Identification and validation of a novel prognostic index based on m6A RNA methylation regulators in esophageal carcinoma

Wanli Yang
Fourth Military Medical University

Lili Duan
Fourth Military Medical University

Xinhui Zhao
Northwest University

Xiaoqian Wang
Fourth Military Medical University

Yiding Li
Fourth Military Medical University

Liaoran Niu
Fourth Military Medical University

Yujie Zhang
Fourth Military Medical University

Wei Zhou
Fourth Military Medical University

Jinqiang Liu
Fourth Military Medical University

Qingchuan Zhao
Fourth Military Medical University

Yu Han
Fourth Military Medical University

Daiming Fan
Fourth Military Medical University

Liu Hong (✉ Hongliufmmu@163.com)
Fourth Military Medical University   https://orcid.org/0000-0002-8276-2345

Primary research

Abstract

Background

The incidence and mortality rate of esophageal carcinoma (ESCA) remains high. This study proposed to explore the promising prognostic markers based on m^6A RNA methylation regulators, and finally to improve the prognostic assessment for ESCA patients.

Methods

The RNA sequencing and relevant clinical data of ESCA and normal tissues were obtained from The Cancer Genome Atlas (TCGA) database. Then, we evaluated the expression pattern of 13 m^6A methylation regulators in ESCA and normal samples. Two groups of ESCA were divided by the consensus clustering analysis. STRING database and R package were used to construct the protein-protein interaction network and conduct correlation analysis, respectively. Cox regression and least absolute shrinkage and selection operator (LASSO) regression analyses were performed to develop the multiple-gene risk signature. Kaplan-Meier method and receiver operating characteristic (ROC) curves were used to assess the accuracy of the model. The clinical nomogram combining clinicpathological factors and gene signature was built to predict survival rate of ESCA patients. Gene set enrichment analysis (GSEA) and networks prediction analysis were conducted to explore the signaling pathways that related to the risk genes.

Results

Eight m^6A methylation regulators (HNRNPC, YTHDF1, METTL3, YTHDF2, WTAP, YTHDC1, KIAA1429, and RBM15) were significantly upregulated in ESCA. Two clusters of ESCA with obvious differences in tumor stage were identified via the consensus clustering analysis. A two-gene signature, ALKBH5 and HNRNPC, was established for predicting the prognosis of ESCA patients. Kaplan-Meier curves illustrated that the overall survival of patients in low-risk group was obviously longer than that of patients in high-risk group (P = 1.411e-02). Importantly, risk score and tumor stage were identified as the independent prognostic indicators. The testing dataset GSE13898 showed that the nomogram had a good capacity to assess the prognosis of ESCA patients. Cell cycle, mTOR pathway, and p53 signaling pathway were found to be related to the dysregulation of risk genes.

Conclusions

m^6A RNA methylation regulators ALKBH5 and HNRNPC could act as prognostic indicators and therapeutic targets for prognostic analysis and cancer treatment of ESCA.

1. Background

Esophageal carcinoma (ESCA) is the seventh most commonly diagnosed cancer worldwide, with 572,000 new cases and 509,000 deaths estimated in 2018[1]. The development of ESCA is a complex process which is influenced by many factors[2]. The major risk factors of ESCA are obesity, heavy drinking, gastric reflux, and smoking[3]. Although much progress has been made in the early screen and treatment strategy of ESCA, this disease remains a profound public health burden [3]. Radical esophageal resection is the mainstay of treatment for patients with resectable ESCA. However, the overall survival (OS) of ESCA patients has remained poor over the past decades[4]. Thus, it is of great importance to identify the prognostic biomarkers for ESCA patients, especially for these with high risk and may receive benefit from the timely treatment.

RNAs play a crucial role in various cellular processer, and its emerging roles in the regulation of tumor progress have been gradually revealed during the past decade[5]. To date, RNA modifications have been identified in various RNAs, including mRNA, non-coding RNA, and others [5, 6]. So far, there are more than 150 known RNA modifications forms have been reported, including 5-methylcytosine, N6-methyladenosine (m^6A), N7-methyladenosine, etc[7]. Among these, N6-methyladenosine (m^6A) is widely found in various RNAs and is regarded as the most abundant types of mRNA modifications [8]. Generally, RNA m^6A modification can modulate RNA transcription, translocation, metabolism, and RNA splicing [8–10]. The process of RNA m^6A modification is dynamic and reversible[10], which is regulated by a methyltransferase complex called "writers" (KIAA1429, RBM15, METTL3, METTL14, ZC3H13, and WTAP), "readers" (YTHDF1, YTHDF2, YTHDC1, YTHDC2, and HNRNPC), and "erasers" (ALKBH5 and FTO). The dysregulation of m^6A RNA methylation regulators is found to be related to several diseases including obesity, autoimmune diseases, and cancers [11, 12]. For example, METTL14 exerts an oncogenic role through regulating its mRNA targets (MYB and MYC) by m^6A modification in leukemogenesis [13]. Controversy, METTL3-mediated m^6A modification plays a tumor suppressor role in colorectal cancer [14]. Thus, the m^6A methylation regulators may play different roles in different cancers. Considering the limited reports of the role of m^6A methylation in ESCA, studying on the clinical values of m^6A methylation regulators in ESCA is highly needed.

In this study, we comprehensively analyzed the expression pattern of 13 widely studied m^6A RNA methylation regulators in 160 ESCA and 11 normal tissues from the Cancer Genome Atlas (TCGA) datasets. The correlation of these m^6A regulators and its relationship with patients' clinical features were explored. Based on the result of Cox regression and LASSO regression analysis, a two-gene signature based on m^6A RNA regulators was constructed, which presented good performance in predicting the prognosis of ESCA patients. A nomogram combining the clinicalpathological factors and

risk gene signature was established to predict the individual's survival rate in ESCA. Finally, several important pathways (cell cycle, mTOR pathway, and p53 signaling pathway) were found to be associated with the dysregulation of risk genes.

2. Material And Methods

2.1 Data collection and procession

The RNA-sequencing transcriptome data and clinicopathological information of 160 ESCA patients and 11 normal samples were obtained from The Cancer Genome Atlas (https://cancergenome.nih.gov/) database. Data procession was performed as the previous studies[15, 16]. GSE13898 dataset was downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) for validation [17]. This study was conducted according to the flow chart (**Figure. 1**)

2.2 Identification of differentially expressed m^{6}A RNA methylation regulators

According to the recent published literature of m^{6}A RNA methylation regulators in human cancer [15, 16], 13 m^{6}A RNA methylation regulators (METTL3, METTL14, WTAP, KIAA1429, RBM15, ZC3H13, YTHDC1, YTHDC2, YTHDF1, YTHDF2, HNRNPC, FTO, and ALKBH5) with available expression profiles in the TCGA-ESCA dataset were included (**Table 1**). The Wilcoxon signed-rank test was applied to screen the differentially expressed m^{6}A RNA methylation regulators between ESCA and normal tissues with $P < 0.05$. Subsequently, the expression data of the m^{6}A RNA methylation regulators were combined with the corresponding clinical data. The "pheatmap" and "vioplot" packages in R software were used to visualize the expression patterns of m^{6}A regulators.

