by "LiMMA-normalize BetweenArrays", 288 samples of normal lung tissues from GTEx were added into TCGA dataset of lung cancer to increase the number of the normal group. Then wilcox test was applied to differential analysis of 13 m⁶A RNA methylation genes between normal and tumorous samples. Package "corrplot" was used to analyze the correlation among each gene. App ClueGO (http://www.ici.upmc.fr/cluego/version2. 3.3) of Cytoscape (http://cytoscape.org, version3.5.1) was to illuminate the relationship between m⁶A RNA methylation genes and related pathways [47-48]. Then the least absolute shrinkage and selection operator (lasso) regression model was made to diagnose lung adenocarcinoma in the training cohort by the "lars" package [49-50]. Eventually, 11 m⁶A RNA methylation related genes and coefficients were identified with smallest 10-fold cross validated mean square error in the training cohort.

$$\text{Diagnostic score} = \sum_{i=1}^{n} \text{Coefi*Xi}$$

(Coefi is the coefficient of each selected gene, Xi is the expression value.)

The diagnostic score model was further validated in the independent cohorts-GSE75037 and GSE63459. Besides, the distributions of diagnostic score were shown in different tumor stage in training and validation cohort to demonstrate the early-cancer discriminability of the diagnostic score model.

Table 1. The basic information of series in the study

Series accession numbers	Platform used	No. of normal samples	No. of tumorous samples	AJCC stage	Gender	Mean age, [min, max]	Region	Survival Outcome
GSE75037	Illumina Human WG-6 v3.0 expression beadchip	83	83	I:50; II:20; III:11; IV:2	Female:118 Male:48	68,[39,90]	USA	NA
GSE63459	Illumina Human Ref-8 v3.0 expression beadchip	32	33	I:28; II:5	Female:34 Male:31	66,[47,88]	USA	NA
GSE29013	Affymetrix Human Genome U133 Plus 2.0 Array	0	30	I:16; II:6; III:8	Female:10 Male:20	64,[32,76]	USA	OS
GSE30219	Affymetrix Human Genome U133 Plus 2.0 Array	0	85	NA	Female:19 Male:66	61,[44,84]	France	OS
GSE37745	Affymetrix Human Genome U133 Plus 2.0 Array	0	106	I:70; II:19; III:13; IV:4	Female:60 Male:46	63,[40,83]	Sweden	OS
GSE50081	Affymetrix Human Genome U133 Plus 2.0 Array	0	129	I:93; II:36	Female:62 Male:67	69,[40,86]	Canada	OS
TCGA-LUAD	Illumina RNAseq	59	526	I:172 II:273 III:45 IV:18 NA:18	Female:274 Male:237 NA:15	65,[33,88]	NA	OS
TCGA-LUSC	Illumina RNAseq	49	501	I:114 II:295	Female:131 Male:370	67,[39,90]	NA	OS

Series accession numbers	Platform used	No. of normal samples	No. of tumorous samples	AJCC stage	Gender	Mean age, [min, max]	Region	Survival Outcome
				III:71 IV:21				
GTEX	NA	288	0	NA	NA	NA	NA	NA

Construction of the clusters stratified by m⁶A RNA methylation genes and correlation analysis

We used "ConsensusClusterPlus" to cluster tumor samples into two to five groups [51]. Then Kaplan-Meier analysis and correlation analysis were applied to explain the correlation between clusters.

Prognostic evaluation of 13 m⁶A RNA methylation genes in lung adenocarcinoma

To assess the prognostic evaluation of each m⁶A RNA methylation gene, univariate Cox regression analysis was performed. Besides, lasso regression model was also made with 13 m6A RNA methylation genes to optimize the prognostic meaning.

$$\text{Risk score} = \sum_{i=1}^{n} \text{Coefi} \ast \text{Xi}$$

(Coefi is the coefficient of each selected gene, Xi is the expression value.)

From this risk score model, each patient could get a risk score. We defined patient into high risk group with the risk score≥ median value of all patients, and patient into low risk group with the risk score< median value. Then Kaplan-Meier analysis was performed to display the prognostic performance of Risk score model in both training and validation cohorts. And correlation analysis was applied to explain the correlation between subgroups stratified by the risk score model and clinicopathological characters, including smoking history, TNM stage, metastasis, lymph node stage, gender, diagnosed age, and status in both training and validation cohorts.

Univariate and multivariate Cox regression analyses of clinicopathological characters and risk score

Univariate and multivariate Cox regression analyses of risk score and clinicopathological characters, including smoking history, TNM stage, metastasis, lymph node stage, gender, diagnosed age, and tumor stage, were performed to identify the prognostic performance of these characters. "Rms" package of R was used to build a nomogram with 1-year, 2-year, and 3-year overall survival (OS) as endpoints.

Statistical analysis

All figures and data were accomplished by R (Version 3.5.3) and Graphad Prism 5.0 (La Jolla, CA).

To estimate the diagnostic role of diagnostic model, operating characteristic (ROC) curves were plotted. One-way ANOVA or t-test was conducted to compare different clinicopathological characters in subclusters or risk groups. Kaplan-Meier plots, displayed as hazard ratio (HR) with 95% confidence intervals (CI), were performed to compare the OS of patients of different subclusters and risk groups. All results with P<0.05 were considered significant.

Results

Differential expression of m⁶A RNA methylation genes in lung adenocarcinoma

Comparing 13 m⁶A RNA methylation genes between 347 normal samples and 526 tumorous samples of TCGA and GTEx, we found that 12 of these genes had statistic difference. Among these genes, KIAA1429, HNRNPC, YTHDC2, METTL3, WTAP, YTHDC1, and FTO were down expressed in lung adenocarcinoma, RBM15, ZC3H13, YTHDF1, YTHDF2, and ALKBH5 were up expressed (Figure 1A-B). These 13 m⁶A RNA methylation genes were mainly associated with RNA modification, dosage compensation by inactivation of X chromosome, and RNA destabilization (Figure 1C). In spearman correlation analysis, YTHDC1, YTHDC2, FTO, WTAP, HNRNPC, and METTL3 were positively related with each other. Besides, ZC3H13, ALKBH5, YTHDF1, and YTHDF2 were also positively related with each other in lung adenocarcinoma. Furthermore, YTHDF1, and YTHDF2 were negatively related with FTO, WTAP, HNRNPC, and METTL3 (Figure 1D).

