Clinical Implication of N6-Methylandenosine-Related lncRNAs in Non-Small Cell Lung Cancer

Wenjing Xiao
wuhan union hospital
Wei Geng
Wuhan Union Hospital
Juanjuan Xu
Wuhan Union Hospital
Qi Huang
Wuhan Union Hospital
Jinshuo Fan
Wuhan Union Hospital
Qi Tan
Wuhan Union Hospital
Zhengrong Yin
Wuhan Union Hospital
Yumei Li
Wuhan Union Hospital
Guanghai Yang
Wuhan Union Hospital
Yang Jin (whuhjy@126.com)
Wuhan Union Hospital https://orcid.org/0000-0003-2409-7073

Research

Keywords: Non-small-cell lung cancer, N6-methyladenosine, long non-coding RNA, prognostic signature, survival, chemotherapeutic sensitivity, radiotherapeutic response.

Posted Date: July 23rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-734615/v1

Abstract

Background: Non-small-cell lung cancer (NSCLC) represents the leading cause of cancer-related deaths worldwide and is highly heterogeneous. The N6-methyladenosine (m^6A) RNA, a vital contributor to the outcomes of cancer, can modify long non-coding RNAs (lncRNAs), thereby influencing the transcript stability, gene expression, and serving a wide array of biological functions. However, the role of m^6A-related lncRNAs in NSCLC remains largely unknown.

Methods: We identified a group of m^6A-related lncRNAs (m^6ARLncRNAs) by using co-expression analysis in 1835 NSCLC patients from The Cancer Genome Atlas (TCGA) (N=1145) and Gene Expression Omnibus (GEO) (N=690) datasets. We employed consensus unsupervised clustering analysis to explore molecular patterns based on the expression of m^6ARLncRNAs. Then we filtered m^6ARLncRNAs by Cox regression and LASSO regression to construct a m^6ARLncRNAs signature (m^6ARLncSig) and further evaluated m^6ARLncSig with external and experimental validation by using qRT-PCR. We analyzed the correlation between m^6ARLncSig scores groups with clinical features, chemotherapeutic sensitivity and radiotherapeutic response. Finally, we established a nomogram for prognosis prediction in patients with LUAD and validated it in the testing set and the entire TCGA dataset.

Results: We constructed a m^6ARLncSig for prognosis prediction of patients consisting of 12 m^6ARLncRNAs. The m^6ARLncSig divided patients into a high-risk group and a low-risk one, which had significantly different OS and could independently predict the prognosis of patients. Meanwhile, we revealed that patients in the high- and low-risk groups differed in tumor-infiltrating immune cells, and chemotherapeutic sensitivity, and biological pathways. Of note, we found that m^6ARLncSig was associated with age, tumor stage, and radiotherapeutic response, indicating they were clinically relevant.

Conclusions: Our study demonstrated that m^6ARLncSig could act as a potential biomarker for evaluating the prognosis and therapeutic efficacy in NSCLC patients.

Research Highlights

1. The m^6ARLncSig divided patients into a high-risk group and a low-risk one, which had significantly different OS and could independently predict the prognosis of patients.

2. m^6ARLncSig could act as a potential biomarker for evaluating the prognosis and therapeutic efficacy in NSCLC patients.

Introduction

Lung cancer, the most common and highly recurrent malignancy, represents the leading cause of cancer-related deaths around the world and non-small cell lung cancer (NSCLC) accounts for 85% of the mortality [1–3]. NSCLC is usually diagnosed at the late stage and the 5-year survival rate stands

somewhere between 10–20%, despite chemotherapeutic and immunotherapeutic interventions [4, 5]. Though early-staged NSCLC can be surgically treated, a great many patients relapse with metastasis in spite of aggressive treatment [6]. Recently, mounting clinical indicators, including age, male gender, smoking status, performance status, underlying diseases, and surgical techniques were believed to be intimately associated with the outcomes of lung cancer [7–9].

The N6-methyladenosine (m^6A) RNA modification, i.e., the methylation at the 6th N atom of adenine, is the most common post-transcriptional modification of mRNAs and is widely seen in eukaryotic species of mammals, plants, and yeast [10, 11]. Moreover, m^6A is also found in ribosomal RNA, transfer RNA, and small nuclear RNA as well as non-coding RNA (ncRNA) [12]. To date, more than 7,600 genes and 300 ncRNA in mammals are found to be modified by m^6A modification [12]. As an invertible and dynamic RNA epigenes, m^6A RNA methylation regulators contain “writers” (methyltransferases), “readers” (signal transducers), and “erasers” (demethylases), which regulate gene expression and are linked to various biological functions, such as RNA splicing, export, stabilization, translation, and the biogenesis of ncRNAs [11, 13]. Emerging evidence has suggested that m^6A is implicated in the development and progression of various malignancies, including liver cancer,[14] glioblastoma [15], breast cancer [16, 17], colorectal cancer [18], and lung cancer [19, 20] by taking part in the regulation of self-renewal of cancer stem cells and tumor immune response [18], promotion of cancer cell proliferation [21], and resistance to radiotherapy or chemotherapy [22]. For instance, METTL3 can increase the miR-143-3p/VASH1 axis and serves as a prognosis factor for *in vivo* progression and overall survival (OS) rate of lung cancer [23]. METTL3 directly promotes YAP translation and increases YAP activity by modulating the MALAT1-miR-1914-3p-YAP axis in NSCLC drug resistance and metastasis [24]. And ALKBH5 inhibits tumor growth and metastasis by suppressing YTHDFs-mediated YAP expression and miR-107/LATS2-mediated YAP activity in NSCLC [25]. Meanwhile, bioinformatic studies have revealed m^6A regulators, as a biomarker for the prognostic evaluation, are closely associated with the development of lung cancer [26, 27].

Long non-coding RNAs (lncRNAs) are transcripts longer than 200 nucleotides that possess no or only limited protein-coding capacity [28]. Recently, mounting studies have demonstrated that lncRNAs are involved in the signaling pathways of cancers and could impact cancer cell activities at both transcriptional and post-transcriptional levels, which makes lncRNAs good biomarkers of cancer [29, 30]. Of note, lncRNAs, which are extensively modified by m^6A, can change the transcript stability and gene expression, resulting in regulatory abnormalities, which, in turn, influence the tumorigenesis and progression of cancer [12]. For instance, METTL3-mediated N6-methyladenosine modification leads to LINC00958 up-regulation by stabilizing its RNA transcript, thereby facilitating HCC lipogenesis and progression of HCC [31]. METTL3, has also been shown to increase the stability of FAM225A, an oncogenic lncRNA, to promote tumorigenesis of nasopharyngeal carcinoma (NPC) [32]. Moreover, post-translational upregulation of m^6A-induced lncRNA RP11 could promote the mRNA degradation of two E3 ligases, Siah1 and Fbxo45, and subsequently prevented the proteasomal degradation of Zeb1 via RP11/hnRNPA2B1/mRNA complex [33]. Up to now, the role of m^6A regulators in the dysregulation of lncRNAs in NSCLC has not been fully elucidated.

In the present study, we examined the prognostic relevance of m^{6}ARLncRNAs in 1835 patients with NSCLC from the Cancer Genome Atlas (TCGA, N = 1145) and Gene Expression Omnibus (GEO, N = 690) datasets. We constructed an m^{6}ARLncRNAs prognostic signature (m^{6}ARLncSig), examined the relationship between m^{6}ARLncSig and clinical features and outcomes of NSCLC, tumor immune, radiotherapeutic response, and chemotherapeutic sensitivity. We also validated m^{6}ARLncSig in three GEO datasets, i.e., GSE37745 (N = 196), and GSE31210 (N = 226), and GSE30219 (N = 268) and in NSCLC samples by experimental validation using qRT-PCR. The study showed that m^{6}ARLncSig could serve as a predictor for the OS and an efficacy measure for individualized treatment for NSCLC.

Materials And Methods

Research roadmap

The entire workflow of this study is illustrated in Fig. 1a. Briefly, the transcriptome expression profiles of lncRNAs and m^{6}A regulators were extracted. Co-expression analyses between lncRNAs and m^{6}A regulators were utilized to identify m^{6}A-related lncRNAs (m^{6}ARLncRNAs). Univariate Cox regression analyses were employed to screen out prognostic m^{6}ARLncRNAs. The molecular clustering patterns were explored based on the expression levels of prognostic m^{6}ARLncRNAs. Then, the patients were divided, at random, into two datasets: a training set and a testing set. We constructed a multivariable Cox regression model following variables selection using LASSO regression, which generated a prognostic m^{6}A-related lncRNAs signature (m^{6}ARLncSig) for predicting the overall survival of NSCLC patients. Subsequently, m^{6}ARLncSig was subjected to an assessment, covering the receiver operating characteristic (ROC) curve, external dataset validation, independent prognostic value analyses, clinical stratification analyses, and immune infiltration level analyses. Finally, a nomogram model was established to improve the practicability of the signature as a prognostic predictor for NSCLC.

Data processing

We downloaded mRNA and lncRNA transcriptional profiles (fragments per kilobase million, FPKM) and the clinical features of 1145 patients from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/). GTF files were downloaded from Ensembl (http://asia.ensembl.org/index.html) for accurate differentiation between mRNAs and lncRNAs. mRNAs and lncRNAs symbols were annotated by using the HUGO Gene Nomenclature Committee (HGNC2) database (https://www.genenames.org/). A total of 23 recognized m^{6}A regulators were obtained from the previous publications, including writers (METTL3, METTL14, METTL16, WTAP, VIRMA, ZC3H13, RBM15, RBM15B), erasers (FTO and ALKBH5), and readers (YTHDC1, YTHDC2, YTHDF1, YTHDF2, YTHDF3, HNRNPC, FMR1, LRPPRC, HNRNPA2B1, IGFBP1, IGFBP2, IGFBP3, and RBMX). In this study, some clinical samples were excluded against the criteria due to the absence of survival data or short survival time (<30 days), to rule out the non-cancer causes of death. Ultimately, 963 patients satisfied the aforementioned inclusion criteria and were included in the study. Then, we merged the expression matrix and the clinical

information matrix into form a new matrix, which included complete survival information, mRNA and lncRNA expression profiles, and other clinical data. In this study, 963 patients were equally randomized into the training dataset for m^{6}ARLncSig model construction and the testing dataset for model validation, respectively, to reach a more compelling and reliable conclusion. The baseline clinical data of the NSCLC patients are listed in Table 1. Three external validation datasets, including GSE37745 (N=196), GSE30219 (N=268), and GSE31210 (N=226), driven largely by large sample size, detailed clinical and survival data, and information resources from the GPL570 Affymetrix, were downloaded from Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) to further confirm our prediction model. The probes of the Affymetrix HG-U133_Plus 2.0 platform were re-annotated to gene symbols by matching the sequence files (HG-U133_Plus_2 Probe Sequences, FASTA format, August 20, 2008) of the probe sets and the annotation files of GENCODE (release 37). Eventually, 3811 lncRNAs were acquired from the GPL570 microarray platform.

Screening of m^{6}A-Related lncRNAs

lncRNA and m^{6}A regulator expression profiles were extracted from the entire annotated transcriptome. The Pearson correlation analysis was performed on the expression levels of lncRNA and m^{6}A regulators using “limma” and “corrplot” package of R software package, to identify the potential m^{6}A-related lncRNAs (m^{6}ARLncRNAs) with the conditions (|Pearson R| > 0.4 and P < 0.001) in the TCGA dataset. Then, we further filtered the m^{6}ARLncRNAs by taking the intersection of lncRNAs in the TCGA and GEO datasets.

Univariate Cox regression analysis

Univariate Cox regression analysis was conducted in the TCGA dataset using the “survival” module of R software to evaluate the relationship between the expression level of m^{6}ARLncRNAs and the overall survival of patients (P < 0.05).

Gene co-Expression analysis

Co-expression analyses were conducted to visualize the correlation of lncRNAs obtained from univariate Cox regression analysis, with 23 m^{6}A regulators using “corrplot”, “psych”, “ggthemes”, and “Hmisc” modules of R software.

Consensus clustering analysis based on the expression of m^{6}ARLncRNAs

We employed consensus unsupervised clustering analysis in the TCGA dataset to explore distinct molecular patterns based on the expression of m^{6}ARLncRNAs obtained from univariate Cox regression analysis using “ConsensusClusterPlus” and “ggplot2” modules of R software [34].

Relationship between consensus clustering and the clinical features, clinical outcomes of NSCLC

We first used the chi-square test to compare the distribution of age, gender, survival status, and stage among different clusters and presented the results utilizing the “pheatmap” module in R software. Subsequently, the differences in overall survival (OS) among different clusters were calculated by employing the Kaplan-Meier method and visualized by using the “survival” and “survminer” modules in R software.