2.3 Consensus clustering analysis

The "Consensus ClusterPlus" package was used to divide the ESCA samples into different clusters[18]. Subsequently, the principal component analysis (PCA) was performed to evaluate the gene expression patterns of m^{6}A RNA methylation regulators in two ESCA clusters. Then, "survival" package was used to compare the OS of patients in two ESCA clusters. The clinicopathologic features and expression patterns of m^{6}A RNA methylation regulators in two groups were presented by "pheatmap" package.

2.4 Protein-protein interaction network construction, correlation analysis, and KEGG enrichment

STRING database (https://string-db.org/; Version: 11.0) was used to construct the protein-protein interaction (PPI) network for these 13 m^{6}A methylation regulators. Subsequently, the hub genes within the PPI network were screened via CytoHubba-plugin in Cytoscape software according to the degree of connectivity. The correlation analysis was conducted by "corrplot" package to explore the relationships among the m^{6}A regulators. KEGG analysis of the 13 m^{6}A methylation regulators was conducted via the FunRich tool (FunRich 3.0)[19].

2.5 Construction of prognostic gene signature

The cBio Cancer Genomics database (https://www.cbioportal.org/) was used to investigate the genetic alteration information of 13 m^6A methylation regulators [20]. Univariate Cox regression analysis was performed to exp lore the relationships between the expression of m^6A RNA methylation regulators and patients' survival [21]. Then, these 13 m^6A methylation regulators were entered into a LASSO regression analysis[15]. Two m^6A regulators were screened as the prognostic factors by LASSO regression analysis. Moreover, the prognostic roles of these two genes were explored using the Kaplan-Meier plotter website (http://kmplot.com/analysis/index.php), a free available tool based on GEO database, European Genome-phenome Archive (EGA) database, and TCGA database[22]. These two m^6A regulators were then used to develop a potential risk signature as previous study[23]. Each patient with ESCA was assigned a risk score according to the formula. The predictive formula was calculated as following: risk score = (Coefficient $_{gene1}$ × Expression $_{gene1}$) + (Coefficient $_{gene2}$ × Expression $_{gene2}$) [24]. The Kaplan-Meier curve and receiver operating characteristic (ROC) curve were applied to assess the predictive efficiency of the gene signature.

2.6 Identification of the independent prognostic indicators in ESCA

Univariate and multivariate Cox regression analyses were performed to identify the independent prognostic factors for ESCA patients (including age, gender, stage, T (primary tumor), M (metastasis), N (lymph nodes), and risk score). In addition, the Kaplan-Meier curves were conducted to compare the OS difference between the low-risk and high-risk groups stratified by gender age, and stage.

2.7 Construction and validation of the clinical nomogram

A nomogram based on the independent prognostic factors (stage and risk score) was develop for predicting the 1-year and 3-year survival rate of ESCA patients [25]. The calibration curves were used to evaluate the performance of the nomogram. GSE13898 ESCA dataset was downloaded and used to validate the accuracy of the nomogram.

2.8 Gene set enrichment analysis (GSEA)

160 ESCA patients were divided into high expression and low expression group according to the median expression value of ALKBH5 or HNRNPC. GSEA (http://software.broadinstitute.org/gsea/index.jsp) was performed to analyze HNRNPC or ALKBH5 gene related biological pathways and biological processes [26]. The normalized enrichment score (NES) and nominal p-value were applied to sort the pathways and processes enriched in each phenotype [27].

2.9 Establishment of transcription factors (TFs)-genes networks and miRNA-genes networks

The NetworkAnalysis database (http://www.networkanalyst.ca) was applied to build the TFs-genes networks and miRNA-genes networks for HNRNPC and ALKBH5 [28]. The ENCODE database with ChIP-

seq data in the NetworkAnalysis platform was used to predict the potential TFs [28]. The miRNA-gene networks were predicted by the TarBase and miRTarBase in the NetworkAnalysis platform.

2.10 Statistical analysis

The Perl language and R statistical package (R version 3.6.3) were used to conduct all the statistical tests and graphics unless otherwise stated. $P < 0.05$ was considered as statistical significance.

3. Results

3.1 Expression patterns of m^{6}A RNA methylated regulators in ESCA

The m^{6}A RNA methylation regulators play a crucial role in the cancer development and progression. Thus, we firstly explored the expression patterns of these 13 genes in 160 ESCA tissues and 11 normal tissues in TCGA dataset. As shown in **Figure. 2a** and **Figure. 2b**, HNRNPC, YTHDF1, METTL3, YTHDF2, WTAP, YTHDC1, KIAA1429, and RBM15 were upregulated in ESCA tissues as compared with normal samples. According to **Figure. 2c**, some of the m^{6}A RNA methylation regulators were positively correlated with others. For example, YTHDC1 gene and METTL14 gene are most relevant. While YTHDC1 is overexpressed, METTL14 is most likely to be overexpressed (**Figure. 2c**). To explore the interaction of the 13 m^{6}A RNA methylation regulators, STRING database (https://string-db.org/; Version: 11.0) was applied to construct the PPI network of the these genes (**Figure. 2d**). The 13 m^{6}A RNA methylation regulators presented complicated interactions among each other. Moreover, 5 genes (METTL14, RBM15, METTL3, WTAP, and KIAA1429) were considered as the hub genes according to the degree connectivity (**Figure. S1**). To better understand the biological functions of m^{6}A RNA methylation regulators in ESCA, KEGG analysis were performed for these 13 genes, as shown in **Figure. S2**, KEGG analysis showed that these genes were mainly enriched in processing of capped intron-containing pre-mRNA, mRNA processing, and formation and maturation of mRNA transcript, gene expression, and regulation of telomerase, etc.

3.2 Clustering of ESCA patients and prognosis of two clusters

According to the expression similarity of m^{6}A RNA methylation regulators, "Consensus ClusterPlus" package in R software was applied to clustered the ESCA cases into different groups. As a result, k = 2 was selected as the most appropriate index, which could cluster the ESCA patients into two cohorts (cluster 1 and cluster 2) with ideal stability (**Figure. 3**). Moreover, the principal component analysis (PCA) demonstrated that cluster 1 and cluster 2 had an obvious distinction of the transcriptional profile (**Figure. S3**). Then, Kaplan-Meier method was used to compare the OS of ESCA patients between cluster 1 and cluster 2. **Figure. S4a** showed that the ESCA patients in cluster 1 presented a trend with shorter OS compared to cluster 2 although it was not statistically significant (P = 0.333). However, the relationship analysis of the clinicopathological features showed obvious difference for the stage factor between the

two clusters (P < 0.01) (**Figure. S4b**). Taking together, the above results indicated that the clustering result was associated with the clinicopathological factors and tumorigenesis of ESCA.