Construction of diagnostic model

The ROC curves of 13 m⁶A RNA methylation genes were shown in Figure S1A. To improve the diagnostic accuracy, 13 genes were served as candidate genes into lasso regression analysis. Figure 2A-B showed the process of model building in the training cohort. Diagnostic score = 0.033*KIAA1429+ 0.116*HNRNPC+0.115*RBM15-0.067*METTL3-0.048* ZC3H13-0.221*WTAP+0.213*YTHDF1-0.132*YTHDC 1-0.135*FTO+0.078*YTHDF2+0.014*ALKBH5. Each people could get a diagnostic score according to the diagnostic score model. The diagnostic scores of the tumor group were significantly higher than the normal group (Figure 2C). Area under the curve (AUC) was 0.996, P<0.0001, indicating high sensitivity and specificity (Figure 2D). The diagnostic score

model was further validated in two independent cohorts-GSE75037 and GSE63459. Figure 2E and Figure S1B showed that the variation trends of HNRNPC, RBM15, YTHDF1, YTHDC1, FTO, and YTHDF2 of GSE75037 were accord with these in the training cohort. Meanwhile, people of tumor group also had extraordinarily higher diagnostic scores than these of normal group (Figure 2F). The AUC of diagnostic score model was 0.971, which was higher than single gene (Figure 2G and Figure S1C). For GSE63459, the trends of RBM15, WTAP, YTHDF1, FTO, and YTHDF2 were consistent with these in the training cohort (Figure 2H and Figure S1D). Likewise, the diagnostic score model had its superiority in diagnosing lung adenocarcinoma (Figure 2I-J and Figure S1E). Besides, diagnostic score model had discrimination of each stage of lung adenocarcinoma, even the very early stage-stage IA in the training and validation cohorts (Figure 2K-L).
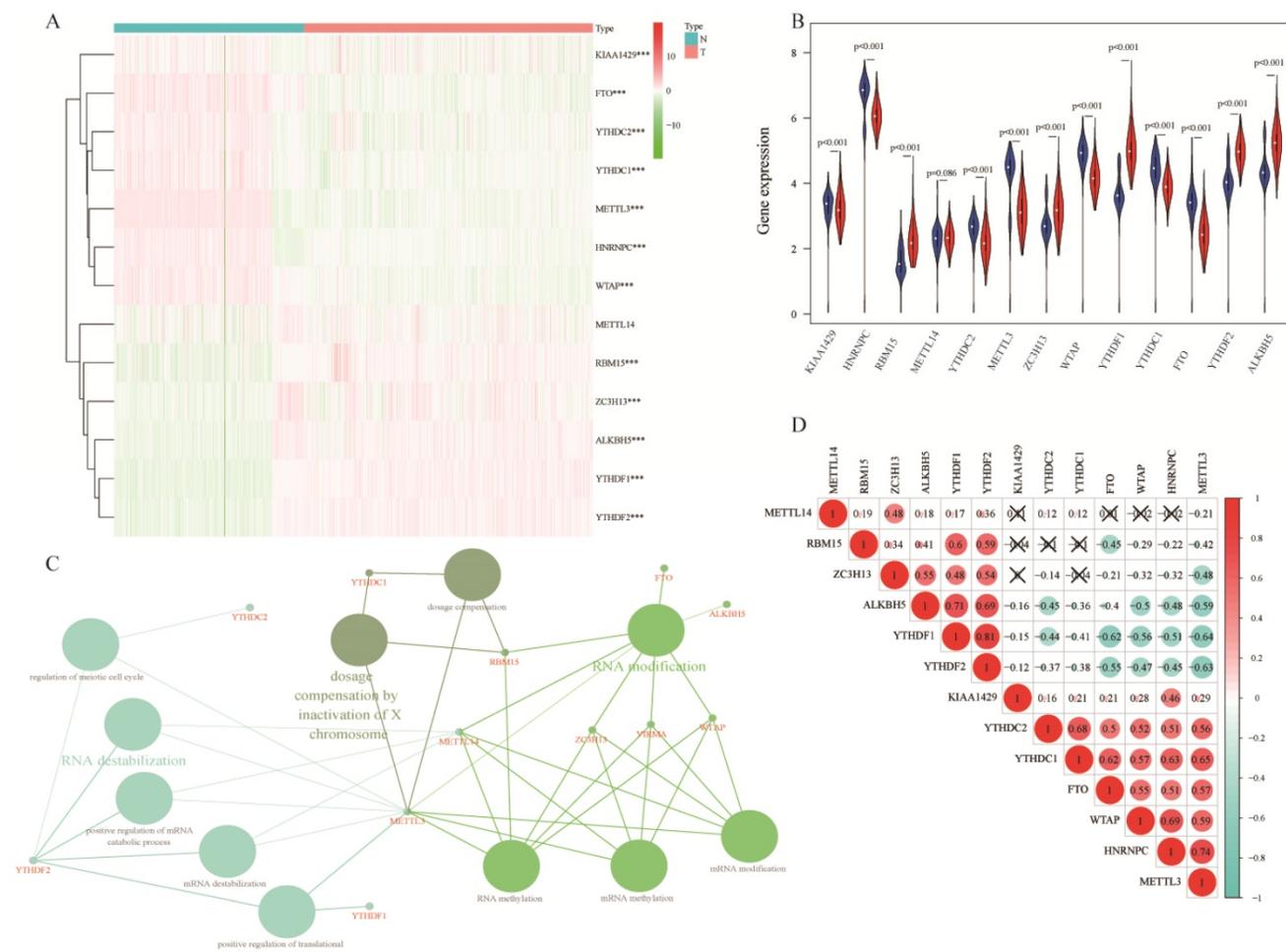
Correlation of m⁶A RNA methylation genes and clinicopathological characters

In both training and validation cohorts, YTHDC2 was associated with tumor stage (P<0.01, Figure 3A-C). Besides, HNRNPC was up expressed in progressed tumor group of training and merged validation cohorts (P<0.05, Figure 3D-F). For mutation, there was no correlation between m⁶A RNA methylation genes and KRAS or EGFR mutation (Figure S2). For TP53 mutation, WTAP, RBM15, KIAA1429, YTHDF1, and YTHDF2 were all up expressed in mutation group (Figure 3G-H).