Immune infiltration level analysis

CIBERSORT was employed to quantitatively determine the leukocyte expression signatures on the expression RPKMs of mRNAs with the validated LM22 gene signature matrix of leukocytes, which identifies 22 types of tumor-infiltrating immune cells at CIBERSORT web tool (http://cibersort.stanford.edu/), setting no quantile normalization and 1.000 permutations as parameters [35]. All samples were performed with two models: relative and absolute modes. The relative mode infers the relative cellular fraction for each cell of the LM22 gene signature matrix and absolute modes calculate a score indicative of the absolute proportion of each cell type.

Gene set enrichment analysis (GSEA)

To identify the possible functions of m^6ARlncRNAs, we conducted GSEA functional enrichment analysis of their mRNA partners among high- and low-risk score groups using GSEA_4.1.0 software [36]. To verify the significant enrichment, a false discovery rate (FDR) of less than 0.25 and a *P* value less than 0.05 were considered statistically significant.

Construction of prognostic m^6ARlncRNA signature (m^6ARLncSig)

The LASSO Cox regression algorithm was carried out to choose and shrink m^6ARLncRNAs obtained from univariate Cox regression analysis with “glmnet” and “survival” modules in R software. Then the screened m^6ARLncRNAs were subjected to multivariate Cox proportional hazard regression analysis to obtain the optimal candidates and construct a prognostic m^6ARLncRNAs signature (m^6ARLncSig) in the training dataset. The receiver operating characteristic (ROC) analysis was conducted and the areas under curve (AUC) values were calculated to evaluate the performance of the prognostic model by using the “survivalROC” package. The m^6ARLncSig score for the signature was calculated by the formula:

$$m^6ARLncSig=\sum_{i=1}^{n} coef * Xi$$

where “coef” is the regression coefficient and “Xi” is the expression level of each m^6ARLncRNA in the prognostic model. In terms of the median m^6ARLncSig score, patients were further divided into two subgroups: a low-risk group and a high-risk group, respectively. The Kaplan-Meier method was performed to plot the survival curves of the two groups using “survival” and “survminer” modules in R software. $P < 0.05$ was considered to be statistically significant.

External validation

Three datasets, including GSE37745 (N=196), GSE30219 (N=268), and GSE31210 (N=226), were used to further validate the prognostic performance of the m^{6}ARLncSig model. We calculated the risk scores of patients based on the above-mentioned formula. Survival curves of the two groups were generated by using Kaplan-Meier analysis (log-rank test) using “survival” and “survminer” modules in the R software package. Statistical significance was set at a P value less than 0.05.

NSCLC samples and quantitative real-time polymerase chain reaction (qRT -PCR)

We included 46 NSCLC patients who had undergone surgical treatments at the Department of Thoracic Surgery of Wuhan Union Hospital from 2014 to 2019. We re-stored the samples in liquid nitrogen for later use. The study was approved by the institutional ethics committees of Wuhan Union Hospital and informed consent was obtained from and signed by each enrolled patient.

To evaluate the expression levels of m^{6}ARlncRNAs, total RNA from clinical samples was extracted by using RNA Trizol reagent (Invitrogen, Carlsbad, CA, United States) according to the instructions and cDNA was synthesized by using a reverse transcription kit (Takara, Dalian, China). Quantification of m^{6}ARlncRNAs was conducted by using an SYBR Green PCR Kit (Takara) and Real-Time PCR System (Applied Biosystems, Carlsbad, CA, USA). The corresponding primers sequences used in our study were as follows: SNHG30 Forward: 5’-GAGAAGCAAAGGAGAGGGAA-3’, and Reverse 5’CGGGAAGACTGAGAGTGAGG-3'; AL021328 Forward 5’-CAGACCCGACCTCTTCCCTC-3', and Reverse 5’-CCTCCTCGCTTTCTCTCCCA-3'; AC024060.2 Forward 5'-TTTTACTTTACCATGCTCAT-3’, and Reverse 5'-AGACTGTATTTCTGTTTTCC-3’; AL137003.1 Forward 5'-CACAAATAAGGCTGGGGAGA-3’, and Reverse 5'-TTGGTGGAGGGTCAAGAAAG-3’; LINC01138 Forward 5’-GACTGTGTCATACTTCCCAT-3’, and Reverse 5’-TTTTGTTGTTCCAACTGCTC-3’; SEPSECS-AS1 Forward 5’-TAAGAATACAACCACGAAAAACA-3’, and Reverse 5’-CAAACTGAGAATCAAATCAATAA-3’; AL034550.1 Forward 5’-TGGGCTGCTTGGAGATGGGC-3’, and

Reverse 5’-CAGGTGGGGTGGGATGGGGG-3’; ITGA9-AS1 Forward 5’-TTTCTCCATTTTCTGTCTTCA-3’, and Reverse 5’-TAGGCAACCGCTAATCTACTT-3’; AP001347.1 Forward 5’-GATAAAGCAGATGAAAATAG-3’, and Reverse 5’-ATCAAAGAAAAAATACACAC-3’; AC083843.2 Forward 5’-GATGGGGAGGAAAGGGATTA-3’, and Reverse 5’-GACACGACAGGGCTTGGAGA-3’; SNHG12 Forward 5’-GGACCTATGGAGTTGGGACAAT-3’, and Reverse 5’-AAGTTCAGTAGCACACTGCATAA-3’; TSPOAP1-AS1, Forward 5’- TCGCCGATGGGAAGAACAAC-3’, and Reverse 5’- AAGAGGAAGTCTAAGCGCCG-3’; GAPDH Forward 5’-ATGTCCCAGCTCTCCTCCACC-3', and Reverse 5’-CTACATTCGGGAGGGCGGGCT-3'.

Relationship between m^{6}ARLncSig and the clinical features, clinical outcomes of NSCLC patients

The clinical features among different groups (high-risk group, low-risk group) were compared using the Chi-square test and visualized using the “pheatmap” package and “ggpubr” in R software. Statistical significance is indicated by *, $P < 0.05$; **, $P < 0.01$; and ***, $P < 0.001$.

Relationship between immunity, gene expression, and m^{6}ARLncSig

We employed the single-sample gene set enrichment analysis (ssGSEA) [37] to quantify the tumor-infiltrating immune cell subgroups and assess their immune function in the high- and low-score groups. Furthermore, the expression of the immune checkpoint was also determined in the two groups.

Independence of the m^{6}ARLncSig from other clinical factors in prognostic value and clinical stratification analysis

To know whether m^{6}ARLncSig can serve as a prognostic factor independent of other factors (age, gender, and tumor stage), univariate and multivariate Cox regression analyses were performed in the training dataset, testing dataset, GSE31210, GSE30219, and GSE37745, by using the “survival” module in the R software package. A P-value less than 0.05 was considered to be of significance. A stratification analysis was conducted to further validate the prognostic performance of m^{6}ARLncSig. Patients in the TCGA set were stratified into a young-patient group (N=414) and an old-patient group (N=534) according to the median age of 65 years. Patients were divided into a male-patient group (N=579) and a female-patient group (N=384). All NSCLC patients were also stratified into an early-stage group (at tumor stage I or II) (N=760) and a late-stage group (at tumor stage III-IV) (N=191). Kaplan-Meier method and log-rank test were used to compare the OS between subgroups in R software with the “survival” and “survminer” package.

Significance of the m^{6}ARLncSig scoring model in the clinical treatment

To evaluate the application of the m^{6}ARLncSig model in the clinical treatment of NSCLC, we calculated the IC50 of chemotherapeutic agents administered in the high-risk group and the low-risk group with “pRRophetic” and “ggpubr” modules of R software.

Establishment and validation of a nomogram scoring system

Based on multivariate Cox regression analysis that assesses the independent prognostic significance of m^{6}ARLncSig and other clinical variables, a nomogram was constructed internally in the training dataset and externally validated in the testing dataset and the whole TCGA dataset for predicting 3-, and 5-year survival. The ROC curve and calibration plot were drawn and the concordance index (C-index) was calculated to estimate the predictive performance and discriminative power of the nomogram scoring system using the “survivalROC” module of R software.

Statistical analysis

We used the Chi-squared test and Fisher's exact test to assess the differences in categorical data between different datasets and groups and Mann–Whitney U test or Student *t*-test to compare the quantitative data. A P value less than 0.05 was considered to be statistically significant. All the statistical analyses

were performed using R version 4.0.2 (https://www.r-project.org) and visualized with the corresponding functional package.

Results

Identification of m^{6}ARLncRNAs in Patients with NSCLC

At first, we downloaded transcriptome information and clinical data from TCGA and GEO datasets and identified 14056 and 3811 lncRNAs, respectively. Thereafter, we extracted the expression levels of 23 m^{6}A regulators from previously published articles. The Pearson correlation analysis was performed based on the expression level of lncRNAs and m^{6}A regulators to identify the potential m^{6}ARLncRNAs with the |Pearson R| > 0.4 and P < 0.001 in the TCGA dataset. As a result, 1651 m^{6}ARLncRNAs were identified. Consequently, a total of 491 m^{6}ARLncRNAs were selected by taking the intersection of GEO and TCGA datasets. Afterward, univariate Cox regression was performed to screen prognostic m^{6}ARLncRNAs, and eventually, 41 m^{6}ARLncRNAs were found to bear a significant association with the overall survival of NSCLC patients (P < 0.05, Fig. S1a). The correlations between the m^{6}ARLncRNAs and 23 m^{6}A regulators in the TCGA dataset are illustrated in Fig. 1b and Supplementary Table 1. Meanwhile, the difference in m^{6}ARLncRNAs between NSCLC and normal tissues was depicted in Fig. S1b.

Identification of distinct molecular patterns mediated by 41 m^{6}ARLncRNAs

To understand the effect of m^{6}ARLncRNAs on the development of NSCLC, consensus unsupervised clustering analysis was conducted in the TCGA dataset based on the expression level of m^{6}ARLncRNAs obtained from univariate Cox regression analysis. The plot showed the relative change in area under the cumulative distribution function (CDF) curve in the k-means (2 to 9) unsupervised clustering of NSCLC (Fig. S2a-b). We also showed the tracking plots of subgroups for k = 2 to k = 9 (Fig. S2c). Lastly, two m^{6}ARLncRNAs-associated clusters were determined and were dubbed cluster 1 (n=705) and cluster 2 (n=258), respectively (Fig. S2d).

To further examine the clinical characteristics of subgroups identified by consensus clustering, we plotted a heatmap, which showed that the expression of m^{6}ARLncRNAs varied and cluster 2 patients had relatively higher expression in most m^{6}ARLncRNAs, such as AC011379.2, AC135050.6, PSMA3-AS1, and AL359921.1. Moreover, the results revealed significant differences in the clinical N stage, age, and gender among clusters (Fig. 2a). To put it simply, cluster 2 was significantly correlated with female sex (P < 0.001), lower N stage (P < 0.01), and younger age (P < 0.05) when compared to cluster 1, as shown in Fig. 2a. In addition, patients with cluster 2 had a longer OS than their counterparts with cluster 1 (P = 0.024) (Fig. 2b). Our findings suggested that the clustering subgroups were closely correlated to the clinical features and clinical outcomes of NSCLC patients.

We investigated the level of immune cell infiltration between the two clusters. The results showed that cluster 1 had a higher infiltration level of neutrophils, macrophages, and activated memory CD4 T cells

(Fig. 2c and Fig. S3a-c). Conversely, cluster 2 was more correlated with the infiltration of monocytes, regulatory T cells, naive B cells, and activated NK cells (Fig. S3d-g).