3.3 Construction of the risk gene signature for ESCA

To investigate the contributions of the 13 m^6A RNA methylation regulators to ESCA, the genetic alteration information of these regulators were explored in the cBio Cancer Genomics database. The Nature 2017 ESCA dataset (265 cases) and Firehose Legacy ESCA dataset (184 cases) of ESCA were both included. The 13 m^6A RNA methylation regulators were altered in 97 (37%) of 265 queried patients (TCGA ESCA, Nature 2017 dataset) (**Figure. S5a**), compared with that altered queried genes were detected in 62 (34%) of 184 queried patients (TCGA ESCA, Firehose Legacy dataset) (**Figure. S5b**). These findings indicated that the 13 m^6A RNA methylation regulators play a crucial role in esophageal carcinogenesis.

To explore the prognostic value of m^6A RNA methylation regulators in ESCA, we conducted the univariate Cox regression analysis to screen the regulators related to OS of ESCA patients. The results showed that ALKBH5 was markedly associated with OS with HR = 0.956 (P = 0.013). In addition, HNRNPC was weekly associated with the OS of ESCA patients (P = 0.168) (**Figure. 4a**). We then explored the prognostic roles of these two genes in Kaplan-Meier plotter website (http://kmplot.com/analysis/index.php), the results showed that overexpression of ALKBH5 might lead to a better prognosis in ESCA, while high level of HNRNPC could predict poor outcome in ESCA patients (**Figure. S6**).

In order to get a better understanding of the relationships between the m^6A RNA methylation regulators expression and prognosis of ESCA patients, LASSO regression analysis was conducted for these 13 genes. **Figure. 4b** presented the regression coefficient of these 13 genes in ESCA. As shown in **Figure. 4c**, when the two genes (ALKBH5 and HNRNPC) were included, the model achieved the best performance. Finally, a stepwise multivariate Cox regression was performed to construct the optimal risk signature based on these two genes. Each ESCA patient received a risk score calculated as follows: risk score = (-0.027844 × expression value of ALKBH5) + (0.008465 × expression value of HNRNPC). Then, the ESCA patients were divided into low-risk and high-risk group according to the median value of risk score. The Kaplan-Meier curves showed that ESCA patients with high risk score had a poorer survival than those with low risk score (P = 1.411e-02) (**Figure. 4d**). The ROC curve showed that the risk score curve had a good feasibility in predicting the individuals' survival rate with AUC of 0.644 (**Figure. 4e**). **Figure. 4f** illustrated that the survival time of ESCA patients decreases along with the rising of risk score. Therefore, the above results suggesting that this risk gene signature could effectively identify the high risk ESCA patients with poor OS.

3.4 The risk score and stage were independent prognostic indicators in ESCA

Univariate and multivariate Cox regression analyses were performed to identify the independent prognostic indicators in ESCA. As shown in **Figure. 5a**, gene signature-based risk score was closely related to the worse OS in EC (HR = 55.941, 95% CI:7.540-415.025, P < 0.001). In addition, stage (HR =

2.625, 95% CI:1.778–3.878, P < 0.001) and N (lymph nodes) (HR = 55.941, 95% CI:1.077–1.888, P = 0.013) were also demonstrated to be associated with the OS. Then, all factors were entered into the multivariate Cox analysis. Importantly, gene signature based risk score (HR = 52.021, 95% CI:6.110-442.934, P < 0.001) and stage (HR = 2.608, 95% CI:1.598–4.255, P < 0.001) were still identified as the independent prognostic factors for poorer OS in ESCA patients (**Figure. 5b**).

Finally, we also performed subgroup analysis to assess the prognostic role of the two-gene risk signature in ESCA patients according to several clinicopathological factors (grade, stage, and age). ESCA patients with high risk score in subgroup of female, stage III-IV, and patients with age > 65 had dramatically lower OS than patients with low risk score (**Figure. 6**) (P < 0.05), suggesting that the two-gene based signature have good ability to discriminate the patients with poor OS. Other subgroups also presented similar trends but not statistically significant.

3.5 Construction and validation of a clinical nomogram

The nomogram was used to quantitatively evaluate the patients' survival rate through integrating the independent prognostic factors (stage and risk score). The total points of risk factors were used to assess the patients' 1-year and 3-year survival rates (**Figure. 7a**). The concordance index (C-index) was 0.70 (95% CI: 0.62–0.79). Moreover, calibration curves showed good concordance between the nomogram-predicted survival and actual survival (**Figure. 7b-7c**), especially for the 3-year survival. Importantly, we validated the nomogram in the GSE13898 ESCA dataset, and the 1-year and 3-year calibration curves was also presented good concordance between the nomogram-predicted survival and actual survival (**Figure. 7d-7e**). These result suggested that the clinical nomogram performs well in predicting the prognosis of ESCA patients.

3.6 Identification of ALKBH5 and HNRNPC related signaling pathway and biological processes via GSEA

To explore the signaling pathways and biological processes that are differentially activated in the ESCA development and progression, GSEA was conducted between high and low ALKBH5 or HNRNPC expression datasets respectively. According to the normalized enrichment score (NES) and NOM p-value < 0.05, the most significantly enriched signaling pathways and biological processes were selected and presented in **Figure. 8** and **Figure. S6**, respectively. **Figure. 8a-8c** showed that cell cycle, DNA replication, and mTOR signaling pathway were differentially enriched in ALKBH5 high expression phenotype, while cell cycle, pentose phosphate pathway, and p53 signaling pathway were activated in HNRNPC high expression phenotype. As for the significant terms of biological processes, DNA biosynthetic process, DNA recombination, and DNA repair were associated with ALKBH5 overexpression (**Figure. 6Sa**), whereas phospholipase C activity and regulation of cellular extravasation were mainly related to low expression phenotype of HNRNPC (**Figure. 6Sb**).

3.7 Construction of TFs-genes networks and miRNA-genes networks for ALKBH5 and HNRNPC

To explore the contributions of ALKBH5 and HNRNPC to the development of ESCA. The TFs-genes networks and miRNA-genes networks of ALKBH5 and HNRNPC were constructed (Fig. 9a-9b). The numbers of TFs and miRNAs in the networks were 60 and 143, respectively. In the TFs-genes networks (Fig. 9a), ALKBH5 and HNRNPC shared 6 TFs (ZNF580, MBD11, GATA4, SP7, HDGF, and KLF9). In the miRNA-genes networks (Fig. 9b), miR-4747-5p, miR-4698, miR-484, miR-5196-5p, miR-8063, miR-4773, miR-615-3p and miR-4531 could regulate both ALKBH5 and HNRNPC. Taking together, the TFs-genes networks and miRNA-genes may provide novel clues for the in-depth studies of ESCA.