Construction of risk score model

Before we build the risk score model, we clustered the patients into two to five subclusters (K=2-5) stratified by 13 m⁶A RNA methylation genes. Regretfully, subclusters stratified by m⁶A RNA methylation genes were not related with OS (Figure S3A-H). Due to subclusters that didn't have a good prognostic performance, we assessed the prognostic performance by a univariate Cox regression analysis in the training cohort. In these genes, HNRNPC was hazard of OS (HR: 1.8, 95%CI (1.22-2.656), P=0.003) (Figure 4A).
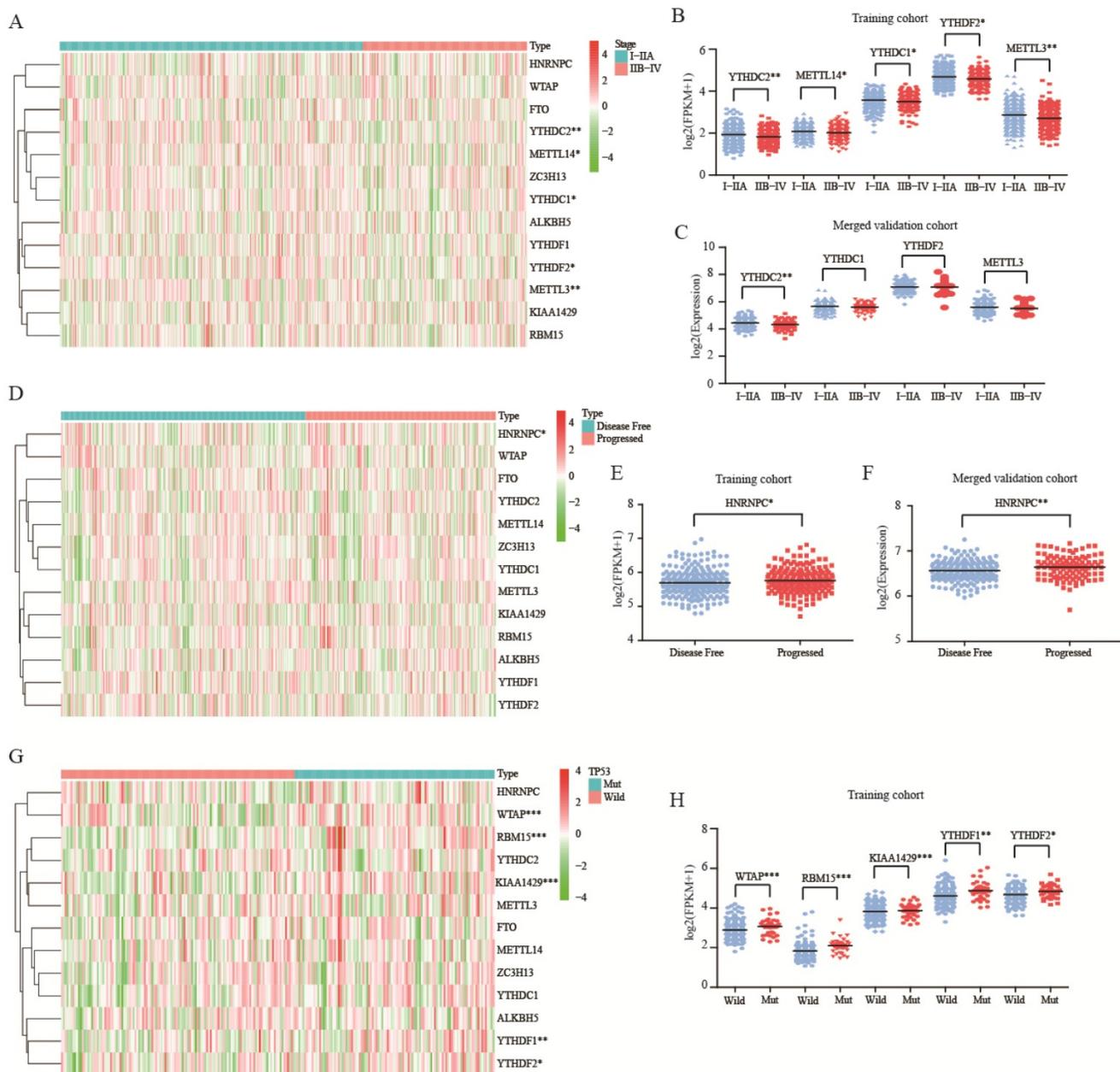

Figure 1. Expression of m⁶A RNA methylation genes in lung adenocarcinoma. (A-B) The differential expression of 13 m⁶A RNA methylation genes between 347 normal samples and 526 tumorous samples in lung adenocarcinoma. **(C)** Visualization of relation between m⁶A RNA methylation genes and related pathways. **(D)** Correlation analysis of 13 m⁶A RNA methylation genes. ***, P<0.001.
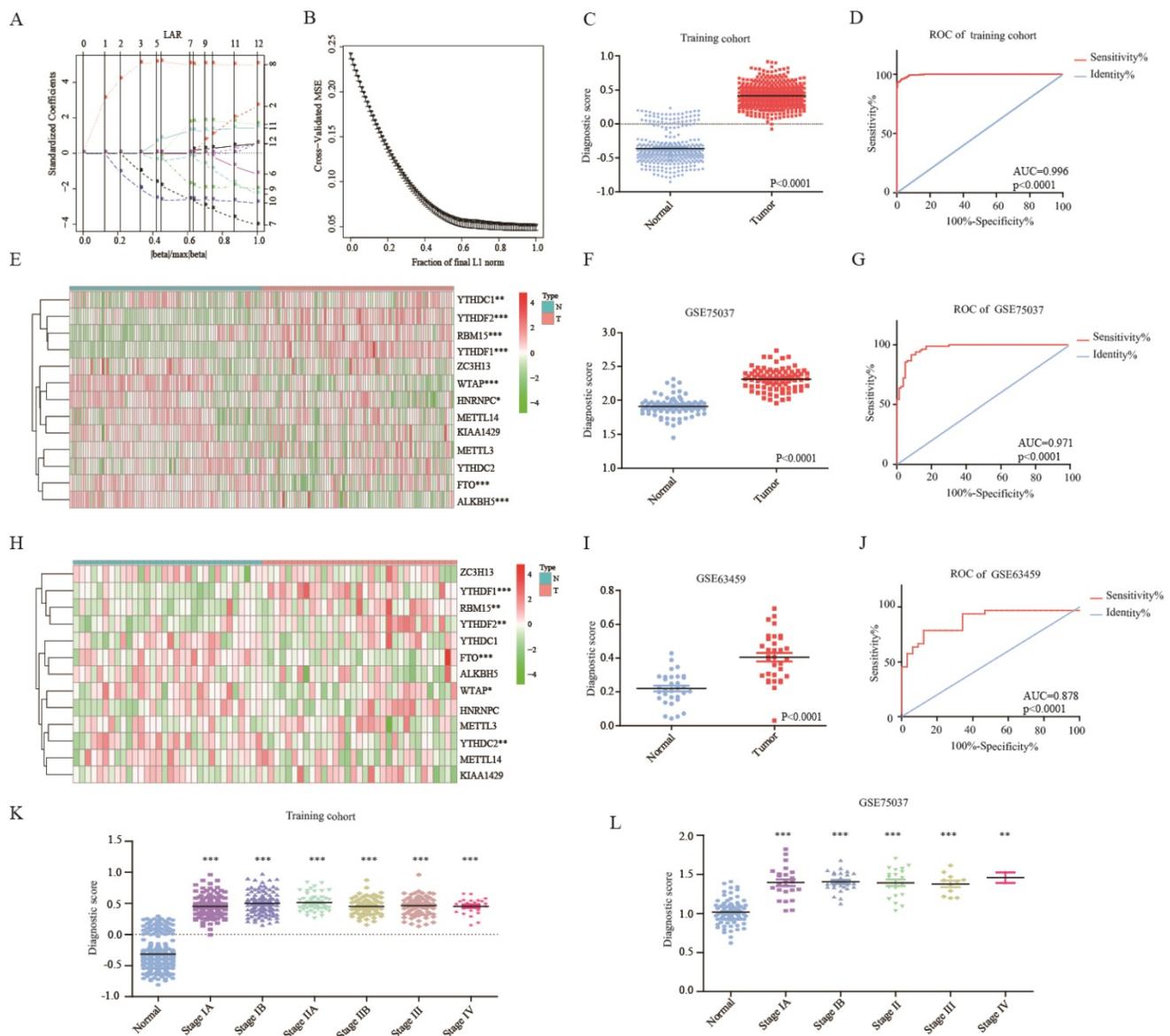
Figure 2. Diagnostic role of m⁶A RNA methylation genes related model. (**A-B**) The solution paths of lasso regression model and the relationship between cross-validated mean square error (CV MSE) and model size. (**C**) The difference in the distribution of diagnostic score of training cohort. (**D**) The ROC curve of diagnostic score of training cohort. (**E**) The expression of 13 m⁶A RNA methylation genes in lung adenocarcinoma of validation cohort-GSE75037. (**F-G**) The distribution and ROC curve of diagnostic score of validation cohort-GSE75037. (**H**) The expression of 13 m⁶A RNA methylation genes in lung adenocarcinoma of validation cohort-GSE63459. (**I-J**) The expression of 13 m⁶A RNA methylation genes in lung adenocarcinoma of validation cohort-GSE63459. (**K**) The difference in the distribution of diagnostic score of tumor stage of training cohort. (**L**) The difference in the distribution of diagnostic score of tumor stage of validation cohort-GSE75037. *, P<0.05; **, P<0.01; ***, P<0.001.