Construction and validation of prognostic m^{6}ARLncSig

The aforementioned results showed that the m^{6}ARLncRNAs could divide patients into two distinct molecular clustering patterns, which were indicative of two significantly different outcomes in terms of survival. To further explore the prognostic implication of m^{6}ARLncRNAs in NSCLC, the 963 patients were randomly assigned into a training dataset (N=483) and a testing dataset (N=480) patients at a 1:1 ratio. LASSO Cox regression analysis and multivariate Cox proportional hazard regression model for those 41 m^{6}ARLncRNAs were conducted to further select a robust and effective risk model for prognosis prediction (Fig. 3a-b). As a result, twelve of 41 m^{6}ARLncRNAs were chosen to construct a prognostic m^{6}ARLncRNAs signature (m^{6}ARLncSig) (Fig. 3c). On the basis of the expression levels of 12 m^{6}ARLncRNAs and their coefficients using the following computational formula: m^{6}ARLncSig score = (0.177 × Expression AC024060.2) + (0.308 × Expression LINC01138) + (0.455 × Expression AL034550.1)+ (0.577 × Expression AP001347.1) + (-0.947 × Expression SEPSECS-AS1) + (-0.895 × Expression ITGA9-AS1) + (-0.832 × Expression TSPOAP1-AS1) + (-0.466 × Expression AL021328.1) + (-0.240 × Expression AC083843.2) + (-0.220 × Expression AL137003.1) + (-0.069 × Expression SNHG30) + (-0.056 × Expression SNHG12), we calculated the risk score of m^{6}ARLncSig in the TCGA training and testing dataset (Fig. 3c and Table S2). In the m^{6}ARLncSig model, four m^{6}ARLncRNAs (AC024060.2, LINC01138, AL034550.1, and AP001347.1) were identified to be high-risk factors, with a value of hazard ratio (HR) more than 1, which is indicative of a poor NSCLC prognosis. On the other hand, eight lncRNAs (SNHG12, ITGA9-AS1, AC083843.2, TSPOAP1-AS1, SNHG30, AL021328.1, AL137003.1, and SEPSECS-AS1) with an HR less than 1 were found to be protective factors, indicating a better survival relevance of their upregulated expression.

Subsequently, with the median risk score as the cut-off, patients in the training and testing dataset were divided into a high- and low-risk group. The distribution of the risk scores, OS, OS status, and expression profiles of the m^{6}ARLncRNAs in the TCGA training and test set was given in Fig. 4a-b. To assess the prognostic value of m^{6}ARLncSig in NSCLC patients, the survival analysis was performed in the high-risk group (N=242) and the low-risk group (N=241) in the training dataset, whose median score was 1.070. The result revealed that patients in the low-risk group had more favorable outcomes than those in the high-risk group ($P < 0.001$, Fig. 4c) and the AUC of the ROC curves in the training dataset was 0.694 for the 1-year OS prediction (Fig. 4d). To verify the prognostic m^{6}ARLncSig value of the risk score for survival prediction, we evaluated survival conditions in the testing dataset and the median risk score in the testing dataset was 1.125. Our findings confirmed that the clinical outcome of patients in the low-risk group (N=240) was better than that of patients in the high-risk group (N=240) in the testing dataset (P=0.030, Fig. 4e), with an AUC value of 0.631 (Fig. 4f). Collectively, our results suggest that m^{6}ARLncSig promises to be a good survival predictor.

External and experimental validation of the prognostic significance of m^{6}ARLncSig

To further determine the prognostic significance of the m^6ARLncSig, we retrieved three independent datasets, including GSE37745, GSE31210, and GSE30219, from the GPL570 microarray platform in GEO datasets to conduct external validation of m^6ARLncSig. According to the above-mentioned formula, Kaplan-Meier survival analysis indicated that all the patients in the low-risk group had a significantly longer survival time when compared with the high-risk group in GSE31210 ($P < 0.001$), GSE37745 ($P = 0.002$), and GSE30219 ($P < 0.001$) and AUC of three datasets was 0.881, 0.700, and 0.643, respectively (Fig. 5a-f). We also conducted qRT-PCR verification for twelve m6ARlncRNAs with our NSCLC tissue samples to validate the prognostic significance. The Kaplan-Meier analysis showed that patients in the low-risk group had better survival outcomes than patients in the high-risk group ($P = 0.012$, Fig. 5g). The AUC of the ROC curves of m^6ARLncSig was 0.763 for the three-year survival prediction of patients (Fig. 5h). Taken together, these results further validated that m^6ARLncSig had high validity for survival prediction in NSCLC.

Correlation between m^6ARLncSig and clinical parameters

Then, we tried to find a correlation between m^6ARLncSig and clinical features. Our results showed that m^6ARLncSig was significantly correlated with age, tumor stage, and molecular clustering patterns ($P < 0.05$, Fig. 6a). Briefly, patients in the advanced-age group tended to have significantly higher m^6ARLncSig scores than those in the young-patient group ($P = 0.016$, Fig. 6b). Patients at the advanced tumor stage had significantly higher m^6ARLncSig scores than those at the early stage ($P = 0.001$, Fig. 6c). Moreover, patients in cluster 1 had significantly higher m^6ARLncSig scores than patients in cluster 2 ($P < 0.001$, Fig. 6d).

Gene set enrichment analysis (GSEA) of the Two m^6ARLncSig subgroups

To understand the regulatory mechanisms, GSEA was employed to conduct the signaling pathway enrichment of the two m^6ARLncSig subgroups (high- and low-risk groups) in the TCGA dataset. The results revealed that pathways such as ECM receptor interaction (NES = 2.143, normalized $P = 0.002$), focal adhesion (NES = 2.052, normalized $P = 0.002$), adherence junction (NES = 1.605, normalized $P = 0.032$) were enriched in high m^6ARLncSig group ($P < 0.05$, Fig. S4). These findings indicated that m^6ARLncRNAs might participate in the progression of NSCLC via the above pathways, which partially explains the poor survival in the high score group and might help us gain insight into the implication of m^6ARLncSig.

Assessment of independent prognostic significance of risk score and clinical stratification analysis

To explore whether m^6ARLncSig score, as a prognostic factor, is independent of other clinical features, univariate and multivariate Cox regression analyses were conducted on four variables: age, gender, tumor stage, and m^6ARLncSig (Table 2). The univariate Cox analysis revealed that m^6ARLncSig ($P < 0.001$, HR = 1.260, 95% CI = 1.195-1.328) was significantly correlated with the OS in the training dataset. Multivariate

Cox analysis further indicated that m^{6}ARLncSig (P < 0.001, HR = 1.222, 95% CI = 1.156-1.291) was an independent risk factor of OS. These results were also validated in the testing dataset (P < 0.001, HR = 1.212, 95% CI = 1.103-1.332), GSE37745 (P < 0.001, HR = 1.626, 95% CI = 1.252-2.113), GSE31210 (P =0.001, HR = 1.006, 95% CI = 1.002-1.009), and GSE30219 (P = 0.006, HR = 1.396, 95% CI = 1.101-1.769) dataset (Table 2).

To assess whether m^{6}ARLncSig is still of prognostic value in various subgroups, we conducted a clinical stratification analysis. Kaplan-Meier analysis showed that patients with high-risk m^{6}ARLncSig scores exhibited worse prognosis compared with those with low-risk m^{6}ARLncSig scores across all six subgroups, i.e., young-patient group (P = 0.002, Fig. 7a), old-patient group (P < 0.001, Fig. 7b), early-stage group (P < 0.001, Fig. 7c), late-stage group (P = 0.019, Fig. 7d), male-patient group (P < 0.001, Fig. 7e), and female-patient group (P = 0.004, Fig. 7f). Collectively, these results indicated that m^{6}ARLncSig was not only strongly associated with clinical outcomes and clinical features, but also was an independent predictor for the prognosis of NSCLC.

Relationship between m^{6}ARLncSig and tumor immunity and therapeutic sensitivity

To know the relationship between the m^{6}ARLncSig and tumor immune microenvironment, we used the CIBERSORT algorithm to estimate the difference in tumor-infiltrating immune cells between high- and low-risk groups. The Wilcoxon signed-rank test showed that the m^{6}ARLncSig score was positively correlated with the infiltration level of the neutrophils (P < 0.001, Fig. 8a), macrophages M0 (P < 0.001, Fig. 8b) and activated memory CD4 T cells (P = 0.049, Fig. 8c). And a significant negative correlation was observed between the m^{6}ARLncSig score and the infiltration level of activated mast cells (P = 0.022, Fig. 8d), monocytes (P =0.02, Fig. 8e), and naive B cells (P =0.022, Fig. 8f). We also conducted a correlation analysis between immune cell subpopulations and related functions by using ssGSEA of TCGA-NSCLC. We found that T cell functions, including APC_co_stimulation, CCR, MHC_class_I, and parainflammation exhibited significant differences between the high-risk and low-risk groups (Fig. 8g). We further found significant differences in the expression at all the critical immune checkpoints between the two groups (P<0.05, Fig. 8h). Overall, these results confirmed that m^{6}ARLncSig was associated with the immune microenvironment, immune response, and immune function.

Then, we examined if there existed an association between m^{6}ARLncSig level and radiotherapeutic response and chemotherapeutic sensitivity in the treatment of NSCLC. The result showed that most patients with high m^{6}ARLncSig did not respond or responded poorly to radiotherapy as compared to the low-risk group (P = 0.017, Fig. 9a). We further looked into the association between m^{6}ARLncSig and half inhibitory concentration (IC50) of chemotherapeutics. m^{6}ARLncSig was positively associated with half inhibitory concentration (IC50) of chemotherapeutics, such as lenalidomide and methotrexate (P < 0.001, Fig. 9b-c). On the other hand, the m^{6}ARLncSig score was negatively correlated with IC50 of gefitinib, gemcitabine, paclitaxel, and docetaxel (P < 0.05, Fig. 9d-g). These findings indicated that the m6ARLncSig was intimately correlated with the radiotherapeutic response and chemotherapeutic

sensitivity, suggesting that m6ARlncRNAs are involved in the mechanisms of radiotherapeutic and chemotherapeutic responses and might, to some extent, dictate the prognosis of NSCLC patients.

Construction and validation of a nomogram for survival prediction of NSCLC

To simplify the model for the clinical prediction of OS in NSCLC patients, we established a nomogram scoring system in the training dataset (Fig. 10a). Our results showed that the nomogram had an improved performance of OS prediction with a C-index of 0.66 and AUCs of ROC for 3-, and 5-year survival predictions of 0.715, 0.694, respectively (Fig. 10b). As expected, our findings were validated in the testing and the entire TCGA dataset with a C-index of 0.671 and 0.666, respectively. The AUCs of ROC for 3-, and 5-year survival predictions were 0.695, 0.636 in the testing dataset and 0.698, 0.666 in the entire TCGA dataset (Fig. 10c-d). Calibration plots showed that consistency was excellent between the observed and predicted values for the prediction of 3- and 5-year OS in the testing, training, and the entire TCGA dataset (Fig. 10e-j). Therefore, these findings indicated that the nomogram possessed a good prospect of clinical application for prognosis evaluation and treatment planning of NSCLC.

Discussion

N6-methyladenosine (m^6A) modification is deemed as the most common, abundant, and conserved internal transcript modification [12, 38]. Research efforts are mainly directed at the association between m^6A RNA modification regulators and the progression of lung cancer and the prognosis. For instance, the prognostic value of 13 common m^6A RNA modification regulators was comprehensively investigated by using The Cancer Genome Atlas database [27]. The correlation between m^6A regulators and the tumor microenvironment was also studied and m^6Sig (a scoring tool) was shown to be not only positively correlated with PD-L1 expression but also to reflect the change in the tumor microenvironment [39].

It is worth noting that m^6A has been discovered in a wider array of non-coding RNAs (ncRNAs), and m^6A peaks were present in approximately 67% lncRNAs of 3' UTRs [27, 40]. m^6A methylation reportedly acted as a lncRNA structural switch, participated in the lncRNA-mediated ceRNA modulation, and enhanced the stability of lncRNA to serve its functions, thereby influencing cancer development and progression [28]. For instance, m^6A methyltransferase-like 3 (METTL3)-induced lncRNA ABHD11-AS1 was closely correlated with the unfavorable prognosis of NSCLC patients [41]. So it is necessary to look into the interaction between the functions of lncRNAs and m^6A modifications. But how m^6A works in a lncRNA-dependent manner during NSCLC progression remains unknown.

In the present study, we identified 491 m^6A-related lncRNAs by using co-expression analysis between 23 m^6A regulators and lncRNAs in 1835 NSCLC patients from TCGA (N = 1145) and GEO (N = 690) datasets and explored their prognostic implication and clinical relevance in NSCLC. We found that the m^6ARLncRNAs could divide patients into two distinct molecular clustering patterns, which showed significantly different survival outcomes and immune cell infiltration. These findings prompted us to

assume whether m^{6}ARLncRNAs might serve as a prognostic predictor for NSCLC patients. Therefore, we conducted LASSO Cox regression and multivariate Cox proportional hazard regressions to construct a prognostic model-m^{6}ARLncRNAs signature, which was dubbed m^{6}ARLncSig. Patients in the high- and low-risk groups divided in terms of m^{6}ARLncSig scores exhibited significantly different survival outcomes, tumor-infiltrating immune cells, radiotherapeutic response, and chemotherapeutic sensitivity. ROC curve evaluation, univariate and multivariate Cox regressions in five datasets (training dataset, testing dataset, GSE37745, GSE31210, and GSE30219) all indicated that m^{6}ARLncSig performed well as a predictor and could serve as a reliable independent prognostic indicator for NSCLC. Moreover, correlation analysis showed that m^{6}ARLncSig bore a significant relationship with clinical features of NSCLC. Last, a nomogram was constructed based on tumor stage, gender, tumor stage, and m^{6}ARLncSig scores to further improve the performance and facilitate the clinical use of the m^{6}ARLncSig model.