4. Discussion

ESCA is one of the major causes of cancer-related mortality worldwide. So far, there are no effective treatment strategies for advanced ESCA patients, leading to a poor 5-year survival rate. To optimize and personalize the treatments for ESCA patients, an accurate prognostic judgment is important during the cancer management. So far, the tumor node metastasis (TNM) classification is the most commonly utilized tool for predicting the prognosis in cancer patients [29]. However, increasing evidences demonstrated the unsatisfactory discriminative capacity of TNM system in predicting clinical outcomes [30, 31]. With the development of life science and technology, more and more prognostic markers have been identified for ESCA [3]. However, the existing markers may lack sufficient sensitivity and specificity in cancer prognosis. The gene signature based prognostic index model holds promising in prognostic prediction, thus may help to improve the personalized medicine.

To the best of our knowledge, this study firstly investigated the prognostic roles of reported m^6A RNA methylation regulators in ESCA. Firstly, 8 out of 13 m^6A RNA regulators were found to be dysregulated in ESCA, implying its crucial roles in the development of ESCA. Moreover, two clusters of ESCA presented significant difference in tumor stage, which also indicated that the expression pattern of these regulators were associated with the malignancy of ESCA. Previously, yang et al reported that genetic variants of m^6A modification genes could affect the ESCA susceptibility [32]. In this study, we also found that genetic alteration of these gene was common in ESCA according to the cBio Cancer Genomics database. These results, taking together, suggested that m^6A RNA methylation play a significant role in modulating the malignant process of ESCA. Subsequently, we conducted the Cox regression and LASSO regression analyses to build a gene signature with two m^6A RNA regulators, ALKBH5 and HNRNPC, which divided the ESCA patients into two groups (high-risk and low-risk groups). The results showed that the two-gene based signature could effectively discriminate the patients with different OS (P = 1.411e-02), and presented a great performance in prognosis prediction. What's more, the univariate and multivariate Cox regression analyses further proved that the risk score calculated by the gene signature was an independent prognostic indicator with the highest HR value than other factors. Subgroup analysis also showed that this two-gene signature have good ability to distinguish the patients with poor OS in group of female, stage III-IV, and patients with age > 65. Other subgroups presented similar trends although with no significant difference. One reasonable explanation for it may due to the limited number of ESCA patients in TCGA database. The clinical nomogram based on the independent prognostic factors was

established and validated to have a good ability in predicting the 1-year and 3-year survival rate of ESCA patients.

In the two-gene signature, ALKBH5 was proved to be a protective gene in ESCA, while high level of HNRNPC was associated with poor OS. The progression of esophageal carcinomas is influenced by the complex gene networks, our predicted TFs-genes networks and miRNA-genes networks of ALKBH5 and HNRNPC may provide powerful bases for in-depth studies regarding the molecular mechanisms of ESCA. Currently, the biological functions of ALKBH5 and HNRNPC have been investigated in various researches. Previously, He et al. demonstrated that ALKBH5 can suppress pancreatic cancer by demethylating lncRNA KCNK15-AS1[33]. Moreover, ALKBH5 can inhibit pancreatic cancer through regulating WIF-1 RNA methylation and Wnt signaling pathway[34]. On the contrary, ALKBH5 was also reported to be an candidate oncogene in other cancers, including ovarian cancer[35], breast cancer[36], and gastric cancer [37]. Given that the opposite role of ALKBH5 in different cancers, more experimental studies are highly demand. In our study, several ALKBH5-related signaling pathways were screened by GESA, including cell cycle, DNA replication, and mTOR signaling pathway. Recently, a study reported that ALKBH5 can activate EGFR-PIK3CA-AKT-mTOR signaling pathway in ovarian cancer [35]. These findings suggested that ALKBH5-mTOR pathway may be an available therapeutic pathway for cancer treatment. HNRNPC mainly function as an oncogenic role in cancers[38, 39]. Previously, studies found that the abnormally expressed HNRNPC was related to the LBX2-AS1 pathway[40] and JAK-STAT1 signaling pathway[39] in several cancers. But the association between the HNRNPC expression and cell cycle related pathway, pentose phosphate pathway, and p53 signaling pathway was firstly reported here, thus the regulatory mechanisms required to be elucidated in future.

There are also some limitations in this study. Firstly, the number of normal tissues (11) is much less than the number of ESCA tissues (160). Secondly, we did not divide the ESCA samples into two subtypes (adenocarcinoma and squamous cell carcinoma) due to the limited number of patients in TCGA dataset. These may influence the reliability of our findings. Thirdly, the prognostic index should be further validated in large clinical cohort. Finally, further studies on the two m^6A RNA regulators may improve the targeted therapy in ESCA patients.

5. Conclusions

In conclusion, the dysregulated m^6A RNA methylation regulators were closely related to clinicopathological factors of ESCA. The m^6A RNA methylation regulators, especially ALKBH5 and HNRNPC, play an essential role in the cancigenesis and progression of ESCA. Importantly, the two-gene based signature was demonstrated to be a promising tool to distinguish ESCA patients with different clinical outcome. The GESA and networks prediction analysis also provided new clues for mechanism research in ESCA.

Abbreviations

ESCA: esophageal carcinoma; LASSO: Least absolute shrinkage and selection operator; OS: Overall survival; TCGA: The Cancer Genome Atlas; KEGG: Kyoto Encyclopedia of Genes and Genomes; PPI: Protein-protein interaction; ROC: Receiver operating characteristic; AUC: The area under the ROC curve; GSEA: Gene set enrichment analysis

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data used in the study were downloaded from The Cancer Genome Atlas (TCGA) database and GEO database.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

Authors' contributions

WY, LD, XZ and XW wrote the manuscript; YH, DF, and LH conceived and designed the study; YL, LN, and YZ collected the literature; QZ, YH, DF, and LH reviewed and revised the manuscript; WZ and JL conducted the statistical analysis. All authors approved the final version of manuscript.

Acknowledgements

Not applicable.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.

2. Domper Arnal MJ, Ferrández Arenas Á, Lanas Arbeloa Á. Esophageal cancer: Risk factors, screening and endoscopic treatment in Western and Eastern countries. World J Gastroenterol. 2015;21(26):7933–43.

3. Yang W, Han Y, Zhao X, Duan L, Zhou W, Wang X, Shi G, Che Y, Zhang Y, Liu J, et al. Advances in prognostic biomarkers for esophageal cancer. Expert Rev Mol Diagn. 2019;19(2):109–19.

4. Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, Bonaventure A, Valkov M, Johnson CJ, Estève J, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37†513†025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet. 2018;391(10125):1023–75.

5. Thapar R, Bacolla A, Oyeniran C, Brickner JR, Chinnam NB, Mosammaparast N, Tainer JA. **RNA Modifications: Reversal Mechanisms and Cancer**. 2019, 58(5):312–329.

6. Roundtree IA, Evans ME, Pan T, He C. Dynamic RNA Modifications in Gene Expression Regulation. Cell. 2017;169(7):1187–200.

7. Ontiveros RJ, Stoute J, Liu KF. **The chemical diversity of RNA modifications**. 2019, 476(8):1227–1245.

8. Hong K. Emerging function of N6-methyladenosine in cancer. Oncol Lett. 2018;16(5):5519–24.

9. He L, Li H, Wu A, Peng Y, Shu G, Yin G. **Functions of N6-methyladenosine and its role in cancer**. 2019, 18(1):176.

10. Yue Y, Liu J, He C. RNA N6-methyladenosine methylation in post-transcriptional gene expression regulation. Genes Dev. 2015;29(13):1343–55.

11. Cao G, Li HB. **Recent advances in dynamic m6A RNA modification**. 2016, 6(4):160003.

12. Tong J, Flavell RA, Li HB. RNA m(6)A modification and its function in diseases. Frontiers of medicine. 2018;12(4):481–9.

13. Weng H, Huang H, Wu H, Qin X, Zhao BS, Dong L, Shi H, Skibbe J, Shen C, Hu C, et al. METTL14 Inhibits Hematopoietic Stem/Progenitor Differentiation and Promotes Leukemogenesis via mRNA m(6)A Modification. Cell stem cell. 2018;22(2):191–205.e199.

14. Deng R, Cheng Y, Ye S, Zhang J, Huang R, Li P, Liu H, Deng Q, Wu X, Lan P, et al. m(6)A methyltransferase METTL3 suppresses colorectal cancer proliferation and migration through p38/ERK pathways. Onco Targets Ther. 2019;12:4391–402.

15. Liu T, Li C, Jin L, Li C, Wang L. The Prognostic Value of m6A RNA Methylation Regulators in Colon Adenocarcinoma. Medical science monitor: international medical journal of experimental clinical research. 2019;25:9435–45.

16. Su Y, Huang J, Hu J. m(6)A RNA Methylation Regulators Contribute to Malignant Progression and Have Clinical Prognostic Impact in Gastric Cancer. Front Oncol. 2019;9:1038.

17. Kim SM, Park YY, Park ES, Cho JY, Izzo JG, Zhang D, Kim SB, Lee JH, Bhutani MS, Swisher SG, et al. Prognostic biomarkers for esophageal adenocarcinoma identified by analysis of tumor transcriptome. PloS one. 2010;5(11):e15074.

18. Chen J, Yu K, Zhong G, Shen W. **Identification of a m(6)A RNA methylation regulators-based signature for predicting the prognosis of clear cell renal carcinoma**. 2020, 20:157.

19. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, Mouradov D, Sieber OM, Simpson RJ, Salim A, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics. 2015;15(15):2597–601.

20. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–4.

21. Zhao X, Liu J, Liu S, Yang F, Chen E. **Construction and Validation of an Immune-Related Prognostic Model Based on TP53 Status in Colorectal Cancer**. 2019, 11(11).

22. Nagy Á, Lánczky A, Menyhárt O, Győrffy B. **Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets**. 2018, 8(1):9227.

23. Zhou J, Wang J, Hong B, Ma K, Xie H, Li L, Zhang K, Zhou B, Cai L, Gong K. Gene signatures and prognostic values of m6A regulators in clear cell renal cell carcinoma - a retrospective study using TCGA database. Aging. 2019;11(6):1633–47.

24. Wang Y, Zhao W, Xiao Z, Guan G, Liu X, Zhuang M. **A risk signature with four autophagy-related genes for predicting survival of glioblastoma multiforme**. 2020, 24(7):3807–3821.

25. Liu Y, Wu L, Ao H, Zhao M, Leng X, Liu M, Ma J, Zhu J. Prognostic implications of autophagy-associated gene signatures in non-small cell lung cancer. Aging. 2019;11(23):11440–62.

26. Luo Y, Chen L, Wang G, Qian G, Liu X, Xiao Y, Wang X, Qian K. PPARα gene is a diagnostic and prognostic biomarker in clear cell renal cell carcinoma by integrated bioinformatics analysis. J Cancer. 2019;10(10):2319–31.

27. Wu H, Zhang J. Decreased expression of TFAP2B in endometrial cancer predicts poor prognosis: A study based on TCGA data. Gynecol Oncol. 2018;149(3):592–7.

28. Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. Nature protocols. 2015;10(6):823–44.

29. Birla R, Gandea C, Hoara P, Caragui A, Marica C, Vasiliu E, Constantinoiu S: **Clinical and Therapeutic Implications of the 8th Edition TNM Classification of Adenocarcinomas of the Esophagogastric Junction**. *Chirurgia (Bucharest, Romania*: 1990) 2018, **113**(6):747–757.

30. Zheng Y, Fu S, He T, Yan Q, Di W, Wang J. Predicting prognosis in resected esophageal squamous cell carcinoma using a clinical nomogram and recursive partitioning analysis. Eur J Surg Oncol. 2018;44(8):1199–204.

31. Du F, Sun Z, Jia J, Yang Y, Yu J, Shi Y, Jia B, Zhao J, Zhang X. Development and Validation of an Individualized Nomogram for Predicting Survival in Patients with Esophageal Carcinoma after Resection. J Cancer. 2020;11(14):4023–9.

32. Yang N, Ying P, Tian J, Wang X, Mei S, Zou D, Peng X, Gong Y, Yang Y, Zhu Y, et al: **Genetic variants in m6A modification genes are associated with esophageal squamous-cell carcinoma in the Chinese population**. *Carcinogenesis* 2020.

33. He Y, Hu H, Wang Y, Yuan H, Lu Z, Wu P, Liu D, Tian L, Yin J, Jiang K, et al. ALKBH5 Inhibits Pancreatic Cancer Motility by Decreasing Long Non-Coding RNA KCNK15-AS1 Methylation. Cell Physiol Biochem. 2018;48(2):838–46.

34. Tang B, Yang Y, Kang M, Wang Y, Wang Y, Bi Y, He S, Shimamoto F. m(6)A demethylase ALKBH5 inhibits pancreatic cancer tumorigenesis by decreasing WIF-1 RNA methylation and mediating Wnt signaling. Mol Cancer. 2020;19(1):3.

35. Zhu H, Gan X, Jiang X, Diao S, Wu H, Hu J. ALKBH5 inhibited autophagy of epithelial ovarian cancer through miR-7 and BCL-2. J Exp Clin Cancer Res. 2019;38(1):163.