To better predict survival of lung adenocarcinoma with m⁶A RNA methylation genes, we conducted lasso regression analysis (Figure 4B-C). Then a ten-gene risk score model was built. Risk score=0.169*ALKBH5-0.159*FTO+0.581*HNRNPC-0.348*YTHDF2-0.265*YTHDF1-0.123*YTHDC2+0.434*RBM15+0.143*KIAA1429-0.200*WTAP-0.310*METTL3 (Figure 4D). We defined patient into high risk group with the risk score ≥ median value of all patients, and patient into low risk group with the risk score < median value. The training cohort (n=500) and merged validation cohort (n=350) were divided into high- or low-risk group respectively. Figure 5A-B showed that risk score model had its superior performance in predicting outcome of lung adenocarcinoma in both training (HR: 1.828, 95%CI (1.347, 2.481)) and validation cohort (HR: 1.812, 95%CI (1.075, 2.563)). Figure 5C showed the expression of 10 m⁶A RNA methylation genes involved in risk score model. There existed correlation between risk group and TNM stage (P<0.01), lymph node stage (P<0.05), gender (P<0.05), status (P<0.001). The distributions of risk score, smoking history, tumor stage, metastasis, lymph node stage, tumor stage, gender, diagnostic age, and status of training cohort were shown in Figure 5D-L. And the correlations between risk score and gender, tumor stage, status were verified in merged validation cohort (Figure 5M-P).

Figure 3. Role of m⁶A RNA methylation genes in lung adenocarcinoma with different clinicopathological characters. (A-B) The expression of m⁶A RNA methylation genes with different tumor stage of training cohort. (C) The expression of m⁶A RNA methylation genes with different tumor stage of merged validation cohort. (D-E) The expression of m⁶A RNA methylation genes with different tumor status of training cohort. (F) The expression of m⁶A RNA methylation genes with different tumor status of merged validation cohort. (G-H).The expression of m⁶A RNA methylation genes with TP53 mutation of training cohort.*, P<0.05; **, P<0.01; ***, P<0.001.

Univariate and multivariate Cox regression analyses of clinicopathological characters and risk score

To investigate whether risk score was an independent risk factor, we implemented univariate and multivariate Cox regression analyses of relevant clinicopathological characters and risk score. In the univariate Cox regression analysis, lymph node stage, TNM stage, tumor stage, and risk score were all associated with bad OS (Figure 6A). While in the multivariate Cox regression analysis, only the risk score was related with bad OS (HR: 2.181, 95%CI (1.594-2.984), P<0.001) (Figure 6B), indicating risk score was an independent risk factor of lung adenocarcinoma. A nomogram was built to facilitate clinicians to predict outcome. Diagnostic age, gender, TNM stage, and risk score were given points according to their impacts to the outcome (Figure 6C). By summing up all the points, we could get a total point for each patient. Then we could be aware of 1-year, 2-year, and 3-year survival of patients respectively.
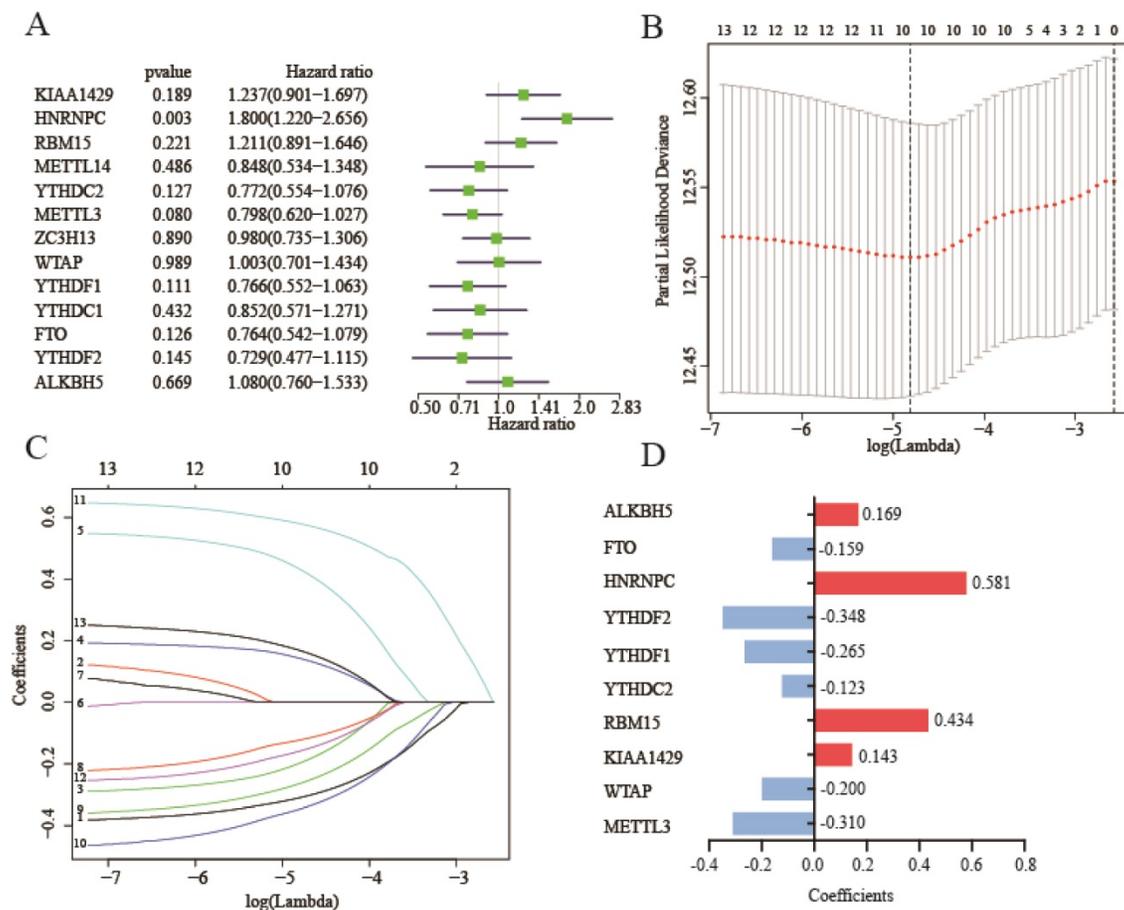
Figure 4. Construction of risk score model with m6A RNA methylation genes. (A) Univariate Cox regression analysis of 13 m6A RNA methylation genes. **(B-C)**The process of building lasso model with size and coefficients by multivariate Cox regression. **(D)** The coefficients of 10 m6A RNA methylation genes involve in lasso risk model. Hazard ratio (95% confidence intervals).