The m^{6}ARLncSig model contained twelve m^{6}ARLncRNAs. Some of the m^{6}ARLncRNAs were found to be involved in malignant phenotypes of various cancers, such as LINC01138, SNHG12, ITGA9-AS1, and TSPOAP1-AS1. LINC01138 reportedly interacted with PRMT5, thus promoting lipid desaturation and cell proliferation in clear cell renal cell carcinoma (ccRCC) [42] and could be used as a prognostic indicator for hepatocellular carcinoma [43] and prostate cancer [44]. Knockdown of SNHG12 suppressed tumor metastasis and epithelial-mesenchymal transition via the Slug/ZEB2 signaling pathway [45] and mediated doxorubicin resistance of osteosarcoma through miR-320a/MCL1 axis [46]. In our study, SNHG12 was found to be a protective factor, which is inconsistent with the findings of previous studies. This inconsistency might be ascribed to the fact that our univariate and multivariate analyses excluded the normal samples, whose overall survival data were not available in the TCGA dataset. It has been reported that ITGA9-AS1 was found to be positively correlated with the survival probability in breast cancer patients and was used as a diagnostic or prognostic marker for tamoxifen resistance [47]. TSPOAP1-AS1 over-expression was found to be correlated with a protracted OS in patients with pancreatic ductal adenocarcinoma (PDAC) [48]. The role of the remaining prognostic m^{6}ARLncRNAs in lung cancer has hardly been studied, and reports on how the m^{6}ARLncRNAs interact with m^{6}A regulators have been scanty. Therefore, further experimental confirmation is warranted to fully understand the functional role of prognostic lncRNAs in NSCLC both *in vivo* and *in vitro*.

It is generally believed that the response to chemotherapy and radiotherapy may vary with patients due to the heterogeneity of NSCLC and only a small subset of NSCLC patients respond well. Therefore, early assessment of treatment efficacy in terms of predictive biomarkers is completely crucial for NSCLC patients to select the best treatment strategy. In our study, we showed that patients with high m^{6}ARLncSig tended to respond poorly to radiotherapy and chemotherapeutics (gemcitabine, paclitaxel, docetaxel, and gefitinib) as compared to those in the low-risk group. This finding can, to some extent, help clinicians individualize treatment for different NSCLC patients based on the m^{6}ARLncSig model.

In this study, we, for the first time, demonstrated that m^{6}ARLncSig was highly associated with clinical features of NSCLC and could independently predict patients' OS, chemotherapeutic sensitivity, and

radiotherapeutic efficacy, by utilizing the public TCGA dataset and validated the model in external GEO datasets. Our study not only provides an indicator for clinical stratified management of patients but also paves the way to future investigation of m^6ARLncRNAs as the therapeutic targets of NSCLC. Nonetheless, our study had some limitations. Firstly, our study was based only on computational reasoning, the biological and functional studies on the mechanism of m^6A-related lncRNAs in the progression of NSCLC worth further validation. Additionally, although the m^6ARLncSig was validated in TCGA testing and three GEO datasets, more independent NSCLC cohorts from different platforms are required to further validate the m^6ARLncSig in terms of robustness and applicability.

Conclusions

In conclusion, we could stratify the patients with NSCLC in terms of m^6ARLncSig as an independent prognostic indicator. We also demonstrated that the m^6ARLncSig model could predict the radiotherapeutic response and chemotherapeutic sensitivity of NSCLC patients.

Abbreviations

NSCLC: non-small cell lung cancer; m^6A: N6-methyladenosine; non-coding RNA: (ncRNA); lncRNA: long non-coding RNAs; TCGA: The Cancer Genome Atlas; GEO: Gene Expression Omnibus; CDF: cumulative distribution function; OS: overall survival; GSEA: Gene set enrichment analyses; m^6ARLncRNA: m^6A-related RNA; m^6ARLncSig: m^6A-related RNA signature; ROC: receiver operating characteristic; AUC: areas under curve; HR: hazard ratio; CI: confidence intervals; ccRCC: clear cell renal cell carcinoma.

Declarations

Ethics approval and consent to participate

This study was approved by the ethical committee of Tongji Medical College, Huazhong University of Science and Technology ([2010]IEC(S202)) and all patients provided informed consent.

Consent for publication

All the authors listed have approved the final manuscript and agreed to publish the manuscript.

Availability of data and materials

Data available upon request.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

The study was supported by grants from the National Major Scientific and Technological Special Project for “Significant New Drugs Development” (Grant No.2019ZX09301-001) and Open Funds for the Hubei Province Key Laboratory of Molecular Imaging (Grant No: 2020fzyx012).

Authors’ Contributions

YJ, WX, and WG contributed to the conception and design of the study. YJ organized the database. WX, and WG performed the statistical analysis. WX and WG wrote the first draft of the manuscript. JX, QH, JF, QT, ZY, YL, and GY provided comments during the writing. All authors contributed to manuscript revision, read, and approved the submitted version.

Acknowledgements

We are sincerely acknowledge the contributions from the TCGA (https://portal.gdc.cancer.gov/) project and the GEO (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) project.

References

1. Wang R, Yamada T, Kita K, Taniguchi H, Arai S, Fukuda K, et al. Transient IGF-1R inhibition combined with osimertinib eradicates AXL-low expressing EGFR mutated lung cancer. Nature communications. 2020; 11: 4607.
2. Hubers AJ, Prinsen CFM, Sozzi G, Witte BI, Thunnissen E. Molecular sputum analysis for the diagnosis of lung cancer. Br J Cancer. 2013; 109: 530-7.
3. Zhen D, Wu Y, Zhang Y, Chen K, Song B, Xu H, et al. m(6)A Reader: Epitranscriptome Target Prediction and Functional Characterization of N (6)-Methyladenosine (m(6)A) Readers. Frontiers in cell and developmental biology. 2020; 8: 741.
4. The L. Lung cancer: some progress, but still a lot more to do. Lancet (London, England). 2019; 394: 1880.
5. Allemani C, Weir HK, Carreira H, Harewood R, Spika D, Wang X-S, et al. Global surveillance of cancer survival 1995-2009: analysis of individual data for 25,676,887 patients from 279 population-based registries in 67 countries (CONCORD-2). Lancet (London, England). 2015; 385.
6. Öjlert ÅK, Halvorsen AR, Nebdal D, Lund-Iversen M, Solberg S, Brustugun OT, et al. The immune microenvironment in non-small cell lung cancer is predictive of prognosis after surgery. Mol Oncol. 2019; 13: 1166-79.
7. Woodard GA, Jones KD, Jablons DM. Lung Cancer Staging and Prognosis. Cancer Treat Res. 2016; 170: 47-75.
8. Bugge AS, Lund MB, Valberg M, Brustugun OT, Solberg S, Kongerud J. Cause-specific death after surgical resection for early-stage non-small-cell lung cancer. European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery. 2018; 53: 221-7.

9. Wei S, Tian J, Song X, Wu B, Liu L. Causes of death and competing risk analysis of the associated factors for non-small cell lung cancer using the Surveillance, Epidemiology, and End Results database. J Cancer Res Clin Oncol. 2018; 144: 145-55.
10. Bodi Z, Button JD, Grierson D, Fray RG. Yeast targets for mRNA methylation. Nucleic Acids Res. 2010; 38: 5327-35.
11. Fu Y, Dominissini D, Rechavi G, He C. Gene expression regulation mediated through reversible m□A RNA methylation. Nat Rev Genet. 2014; 15: 293-306.
12. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell. 2012; 149: 1635-46.
13. Song Y, Xu Q, Wei Z, Zhen D, Su J, Chen K, et al. Predict Epitranscriptome Targets and Regulatory Functions of N (6)-Methyladenosine (m(6)A) Writers and Erasers. Evolutionary bioinformatics online. 2019; 15: 1176934319871290.
14. Chen M, Wei L, Law C-T, Tsang FH-C, Shen J, Cheng CL-H, et al. RNA N6-methyladenosine methyltransferase-like 3 promotes liver cancer progression through YTHDF2-dependent posttranscriptional silencing of SOCS2. Hepatology (Baltimore, Md). 2018; 67: 2254-70.
15. Zhang S, Zhao BS, Zhou A, Lin K, Zheng S, Lu Z, et al. mA Demethylase ALKBH5 Maintains Tumorigenicity of Glioblastoma Stem-like Cells by Sustaining FOXM1 Expression and Cell Proliferation Program. Cancer cell. 2017; 31.
16. Sun T, Wu Z, Wang X, Wang Y, Hu X, Qin W, et al. LNC942 promoting METTL14-mediated mA methylation in breast cancer cell proliferation and progression. Oncogene. 2020; 39: 5358-72.
17. Anita R, Paramasivam A, Priyadharsini JV, Chitra S. The m6A readers and aberrations associated with metastasis and predict poor prognosis in breast cancer patients. Am J Cancer Res. 2020; 10: 2546-54.
18. Chong W, Shang L, Liu J, Fang Z, Du F, Wu H, et al. m(6)A regulator-based methylation modification patterns characterized by distinct tumor microenvironment immune profiles in colon cancer. Theranostics. 2021; 11: 2201-17.
19. Li F, Wang H, Huang H, Zhang L, Wang D, Wan Y. m6A RNA Methylation Regulators Participate in the Malignant Progression and Have Clinical Prognostic Value in Lung Adenocarcinoma. Front Genet. 2020; 11: 994.
20. Sheng H, Li Z, Su S, Sun W, Zhang X, Li L, et al. YTH domain family 2 promotes lung cancer cell growth by facilitating 6-phosphogluconate dehydrogenase mRNA translation. Carcinogenesis. 2020; 41: 541-50.
21. Zhang C, Huang S, Zhuang H, Ruan S, Zhou Z, Huang K, et al. YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation. Oncogene. 2020; 39: 4507-18.
22. Jin D, Guo J, Wu Y, Du J, Yang L, Wang X, et al. m(6)A mRNA methylation initiated by METTL3 directly promotes YAP translation and increases YAP activity by regulating the MALAT1-miR-1914-3p-YAP axis to induce NSCLC drug resistance and metastasis. J Hematol Oncol. 2019; 12: 135.

23. Wang H, Deng Q, Lv Z, Ling Y, Hou X, Chen Z, et al. N6-methyladenosine induced miR-143-3p promotes the brain metastasis of lung cancer via regulation of VASH1. Molecular cancer. 2019; 18: 181.
24. Jin D, Guo J, Wu Y, Du J, Yang L, Wang X, et al. mA mRNA methylation initiated by METTL3 directly promotes YAP translation and increases YAP activity by regulating the MALAT1-miR-1914-3p-YAP axis to induce NSCLC drug resistance and metastasis. J Hematol Oncol. 2019; 12: 135.
25. Jin D, Guo J, Wu Y, Yang L, Wang X, Du J, et al. mA demethylase ALKBH5 inhibits tumor growth and metastasis by reducing YTHDFs-mediated YAP expression and inhibiting miR-107/LATS2-mediated YAP activity in NSCLC. Molecular cancer. 2020; 19: 40.
26. Zhuang Z, Chen L, Mao Y, Zheng Q, Li H, Huang Y, et al. Diagnostic, progressive and prognostic performance of m(6)A methylation RNA regulators in lung adenocarcinoma. Int J Biol Sci. 2020; 16: 1785-97.
27. Liu Y, Guo X, Zhao M, Ao H, Leng X, Liu M, et al. Contributions and prognostic values of m(6) A RNA methylation regulators in non-small-cell lung cancer. J Cell Physiol. 2020; 235: 6043-57.
28. Chen Y, Lin Y, Shu Y, He J, Gao W. Interaction between N(6)-methyladenosine (m(6)A) modification and noncoding RNAs in cancer. Mol Cancer. 2020; 19: 94.
29. Li Y, Yin Z, Fan J, Zhang S, Yang W. The roles of exosomal miRNAs and lncRNAs in lung diseases. Signal Transduct Target Ther. 2019; 4: 47.
30. Dai F, Wu Y, Lu Y, An C, Zheng X, Dai L, et al. Crosstalk between RNA m(6)A Modification and Non-coding RNA Contributes to Cancer Growth and Progression. Mol Ther Nucleic Acids. 2020; 22: 62-71.
31. Zuo X, Chen Z, Gao W, Zhang Y, Wang J, Wang J, et al. M6A-mediated upregulation of LINC00958 increases lipogenesis and acts as a nanotherapeutic target in hepatocellular carcinoma. J Hematol Oncol. 2020; 13: 5.
32. Zheng Z-Q, Li Z-X, Zhou G-Q, Lin L, Zhang L-L, Lv J-W, et al. Long Noncoding RNA FAM225A Promotes Nasopharyngeal Carcinoma Tumorigenesis and Metastasis by Acting as ceRNA to Sponge miR-590-3p/miR-1275 and Upregulate ITGB3. Cancer research. 2019; 79: 4612-26.
33. Wu Y, Yang X, Chen Z, Tian L, Jiang G, Chen F, et al. mA-induced lncRNA RP11 triggers the dissemination of colorectal cancer cells via upregulation of Zeb1. Molecular cancer. 2019; 18: 87.
34. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010; 26: 1572-3.
35. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015; 12: 453-7.
36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102: 15545-50.
37. Yi M, Nissley DV, McCormick F, Stephens RM. ssGSEA score-based Ras dependency indexes derived from gene expression data reveal potential Ras addiction mechanisms with possible clinical