36. Zhang C, Zhi WI, Lu H, Samanta D, Chen I, Gabrielson E, Semenza GL. Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. Oncotarget. 2016;7(40):64527–42.

37. Zhang J, Guo S, Piao HY, Wang Y, Wu Y, Meng XY, Yang D, Zheng ZC, Zhao Y: **ALKBH5 promotes invasion and metastasis of gastric cancer by decreasing methylation of the lncRNA NEAT1**. *Journal of physiology and biochemistry* 2019, **75**(3):379–389.

38. Wu Y, Zhao W, Liu Y, Tan X, Li X, Zou Q, Xiao Z, Xu H, Wang Y, Yang X: **Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response**. 2018, **37**(23).

39. Yan M, Sun L, Li J, Yu H, Lin H, Yu T, Zhao F, Zhu M, Liu L, Geng Q, et al. RNA-binding protein KHSRP promotes tumor growth and metastasis in non-small cell lung cancer. J Exp Clin Cancer Res. 2019;38(1):478.

40. Zhang Y, Chen W, Pan T, Wang H, Zhang Y, Li C. 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(3):566–72.
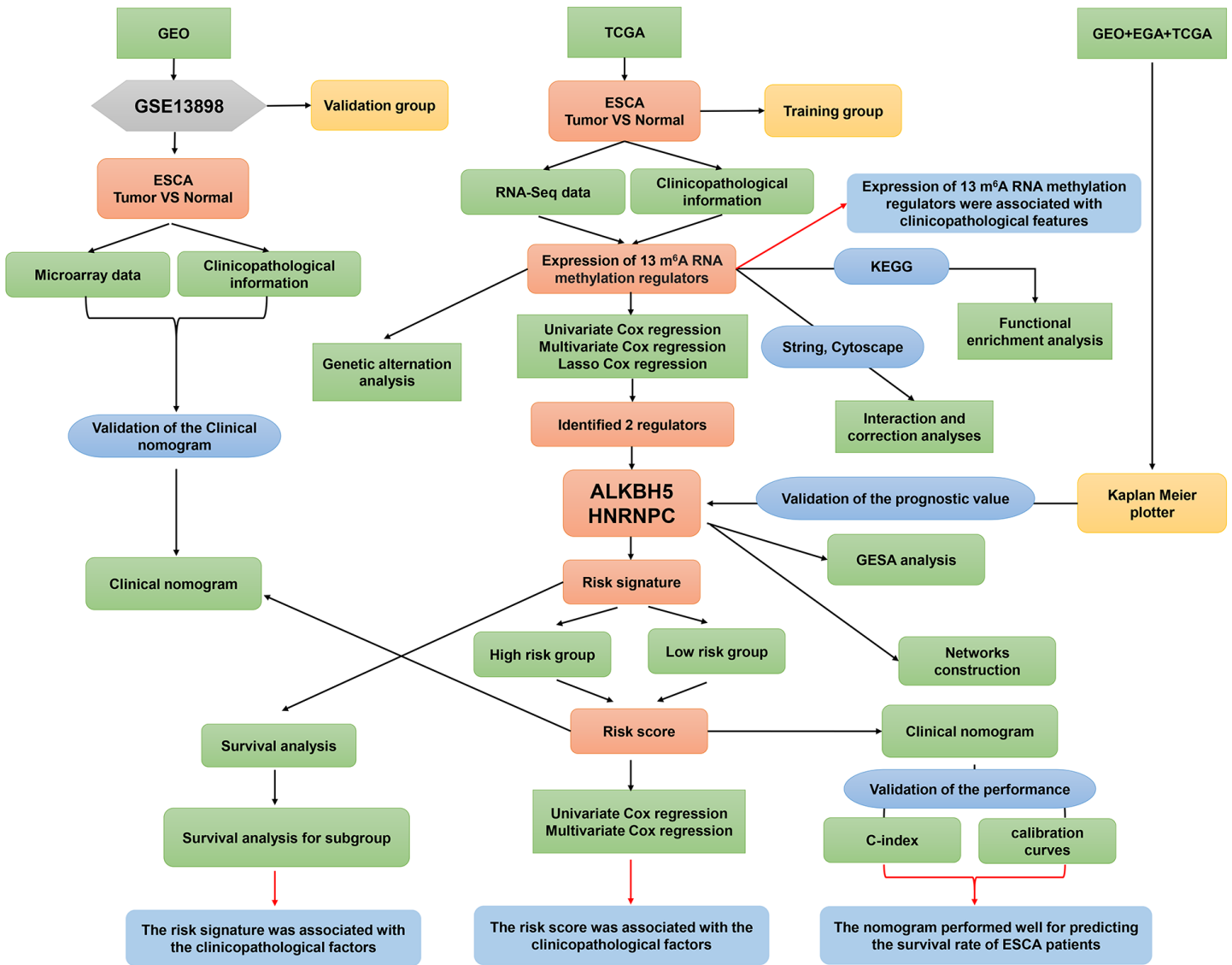
Figures