Role of the diagnostic score model and prognostic risk model in squamous cell lung carcinoma

We investigated the expression of 13 m6A RNA methylation genes in squamous cell lung carcinoma. From Figure S4A-B, the variation trends of 12 m6A RNA methylation genes were in line with these in lung adenocarcinoma. Besides, the result of correlation analysis of squamous cell lung carcinoma was roughly consistent with that in lung adenocarcinoma (Figure S4C). The diagnostic score could diagnose squamous cell lung carcinoma as well (AUC=0.992, P<0.0001) (Figure S4D). However, the risk score that enabled clinicians to predict outcome of lung adenocarcinoma couldn't be applied to squamous cell lung carcinoma (Figure S4E).

Discussion

Lung cancer is the most common malignant tumor with high morbidity and mortality worldwide, of which NSCLC accounts for 85-90%, including LUSC, LUAD and large cell lung cancer (LCLC) [52]. The underlying cause of the low lung cancer survival rate is its difficulty in early diagnosis. More than 70% patients are diagnosed at an advanced and usually fatal stage, lacking the effective therapeutic measures [53]. Thus, it becomes great significance to understand the development mechanism for its well diagnosis and treatment. Among these subtypes, LUAD is responsible for approximately 40% of all NSCLC cases [33]. The early diagnosis and treatment can greatly improve the survival rate of patients with LUAD. Recently, emerging evidence has indicated that the development of lung cancer is both affected by genetic variation and epigenetic variation [54]. Epigenetic regulates gene expression from multiple levels, including DNA methylation, RNA regulation, histone modification, and chromosome remodeling [55]. However, up to now, in lung cancers, only FTO and METTL3 have been reported as potential targets for its diagnosis and treatment [31,32]. In this study, except FTO and METTL3, we also found that KIAA1429, HNRNPC, YTHDC2, WTAP, YTHDC1, RBM15, ZC3H13, YTHDF1, YTHDF2, and ALKBH5 had differential expression in LUAD.

In exploring the clinical value of these differential expression regulators, it was found that

the joint detection of these regulators was more significant for the diagnosis of LUAD in comparison to the detection of mostly single regulator. Such combination provided higher predictive accuracy of 99.6% (AUC) for LUAD, while the most single regulator was less than 90%. The diagnostic score model we constructed involved 11genes. From the results of Figure S1, in the training and validation cohorts, YTHDF1 had a good diagnostic performance. So far, many studies in progress with regard

toYTHDF1 in colorectal carcinoma and hepatocellular carcinoma, which was usually regarded as an independent prognostic factor [56-58]. However, there has been no related reports on LUAD, and no has been conducted as a diagnostic factor. Though the diagnostic score model had a unique superiority in the diagnosis of LUAD, we could use YTHDF1 alone to diagnose disease in consideration of operability and simplification.
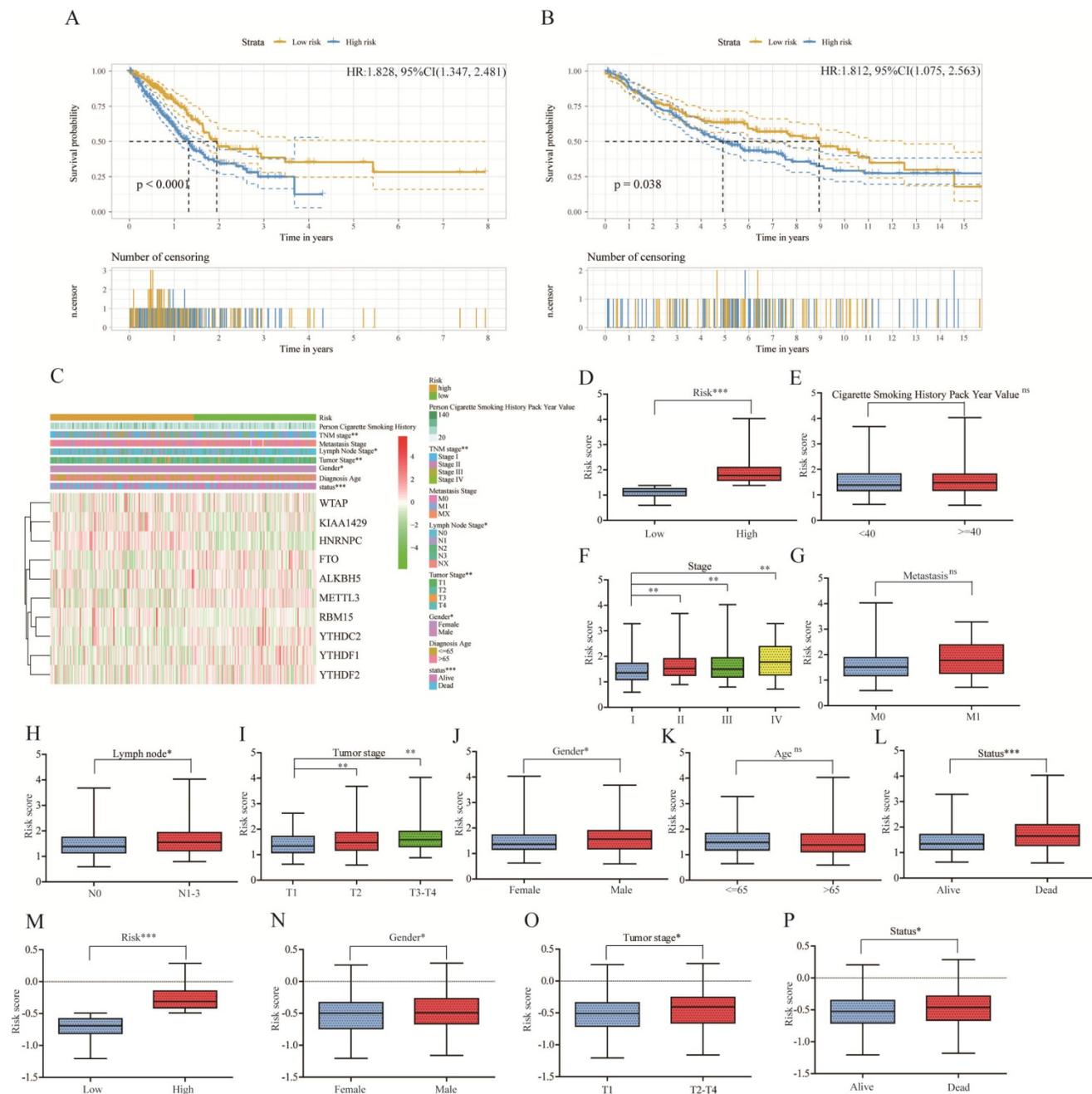
Figure 5. Relationship between clinicopathological characters and risk score model in lung adenocarcinoma of training cohort and merged validation cohort. (**A**) Kaplan-Meier plot of high risk and low risk subgroup in training cohort. (**B**) Kaplan-Meier plot of high risk and low risk subgroup in merged validation cohort. (**C**).The heatmap of 10 m⁶A RNA methylation genes involve in lasso risk model and relationship between clinicopathological characters and risk subgroup. (**D-L**) The distribution of risk score in different risk (D), smoking history (E), stage (F), metastasis (G), lymph node (H), tumor stage (I), gender (J), age (K), status (L) subgroup of training cohort. (**M-P**) The validation of distribution of risk score in different risk (M), gender (N), tumor stage (O), status (P) subgroup. *, P<0.05; **, P<0.01; ***, P<0.001, ns, no sense; HR, hazard ratio; 95%CI, 95% confidence intervals.
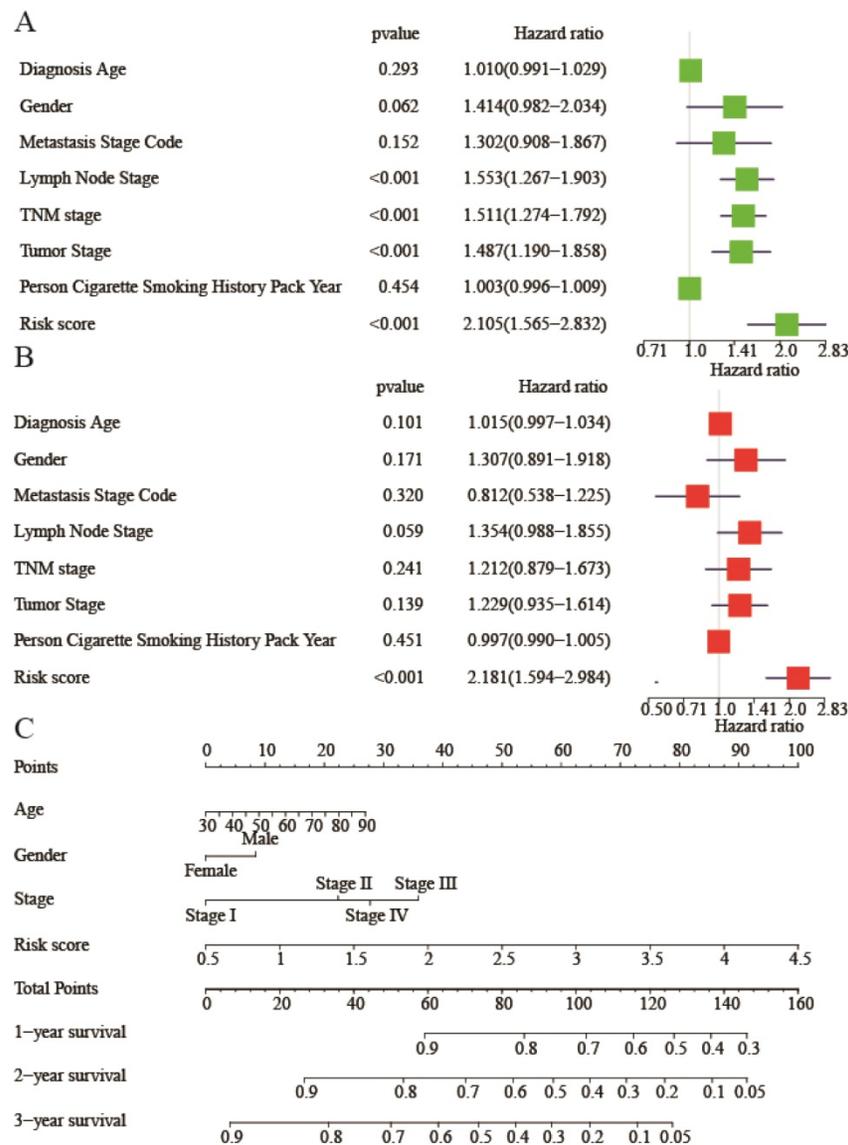
Figure 6. Univariate and multivariate Cox regression analysis of clinicopathological characters and risk score. (**A**) Univariate Cox regression analysis for evaluating the prognostic role of clinicopathological characters and risk score in lung adenocarcinoma. (**B**) Multivariate Cox regression analysis for evaluating the prognostic role of clinicopathological characters and risk score in lung adenocarcinoma. (**C**) The nomogram plot to predict 1-year, 2-year, 3-year overall survival. Summing up each point of clinicopathological characters and risk score to predict overall survival. Hazard ratio (95% confidence intervals).