implications. Sci Rep. 2020; 10: 10258.
38. Yang Y, Hsu PJ, Chen Y-S, Yang Y-G. Dynamic transcriptomic mA decoration: writers, erasers, readers and functions in RNA metabolism. Cell research. 2018; 28: 616-24.
39. Li Y, Gu J, Xu F, Zhu Q, Chen Y, Ge D, et al. Molecular characterization, biological function, tumor microenvironment association and clinical significance of m6A regulators in lung adenocarcinoma. Brief Bioinform. 2020.
40. Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell. 2012; 149: 1635-46.
41. Xue L, Li J, Lin Y, Liu D, Yang Q, Jian J, et al. m(6) A transferase METTL3-induced lncRNA ABHD11-AS1 promotes the Warburg effect of non-small-cell lung cancer. J Cell Physiol. 2021; 236: 2649-58.
42. Zhang X, Wu J, Wu C, Chen W, Lin R, Zhou Y, et al. The LINC01138 interacts with PRMT5 to promote SREBP1-mediated lipid desaturation and cell growth in clear cell renal cell carcinoma. Biochemical and biophysical research communications. 2018; 507: 337-42.
43. Li Z, Zhang J, Liu X, Li S, Wang Q, Di Chen, et al. The LINC01138 drives malignancies via activating arginine methyltransferase 5 in hepatocellular carcinoma. Nature communications. 2018; 9: 1572.
44. Wan X, Huang W, Yang S, Zhang Y, Pu H, Fu F, et al. Identification of androgen-responsive lncRNAs as diagnostic and prognostic markers for prostate cancer. Oncotarget. 2016; 7: 60503-18.
45. Wang Y, Liang S, Yu Y, Shi Y, Zheng H. Knockdown of SNHG12 suppresses tumor metastasis and epithelial-mesenchymal transition via the Slug/ZEB2 signaling pathway by targeting miR-218 in NSCLC. Oncology letters. 2019; 17: 2356-64.
46. Zhou B, Li L, Li Y, Sun H, Zeng C. Long noncoding RNA SNHG12 mediates doxorubicin resistance of osteosarcoma via miR-320a/MCL1 axis. Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie. 2018; 106: 850-7.
47. Zhang X, Gao S, Li Z, Wang W, Liu G. αIdentification and Analysis of Estrogen Receptor Promoting Tamoxifen Resistance-Related lncRNAs. BioMed research international. 2020; 2020: 9031723.
48. Giulietti M, Righetti A, Principato G, Piva F. LncRNA co-expression network analysis reveals novel biomarkers for pancreatic cancer. Carcinogenesis. 2018; 39: 1016-25.

Tables

Table 1 Clinical information of patients with NSCLC in TCGA cohort.

Covariates	Type	Total (N=963)	Training dataset (N=483)	Testing dataset (N=480)	P value
Age (%)	<=65	414 (42.99%)	207 (42.86%)	207 (43.12%)	1
	>65	534 (55.45%)	267 (55.28%)	267 (55.62%)	
	unknow	15 (1.56%)	9 (1.86%)	6 (1.25%)	
Gender (%)	Female	384 (39.88%)	194 (40.17%)	190 (39.58%)	0.9055
	Male	579 (60.12%)	289 (59.83%)	290 (60.42%)	
Tumor Stage (%)	Stage I-II	760 (78.92%)	391 (80.95%)	369 (76.88%)	0.2124
	Stage III-IV	191 (19.83%)	88 (18.22%)	103 (21.46%)	
	unknow	12 (1.25%)	4 (0.83%)	8 (1.67%)	
T Stage (%)	T1-2	810 (84.11%)	398 (82.4%)	412 (85.83%)	0.2479
	T3-4	150 (15.58%)	82 (16.98%)	68 (14.17%)	
	unknow	3 (0.31%)	3 (0.62%)	0 (0%)	
M Stage (%)	M0	717 (74.45%)	347 (71.84%)	370 (77.08%)	0.7209
	M1	30 (3.12%)	13 (2.69%)	17 (3.54%)	
	unknow	216 (22.43%)	93 (19.38%)	123 (25.47%)	
N Stage (%)	N0	617 (64.07%)	313 (64.8%)	304 (63.33%)	0.7131
	N1-3	331 (34.37%)	163 (33.75%)	168 (35%)	
	unknow	15 (1.56%)	7 (1.45%)	8 (1.67%)	

Chi-squared test, $P < 0.05$ means significantly different.

Table 2 Univariate and multivariate Cox regression analysis of the m^6ARLncSig and clinical features for the independent prognostic significance in five datasets.

Variables	Univariable model				Multivariable model			
	HR	95% CI lower	95% CI Higher	P-value	HR	95% CI lower	95% CI higher	P-value
TCGA training set (N=483)								
Age	1.015	0.999	1.031	0.062				
Gender	0.864	0.646	1.155	0.323				
Tumor Stage	1.491	1.273	1.746	<0.001	1.390	1.179	1.640	<0.001
m^6ARLncSig	1.260	1.195	1.328	<0.001	1.222	1.156	1.291	<0.001
TCGA testing set (N=480)								
Age	1.002	0.986	1.019	0.801				
Gender	1.504	1.104	2.050	0.010	1.380	1.005	1.894	0.046
Tumor Stage	1.530	1.316	1.778	<0.001	1.451	1.245	1.692	<0.001
m^6ARLncSig	1.244	1.132	1.366	<0.001	1.212	1.103	1.332	<0.001
GSE37745 (N=196)								
Age	1.025	1.006	1.045	0.010	1.024	1.004	1.044	0.016
Gender	1.096	0.789	1.523	0.585				
Tumor Stage	1.270	1.049	1.539	0.014	1.281	1.054	1.557	0.013
m^6ARLncSig	1.699	1.297	2.226	<0.001	1.626	1.252	2.113	<0.001
GSE31210 (N=226)								
Age	1.025	0.977	1.075	0.306				
Gender	1.519	0.780	2.955	0.219				
Tumor Stage	4.232	2.175	8.236	<0.001	3.153	1.580	6.291	0.001
m^6ARLncSig	1.007	1.004	1.010	<0.001	1.006	1.002	1.009	0.001
GSE30219 (N=268)								
Age	1.038	1.023	1.054	<0.001	1.037	1.022	1.053	<0.001

Gender	1.646	1.033	2.622	0.036	1.300	0.810	2.085	0.2770
Tumor Stage	1.690	1.424	2.007	<0.001	1.692	1.416	2.023	<0.001
m^6ARLncSig	1.732	1.366	2.196	<0.001	1.396	1.101	1.769	0.006

HR, hazard ratio; CI, confidence interval.

Figures

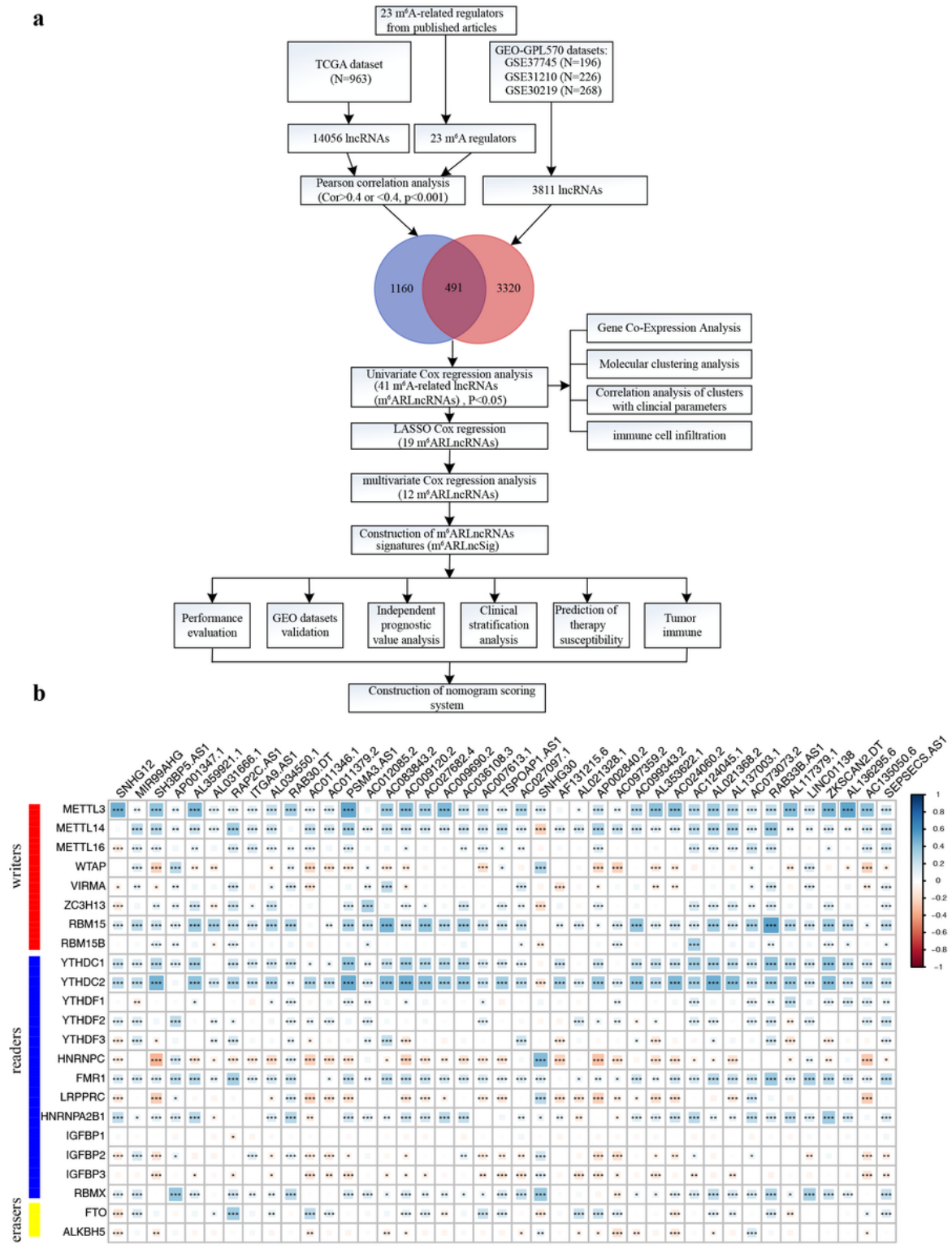

Figure 1

A The study flow chart. b Gene co-expression analysis were performed to visualize the correlation of 41 prognostic lncRNAs and 23 m6A regulators with the | Pearson R| > 0.4 and P < 0.001. *p < 0.05, **p < 0.01, and ***p < 0.001.