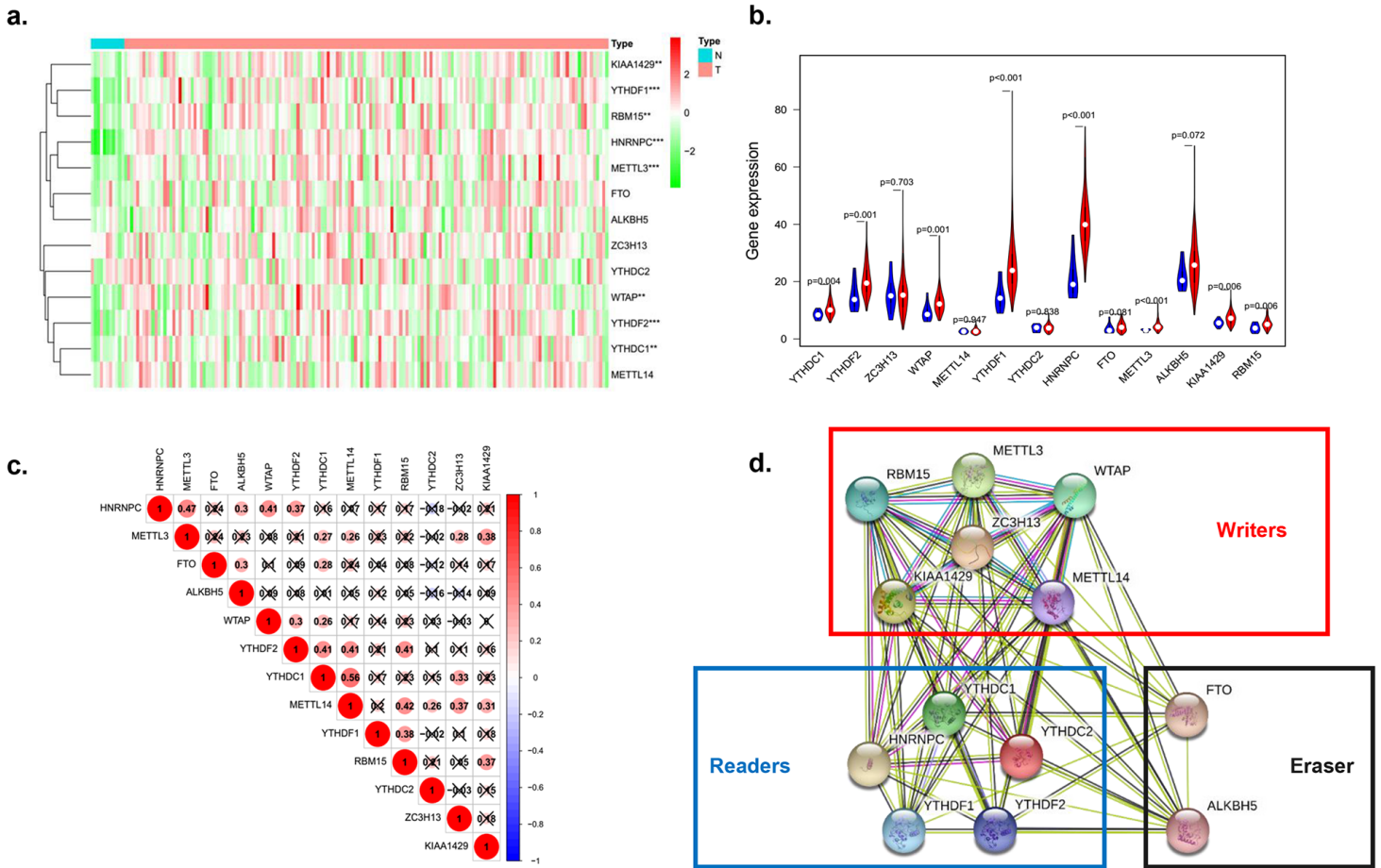**Figure 1**

The flow chart of the study analysis.

Figure 2

The expression pattern of 13 m6A RNA methylation regulators in ESCA. a. Heatmap of expression levels of 13 m6A RNA methylation regulators in ESCA tissues and normal tissues; b. Vioplot of expression levels of 13 m6A RNA methylation regulators in ESCA and normal samples; c. The Pearson correlation analysis of the 13 m6A RNA methylation regulators in ESCA; d. PPI network of the 13 m6A RNA methylation regulators. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$
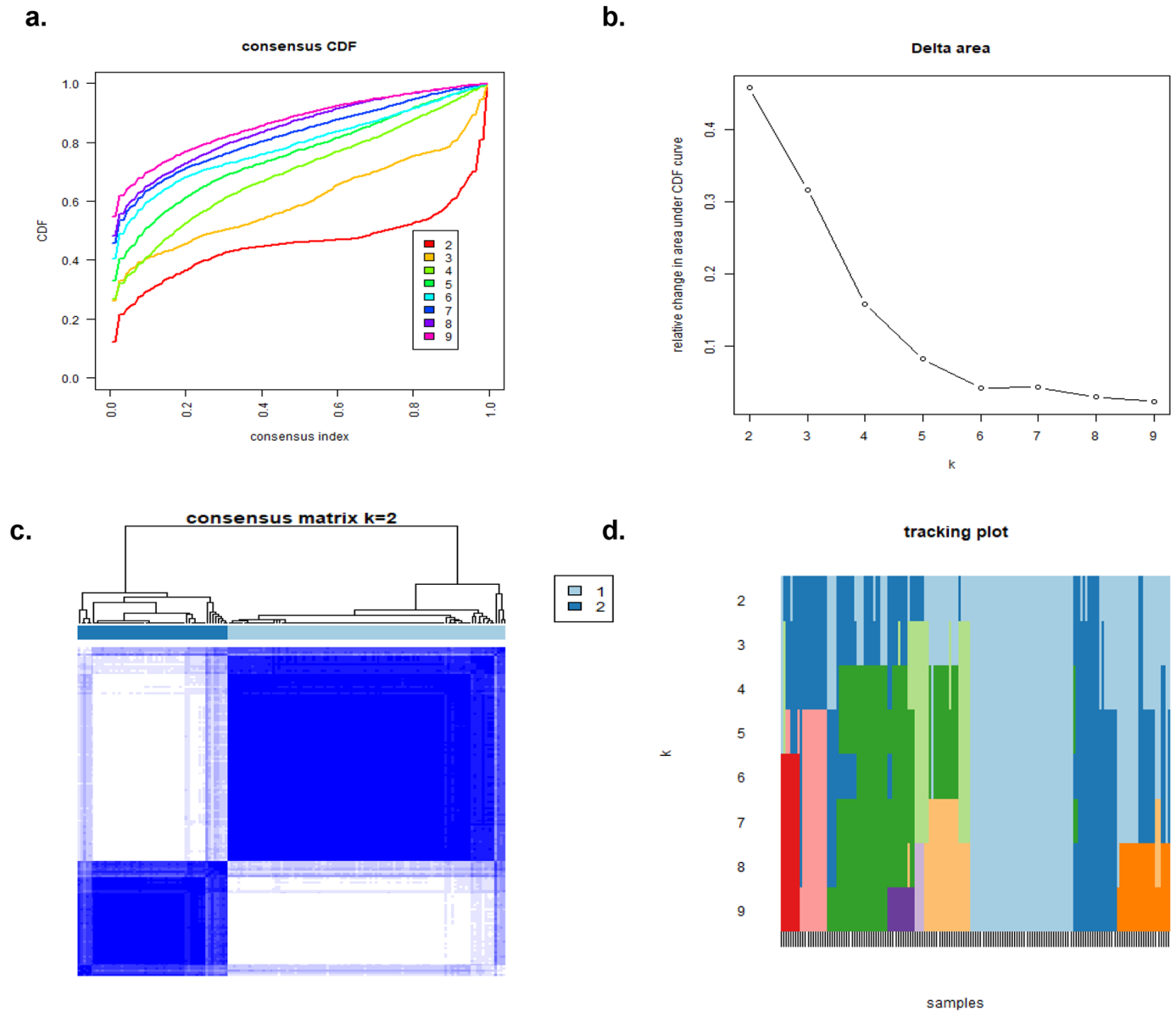
Figure 3