In the absence of diagnostic characteristics, YTHDC2 was then found to be associated with tumor stage. Accurate pathological staging is the basis for setting on ideal treatment. Increasing researchers are also devoted in the research of molecular staging using gene or protein expression. Previously, in colon cancer, it has been found a significantly positive correlation between YTHDC2 expression and the tumor stage [59]. However, there is still lack of information for LUAD. The finding thus will provide new insight into our understanding of the function of YTHDC2 in LUAD. HNRNPC, except as a joint regulator in the diagnostic score model, it even exhibited differential expression in progressed tumor group and disease free group. Many studies have suggested that several types of solid tumor cells

including breast cancer, gastric cancer, and esophageal squamous cell carcinoma acquired generation and development characteristics through HNRNPC disorder [60-62]. In lung epithelial cells, the stability of urokinase receptor (uPAR) mRNA was regulated by HNRNPC [63]. Increased uPAR expression as well as stabilization of uPAR mRNA would contribute to the pathogenesis of lung inflammation and neoplasia [63]. In this study, we preliminary confirmed that HNRNPC played a critical role in LUAD progression, which was expected to provide a potential therapeutic target.

The consensus clustering stratified by 13 m^{6}A RNA methylation genes was not associated with the most clinicopathologic characters and survival. The inconsistent aberrated trends of these m^{6}A RNA

methylation genes might account for the frustrated utilization of the stratification. Then, we performed univariate Cox regression analysis of each selected m6A RNA methylation genes. We found that HNRNPC also was hazard of OS (HR: 1.8, 95%CI (1.22-2.656), P=0.003). As we have demonstrated, for breast cancer cells, the repression of HNRNPC could inhibit cell proliferation and tumor growth [60]. In gastric cancer, HNRNPC has been identified as a prognostic and therapeutic marker [61]. However, to the best of our knowledge, there has been rarely reported that HNRNPC was involved in the prognosis of LUAD.

As reported, a number of monogenic prognostic markers that were found significantly associated with prognosis, such as p53, HER2, and ERBB3 [64]. However, due to the heterogeneity of NSCLC, the genes involved might vary greatly among those different individuals [64]. Thus, it greatly prompted the researchers to construct gene expression profiles composed of multiple prognostic genes for using in the patients risk stratification. For instance, Kratz et al. ever have constructed a prognostic model using 14 genes which divided the patients with NSCLC into low-risk, intermediate-risk, and high-risk groups [65]. The results showed that the 5-year survival rate among three groups was significantly different, and the model exhibited satisfactory predictability in prognosis than traditional clinical pathology staging in multiple validation cohorts. Similarly, in this study, A ten-gene risk score model (Risk score=0.169*ALKBH5-0.159*FTO+0.581*HNRNPC-0.348*YTHDF2-0.265*YTHDF1-0.123*YTHDC2+0.434*RBM15+0.143*KIAA1429-0.200*WTAP-0.310*METTL3)w as built. It stratified the OS of patients with LUAD into high and low risk categories, which exhibited entirely differences in survival. Following the discovery of these prognostic genes, the risk group distinguished through risk score was found existed correlation with TNM stage, lymph node stage, gender, status, and tumor stage in both training and validation cohort. Besides, these clinicopathological features were associated with bad OS, and risk score was an independent risk factor of LUAD. At present, TNM stage is commonly performed to assess the prognosis of patients. Due to the limited risk factors included in the TNM staging system, it is not possible to conduct a precise prediction for NSCLC. Validating and combining more clinical pathology and molecular biology risk factors to establish an effective prognosis model becomes an important direction for future research in the development of multidisciplinary comprehensive treatment. In this study, based on characteristics of diagnostic age, gender, TNM stage, and risk score, a nomogram was built to facilitate

clinicians to predict outcome. To some extent, it was expected to address the issue of prognostic heterogeneity caused by single factor or insufficient risk factors analysis.