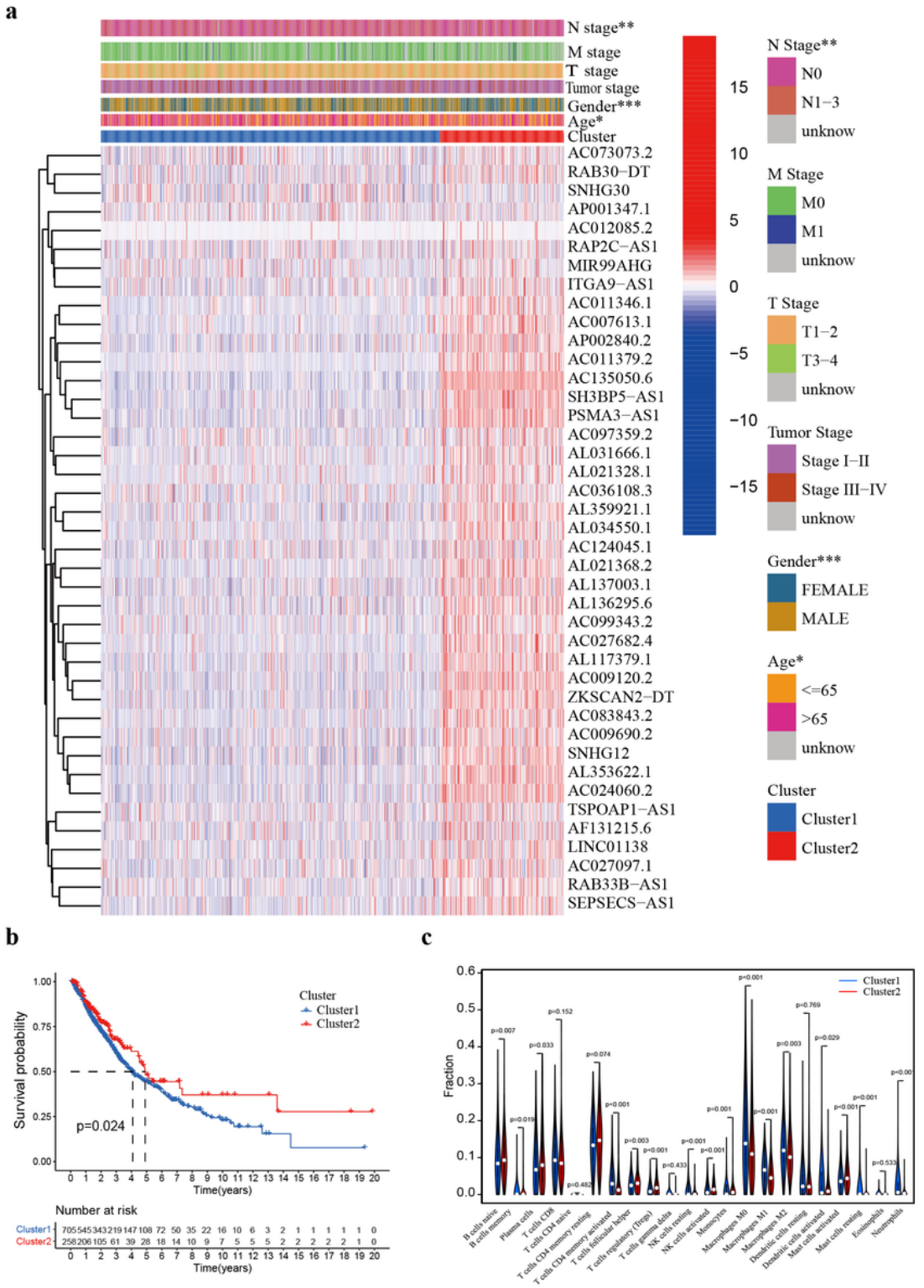

Figure 2

Differential clinical features and clinical outcomes of patients with NSCLC between cluster 1 and cluster 2. a The heatmap was shown the correlation between clinical features and different clusters b Kaplan-Meier method indicated the overall survival (OS) among different clusters. c The infiltrating levels of 22 immune cell types in cluster 1 and cluster 2. *p < 0.05 and **p < 0.01, and ***p < 0.001.

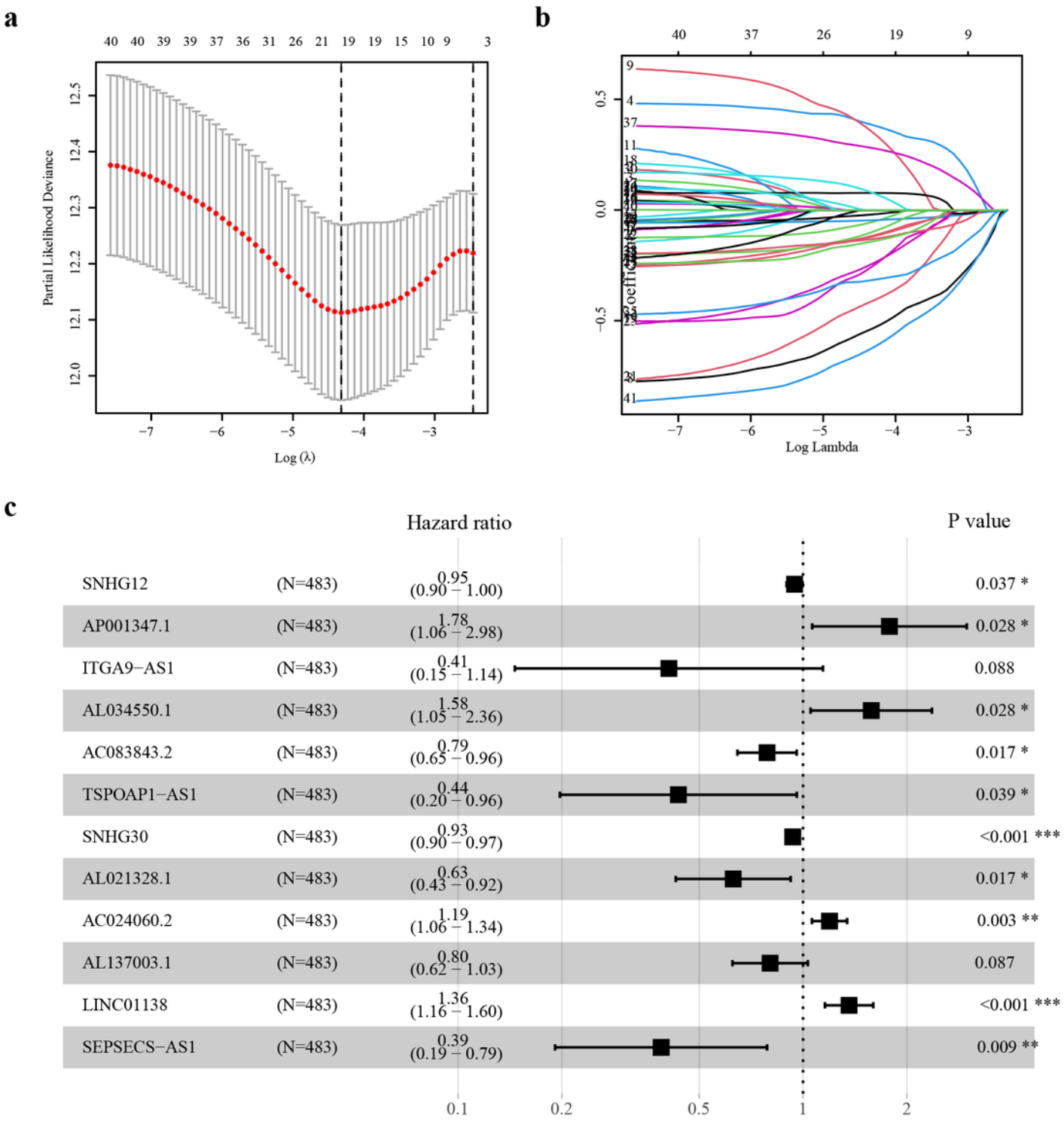

Figure 3

Construction of m6ARLncSig of patients with NSCLC in the training set. a-b The least absolute shrinkage and selection operator (LASSO) regression was performed, calculating the minimum criteria. c Forest plot of multivariate cox regression analysis for prognostic m6ARLncRNAs.

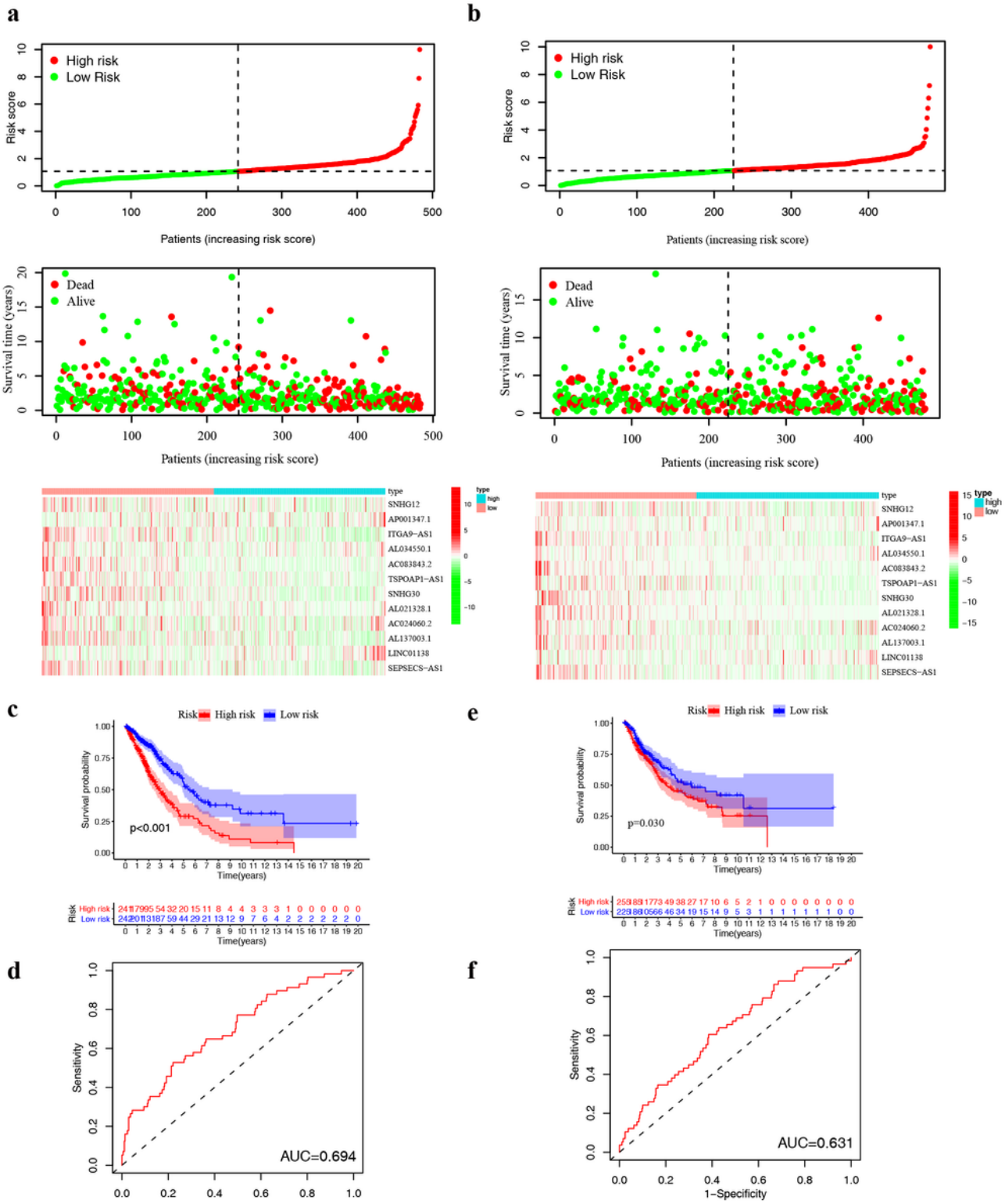

Figure 4

Evaluation and validation of prognostic m6ARLncSig of overall survival (OS) in patients with NSCLC. a-b The distribution of score, OS, and OS status, and heatmap of prognostic m6ARLncRNAs in the TCGA training dataset (a) and the TCGA testing dataset (b). c-f Kaplan–Meier survival curves of the patients in the high- and low-risk groups separated by m6ARLncSig and ROC curves of m6ARLncSig for 1-year survival prediction in the training dataset (c-d) and the testing dataset (e-f).

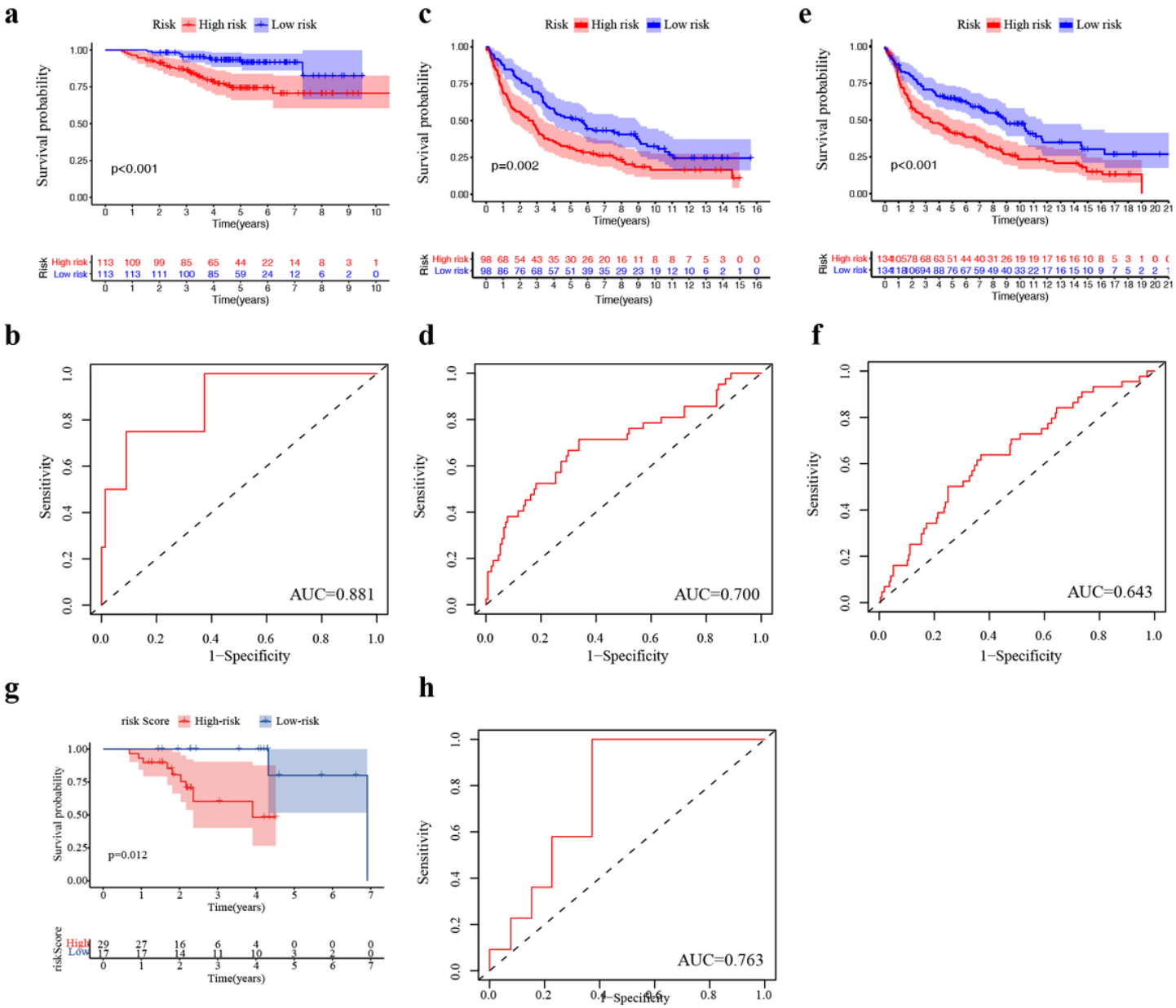

Figure 5

Eternal and experimental validation of m6ARLncSig in patients with NSCLC. a-f Kaplan–Meier survival curves of patients in the high- and low-risk groups separated by m6ARLncSig in GSE31210 (a), GSE37745 (c), and GSE30219 (e). ROC curves for 1-year survival prediction of m6ARLncSig score in GSE31210 (b), GSE37745 (d), and GSE30219 (f). g Kaplan-Meier curve analysis in the high- and low-risk groups identified by the determined cut-off point in the NSCLC tissues. h ROC curves of m6ARLncSig for three-year survival prediction of patients in the NSCLC tissues.

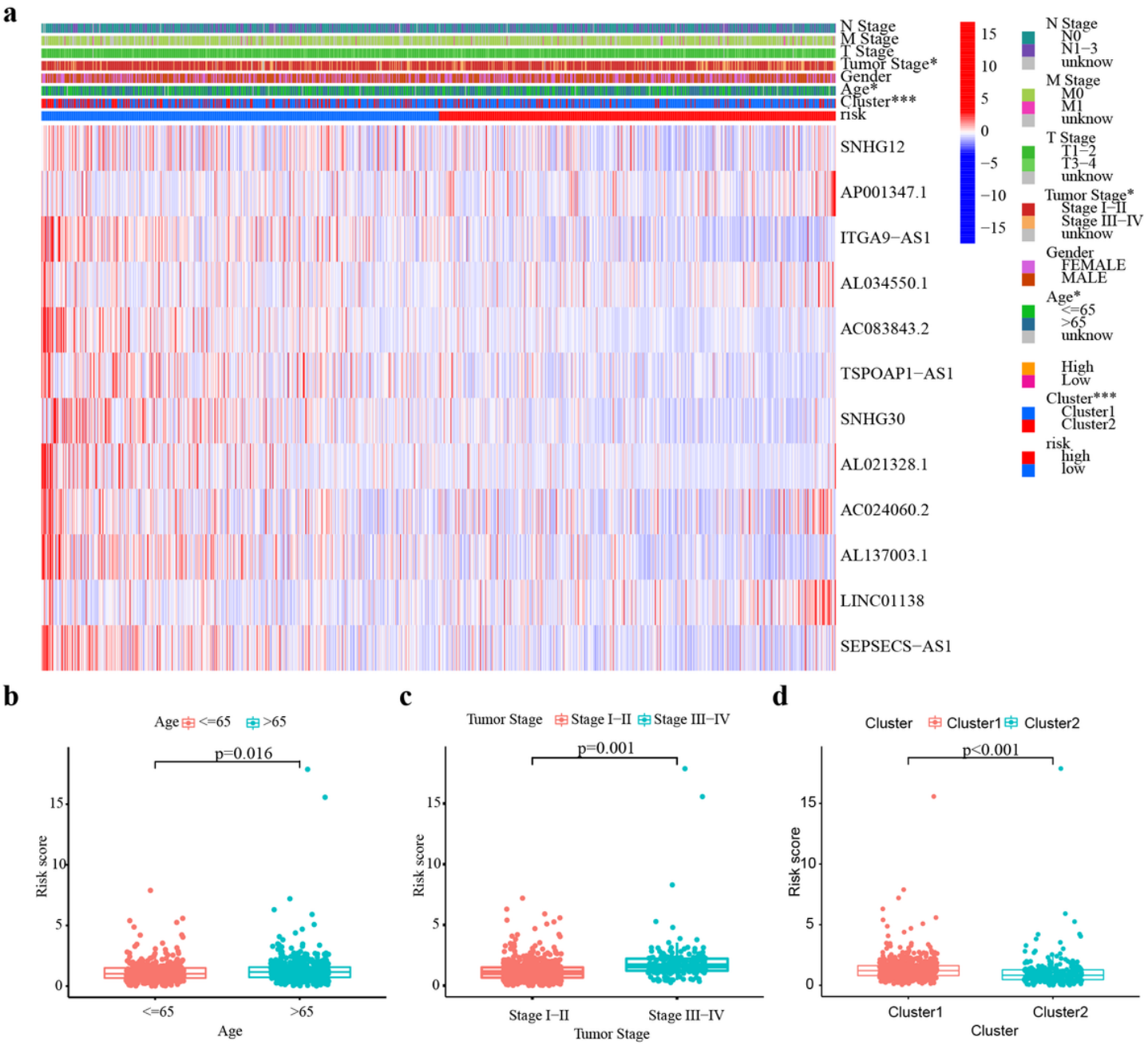

Figure 6

Analysis of the correlation between m6ARLncSig and clinical features. a The heatmap of m6ARLncRNAs of m6AlncSig along with clinical characteristics. b-d the distribution of m6ARLncRNAs in NSCLC patients in the advanced age group and young-patient group (b), patients with advanced tumor stage or early-stage (c), and patients in cluster 1 or cluster 2 (d). *p < 0.05, **p < 0.01, ***p < 0.001.

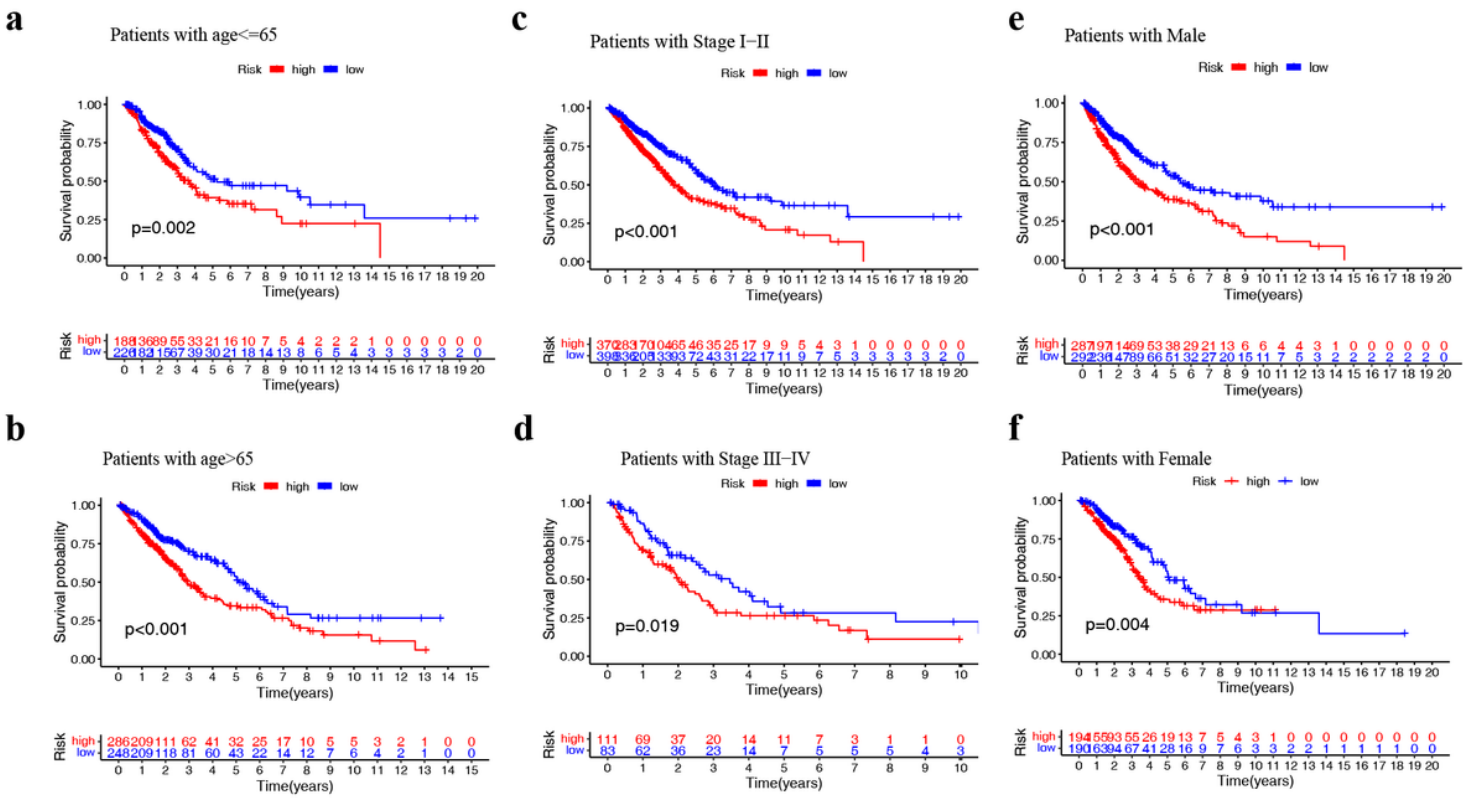

Figure 7

Analysis of the correlation between the m6ARLncSig score and a stratification analysis. a-f Kaplan–Meier survival curves of patients in multiple subgroups of patients with NSCLC including patients aged ≤65 or >65 (a-b) patients with tumor stage I-II or III-IV (c-d), and patients with male or female (e-f). *p < 0.05, **p < 0.01, ***p < 0.001.