In spite of these indelible contributions, as the study went on, it was found that the above regulators exhibited no significant differential expression in KRAS and EGFR mutant LUAD. In TP53 mutant LUAD, the regulators including WTAP, RBM15, KIAA1429, YTHDF1, and YTHDF2 all were highly expressed. Thus, we inferred that those regulators were likely only effective in some types of LUAD, while the specific reasons required following elucidation. Besides, it was note-worthy that YTHDF2 was merely reported to be involved with TP53 in peripheral T-cell lymphoma and gastric cancer yet [66,67]. YTHDF1 was functionally interactional with TP53 in gastric cancer [67]. To the best of our knowledge, it was not clear yet that TP53 mutant impacted the expression of WTAP, RBM15, and KIAA1429, in which more evidences were required to elucidate their mechanistic correlation.

In addition to LUAD, our work also indicated that m6A shared connections with LUSC, and the joint detection also provided higher predictive accuracy of 99.2% (AUC). Unfortunately, as we observed, these m6A RNA methylation regulators exhibited no prognostic value in LUSC. LUAD is more likely to occur in women and non-smokers [68]. LUSC is more common in older men and is closely related to smoking [69]. It is likely that different pathological features determine the different prognostic effects exhibited by the same regulators, which is worth investigating in future studies.

In this study, due to the limited sample size of normal samples, we normalized the lung tissues of GTEx and increased the credibility and accuracy of differential expression analysis. The robust analyses of correlation were performed among m6A RNA methylation regulators and diagnosis, pathological features or prognosis. The sample number of the study approximately reached 1900, which we thought to be enough to offset some confounding factors.

Conclusion

In conclusion, our work systemically elucidated m6A's progressive role in LUAD, and provided insights into its diagnostic and prognostic function. However, m6A's functional details and its relationship with specific gene mutations and other types of NSCLC in controlling tumor progression merit further investigation.

Abbreviations

m6A: N6-methyladenosine; LUAD: lung

adenocarcinoma; ROC: receiver operating characteristic; AUC: area under the curve; lasso: least absolute shrinkage and selection operator; m^{1}A: N1-methyladenosine; lncRNAs: long-chain non-coding RNAs; tRNA: transfer RNA; rRNA: ribosome RNA; CSCC: cervical squamous cell carcinoma; MZF1: myeloid zinc finger 1; LUSC: lung squamous cell carcinoma; NSCLC: non-small cell lung cancer; GTEx: Genotype-Tissue Expression; OS: overall survival; HR: hazard ratio; CI: confidence intervals; LCLC: large cell lung cancer; uPAR: urokinase receptor.

Supplementary Material

Acknowledgements

Funding

Competing Interests

The authors have declared that no competing interest exists.

References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018; 68: 394-424.
2. Verma M. The role of epigenomics in the study of cancer biomarkers and in the development of diagnostic tools. Adv Exp Med Biol. 2015; 867: 59-80.
3. Wang KC, Chang HY. Epigenomics: technologies and applications. Circ Res. 2018; 122: 1191-1199.
4. Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012; 485: 201-206.
5. Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature. 2016; 530: 441-446.
6. Spenkuch F, Motorin Y, Helm M. Pseudouridine: still mysterious, but never a fake (uridine) !. RNA Biol. 2014; 11: 1540-1554.
7. Boccaletto P, Machnicka MA, Purta E, et al. MODOMICS: a database of RNA modification pathways. 2017 update. Nucleic Acids Res. 2018; 46: D303-D307.
8. Dubin DT, Taylor RH. The methylation state of poly A-containing messenger RNA from cultured hamster cells. Nucleic Acids Res. 1975; 2: 1653-1668.
9. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. Proc Natl Acad Sci U S A. 1974; 71: 3971-3975.
10. Wei CM, Gershowitz A, Moss B. Methylated nucleotides block 5'terminus of HeLa cell messenger RNA. Cell. 1975; 4: 379-386.
11. Visvanathan A, Somasundaram K. MRNA traffic control reviewed: N6-methyladenosine (m6 A) takes the driver's seat. Bioessays. 2018; 40.
12. Meyer KD, Jaffrey SR. Rethinking m6A Readers, Writers, and Erasers. Annu Rev Cell Dev Biol. 2017; 33: 319-342.
13. Bokar JA, Shambaugh ME, Polayes D, et al. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. RNA. 1997; 3: 1233-1247.
14. Wang Y, Li Y, Toth JI, et al. N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat Cell Biol. 2014; 16: 191-198.
15. Ping XL, Sun BF, Wang L, et al. Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. 2014; 24: 177-189.
16. Horiuchi K, Kawamura T, Iwanari H, et al. Identification of Wilms' tumor 1-associating protein complex and its role in alternative splicing and the cell cycle. J Biol Chem. 2013; 288: 292-302.
17. Patil DP, Chen CK, Pickering BF, et al. M6A RNA methylation promotes XIST-mediated transcriptional repression. Nature. 2016; 537: 369-373.
18. Schwartz S, Mumbach MR, Jovanovic M, et al. Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. Cell Rep. 2014; 8: 284-296.
19. Wen J, Lv R, Ma H, et al. Zc3h13 regulates nuclear RNA m6A methylation and mouse embryonic stem cell self-renewal. Mol Cell. 2018; 69: 1028-1038.
20. Wang X, Lu Z, Gomez A, et al. N6-methyladenosine-dependent regulation of messenger RNA stability. Nature. 2014; 505: 117-120.
21. Huang Y, Yan J, Li Q, et al. Meclofenamic acid selectively inhibits FTO demethylation of m6A over ALKBH5. Nucleic Acids Res. 2015; 43: 373-384.
22. Meyer KD, Saletore Y, Zumbo P, et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell. 2012; 149: 1635-1646.
23. Ben-Haim MS, Moshitch-Moshkovitz S, Rechavi G. FTO: linking m6A demethylation to adipogenesis. Cell Res. 2015; 25: 3-4.
24. Shen F, Huang W, Huang JT, et al. Decreased N(6)-methyladenosine in peripheral blood RNA from diabetic patients is associated with FTO expression rather than ALKBH5. J Clin Endocrinol Metab. 2015; 100: E148-154.
25. Angelova MT, Dimitrova DG, Dinges N, et al. The emerging field of epitranscriptomics in neurodevelopmental and neuronal disorders. Front Bioeng Biotechnol. 2018; 6:46.
26. Zheng G, Dahl JA, Niu Y, et al. ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. Mol Cell. 2013; 49: 18-29.
27. Li HB, Tong J, Zhu S, et al. m(6)A mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. Nature. 2017; 548: 338-342.
28. Zheng Q, Hou J, Zhou Y, et al. The RNA helicase DDX46 inhibits innate immunity by entrapping m(6)A-demethylated antiviral transcripts in the nucleus. Nat Immunol. 2017; 18: 1094-1103.
29. Zhou S, Bai ZL, Xia D, et al. FTO regulates the chemo-radiotherapy resistance of cervical squamous cell carcinoma (CSCC) by targeting β-catenin through mRNA demethylation. Mol Carcinog. 2018; 57(5): 590-597.
30. Nishizawa Y, Konno M, Asai A, et al. Oncogene c-Myc promotes epitranscriptome m6A reader YTHDF1 expression incolorectal cancer. Oncotarget. 2018; 9(7): 7476-7486.
31. Liu J, Ren D, Du Z, Wang H, Zhang H, Jin Y. m6A demethylase FTO facilitates tumor progression in lung squamous cell carcinoma by regulating MZF1 expression. Biochem Biophys Res Commun. 2018; 502: 456-464.
32. Du M, Zhang Y, Mao Y, et al. MiR-33a suppresses proliferation of NSCLC cells via targeting METTL3 mRNA. Biochem Biophys Res Commun. 2017; 482: 582-589.
33. Li TT, Gao X, Gao L, et al. Role of upregulated miR-136-5p in lung adenocarcinoma: A study of 1242 samples utilizing bioinformatics analysis. Pathol Res Pract. 2018; 214: 750-766.
34. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013; 45: 580-5.
35. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014; 511: 543-50.
36. Goldman M, Craft B, Hastie M, et al. The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. doi: https://doi.org/10.1101/326470.
37. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012; 2: 401-404
38. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013; 6: pl1.
39. Girard L, Rodriguez-Canales J, Behrens C, et al. An Expression Signature as an Aid to the Histologic Classification of Non-Small Cell Lung Cancer. Clin Cancer Res. 2016; 22: 4880-4889.
40. Robles AI, Arai E, Mathé EA, et al. An Integrated Prognostic Classifier for Stage I Lung Adenocarcinoma Based on mRNA, microRNA, and DNA Methylation Biomarkers. J Thorac Oncol. 2015; 10: 1037-48.
41. Xie Y, Xiao G, Coombes KR, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. Clin Cancer Res. 2011; 17: 5705-14.
42. Rousseaux S, Debernardi A, Jacquiau B, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. Sci Transl Med. 2013; 5: 186ra66.
43. Jabs V, Edlund K, König H, et al. Integrative analysis of genome-wide gene copy number changes and gene expression in non-small cell lung cancer. PLoS One. 2017; 12: e0187246.
44. Der SD, Sykes J, Pintilie M, et al. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. J Thorac Oncol. 2014; 9: 59-64.
45. Chai RC, Wu F, Wang QX, et al. m6A RNA methylation regulators contribute to malignant progression and have clinical prognostic impact in gliomas. Aging NY. 2019; 11: 1204-1225.
46. Zhou J, Wang J, Hong B, et al. Gene signatures and prognostic values of m6A regulators in clear cell renal cell carcinoma - a retrospective study using TCGA database. Aging (Albany NY). 2019; 11: 1633-1647.

47. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009; 25: 1091-1093.
48. Bindea G, Galon J, Mlecnik B. Clue Pedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bioinformatics. 2013; 29: 661-3.
49. Zhao Y, Simon R. Development and validation of predictive indices for a continuous outcome using gene expression profiles. Cancer Inform. 2010; 9: 105-14.
50. Xiao W, Wu Y, Zhou H. ConvexLAR: An Extension of Least Angle Regression. J Comput Graph Stat. 2015; 24: 603-626.
51. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26: 1572-1573.
52. Wakelee H, Kelly K, Edelman M J. 50 Years of progress in the systemic therapy of non-small cell lung cancer. Am Soc Clin Oncol Educ Book. 2014; 177-189.
53. Ansari J, Shackelford RE, El-Osta H. Epigenetics in non-small celllung cancer: from basics to therapeutics. Transl Lung Cancer Res. 2016; 5: 155-171.
54. Mehta A, Dobersch S, Romero-Olmedo AJ, et al. Epigenetics in lung cancer diagnosis and therapy. Cancer Metastasis Rev. 2015; 34: 229-241.
55. Sandoval J, Esteller M. Cancer epigenomics: beyond genomics. Curr Opin Genet Dev. 2012; 22: 50-55.
56. Zhao X, Chen Y, Mao Q, et al. Overexpression of YTHDF1 is associated with poor prognosis in patients with hepatocellular carcinoma. Cancer Biomark. 2018; 21: 859-868.
57. Bai Y, Yang C, Wu R, et al. YTHDF1 regulates tumorigenicity and cancer stem cell-like activity in human colorectal carcinoma. Front Oncol. 2019; 9: 332.
58. Nishizawa Y, Konno M, Asai A, et al. Oncogene c-Myc promotes epitranscriptome m6A reader YTHDF1 expression in colorectal cancer. Oncotarget. 2017; 9: 7476-7486.
59. Tanabe A, Tanikawa K, Tsunetomi M, et al. RNA helicase YTHDC2 promotes cancer metastasis via the enhancement of the efficiency by which HIF-1α mRNA is translated. Cancer Lett. 2016; 376: 34-42.
60. Wu Y, Zhao W, Liu Y, et al. Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response. EMBO J. 2018; 37: e99017.
61. Huang H, Han Y, Zhang C, et al. HNRNPC as a candidate biomarker for chemoresistance in gastric cancer. Tumour Biol. 2016; 37: 3527-3534.
62. Zhang Y, Chen W, Pan T, et al. LBX2-AS1 is activated by ZEB1 and promotes the development of esophageal squamous cell carcinoma by interacting with HNRNPC to enhance the stability of ZEB1 and ZEB2 mRNAs. Biochem Biophys Res Commun. 2019; 511: 566-572.
63. Shetty S. Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells. Mol Cell Biochem. 2005; 272: 107-118.
64. Lin J, Beer DG. Molecular predictors of prognosis in lung cancer. Ann Surg Oncol. 2012; 19: 669-676.
65. Kratz JR, He J, Van Den Eeden SK, et al. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. Lancet. 2012; 379: 823-832.
66. Watatani Y, Sato Y, Miyoshi H, et al. Molecular heterogeneity in peripheral T-cell lymphoma, not otherwise specified revealed by comprehensive genetic profiling. Leukemia. 2019. doi: 10.1038/s41375-019-0473-1.
67. Zhang C, Zhang M, Ge S, et al. Reduced m6A modification predicts malignant phenotypes and augmented Wnt/PI3K-Akt signaling in gastric cancer. Cancer Med. 2019. doi: 10.1002/cam4.2360.
68. Zhao J, Li L, Wang Q, et al. CircRNA expression profile in early-stage lung adenocarcinoma patients. Cell Physiol Biochem. 2017; 44: 2138-2146.
69. Yun YD, Back JH, Ghang H, et al. Hazard ratio of smoking on lung cancer in Korea according to histological type and gender. Lung. 2016; 194: 281-289.