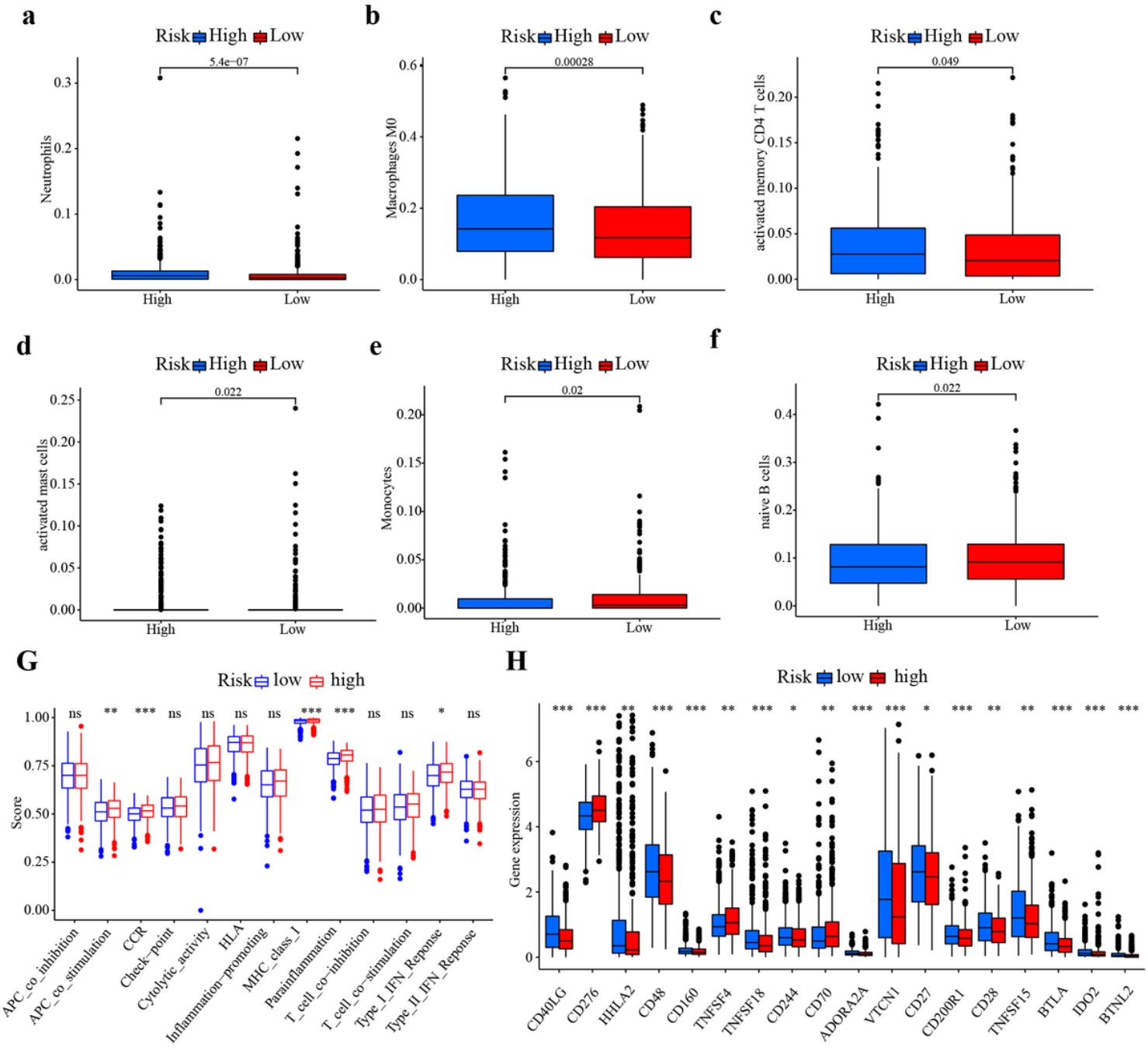

Figure 8

Relationships between the m6ARLncSig score and infiltration abundances of immune Cell Types, immunce function and immunce checkpoints. a-f Infiltration abundances of immune cell types in neutrophils (a), macrophages M0 (b), activated memory CD4 T cells (c), activated mast cells (d), Monocytes (e), and naive B cells (f). *p < 0.05, **p < 0.01, ***p < 0.001. g ssGSEA for the association between immune cell subpopulations and related functions between the two risk groups. h The differential expression of immune checkpoints among high- and low- risk m6ARLncSig groups.

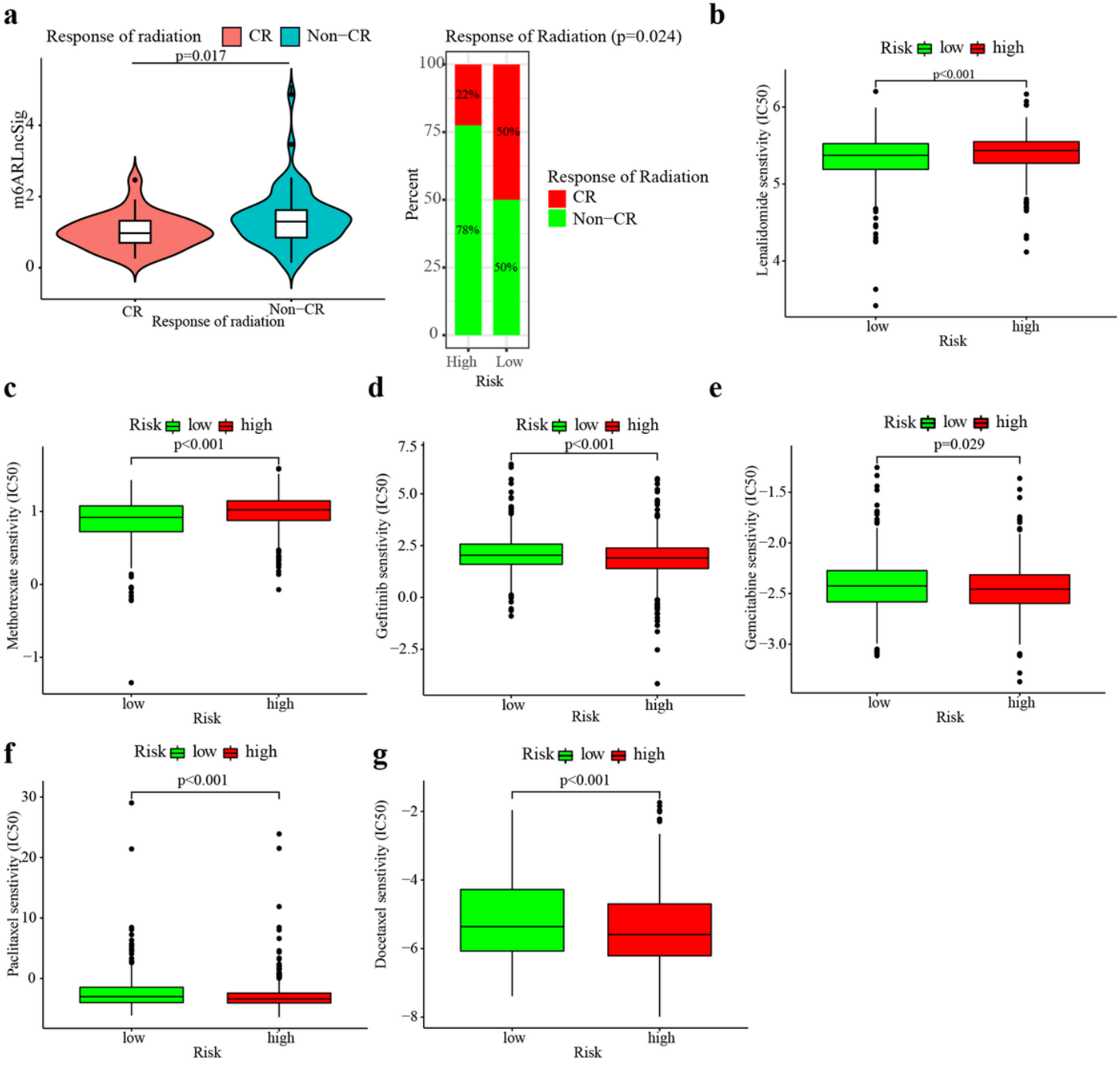

Figure 9

Relationships between m6ARLncSig score and therapeutic sensitivity. a. The boxplots of comparison of the proportion of patients with NSCLC receiving radiotherapy between the high- and low-risk groups in the training set. b-f The distribution of the estimated IC50 of lenalidomide (b), methotrexate (c), gefitinib (d), gemcitabine (e), paclitaxel (f), and docetaxel (g) in patients with NSCLC in the training dataset.

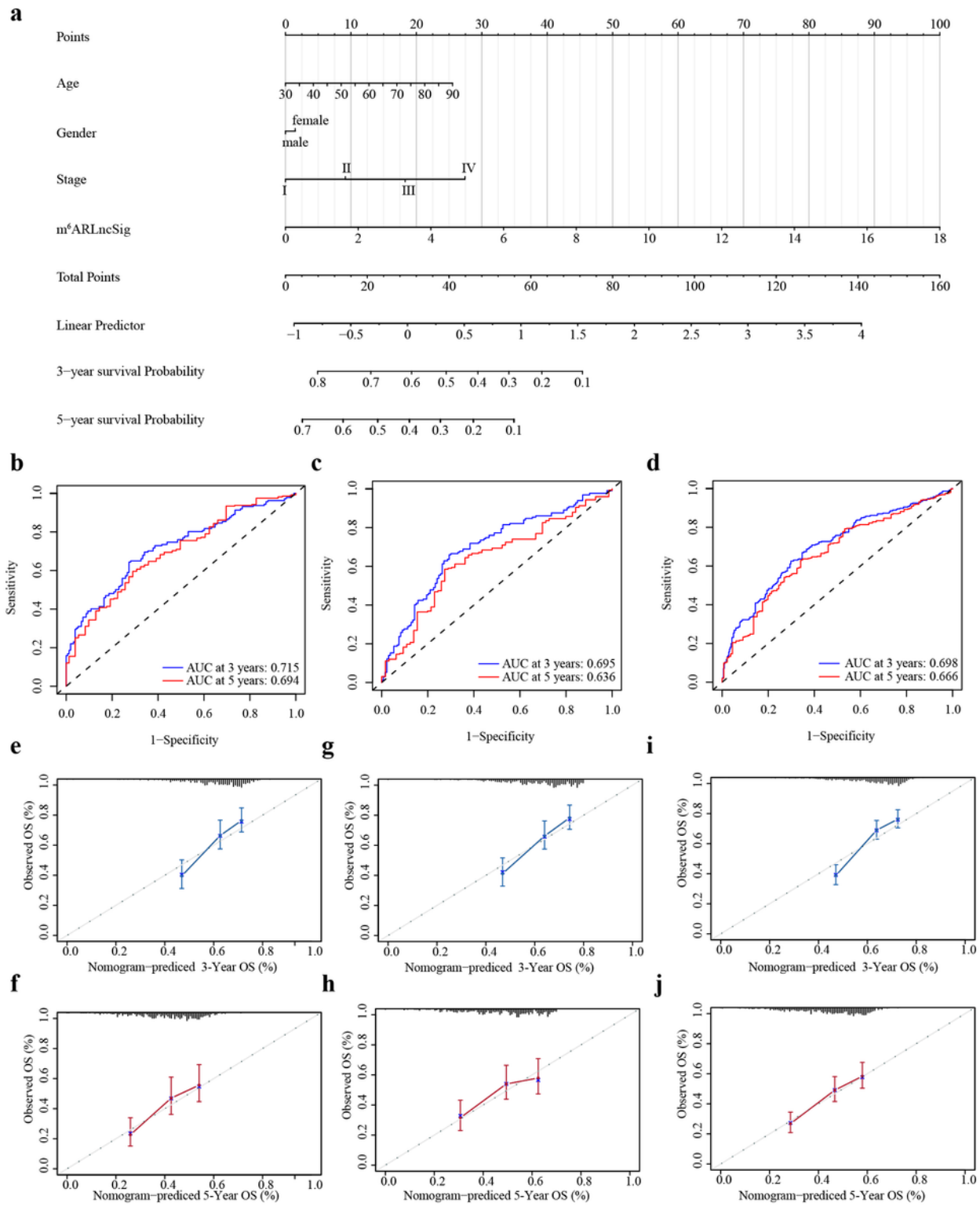

Figure 10

Construction and evaluation of a nomogram for survival prediction of NSCLC patients based on m6ARLncSig and clinical variables a The nomogram was developed in the TCGA training dataset for predicting the 3- and 5-year survival of NSCLC patients. b-d ROC curves for 3-, and 5-year survival prediction of the nomogram in the training dataset (b), the testing dataset (c), and the entire TCGA dataset (d), respectively. e-f The calibration plots of the training dataset in 3-year (e) and 5-year OS (f). g-h

The calibration plots of the testing dataset in 3-year (g) and 5-year OS (h). (I-J) The calibration plots of the whole TCGA dataset in 3-year (i) and 5-year OS (j).

Supplementary Files

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

- Additionalfile.docx