Consensus clustering by m6A RNA methylation regulators. a. Consensus clustering cumulative distribution function (CDF) for k=2-9; b. Relative change in area under CDF curve for k=2-9; c. Consensus clustering matrix for k=2; d. The tracking plot for k=2-10
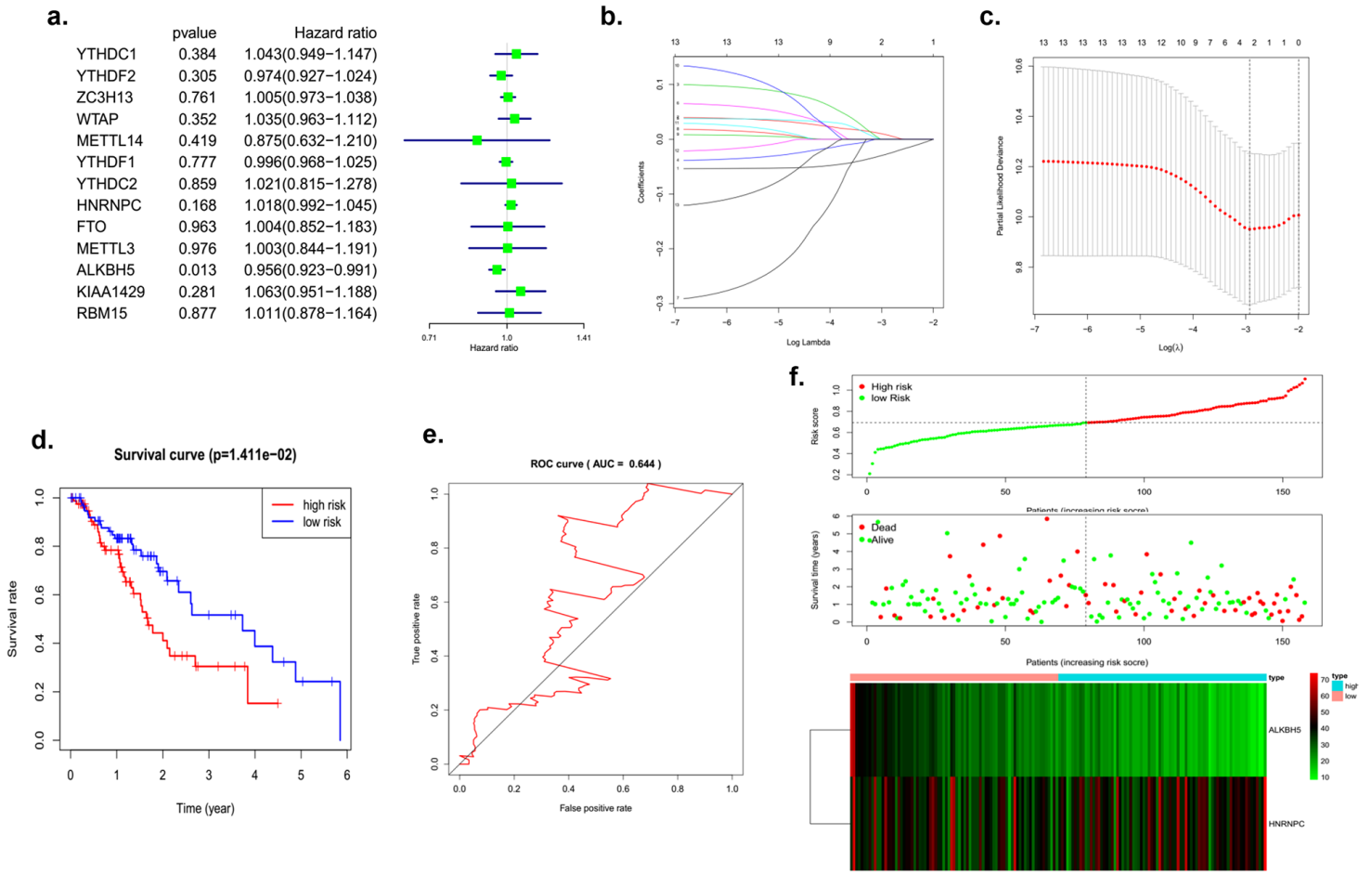
Figure 4

Construction of the prognostic index based on m6A RNA methylation regulators. a. Univariate Cox regression analysis of the 13 m6A RNA methylation regulators in ESCA; b-c. LASSO regression analysis of the 13 m6A RNA methylation regulators; d. Kaplan-Meier curves of ESCA patients in low-risk and high-risk group; e. The ROC curve for assessing the prediction efficiency of the prognostic index; f. Characteristics of prognostic gene signature in ESCA.
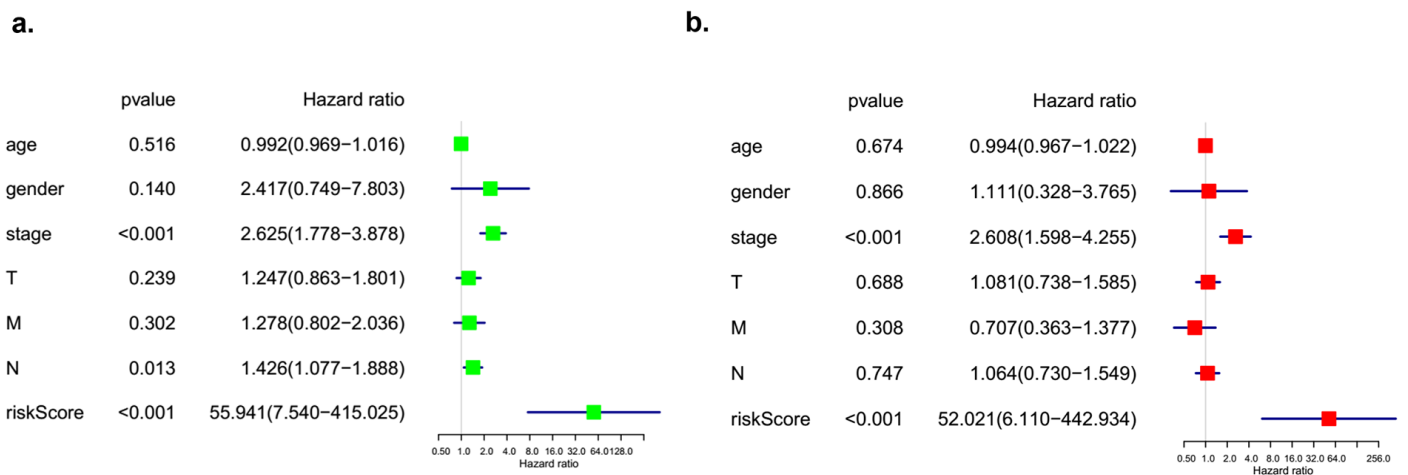
Figure 5

Identification of the independent prognostic indicators in ESCA. a. Univariate Cox regression analysis of the clinicopathological factors and risk score; b. Multivariate Cox regression analysis of the clinicopathological features and risk score.
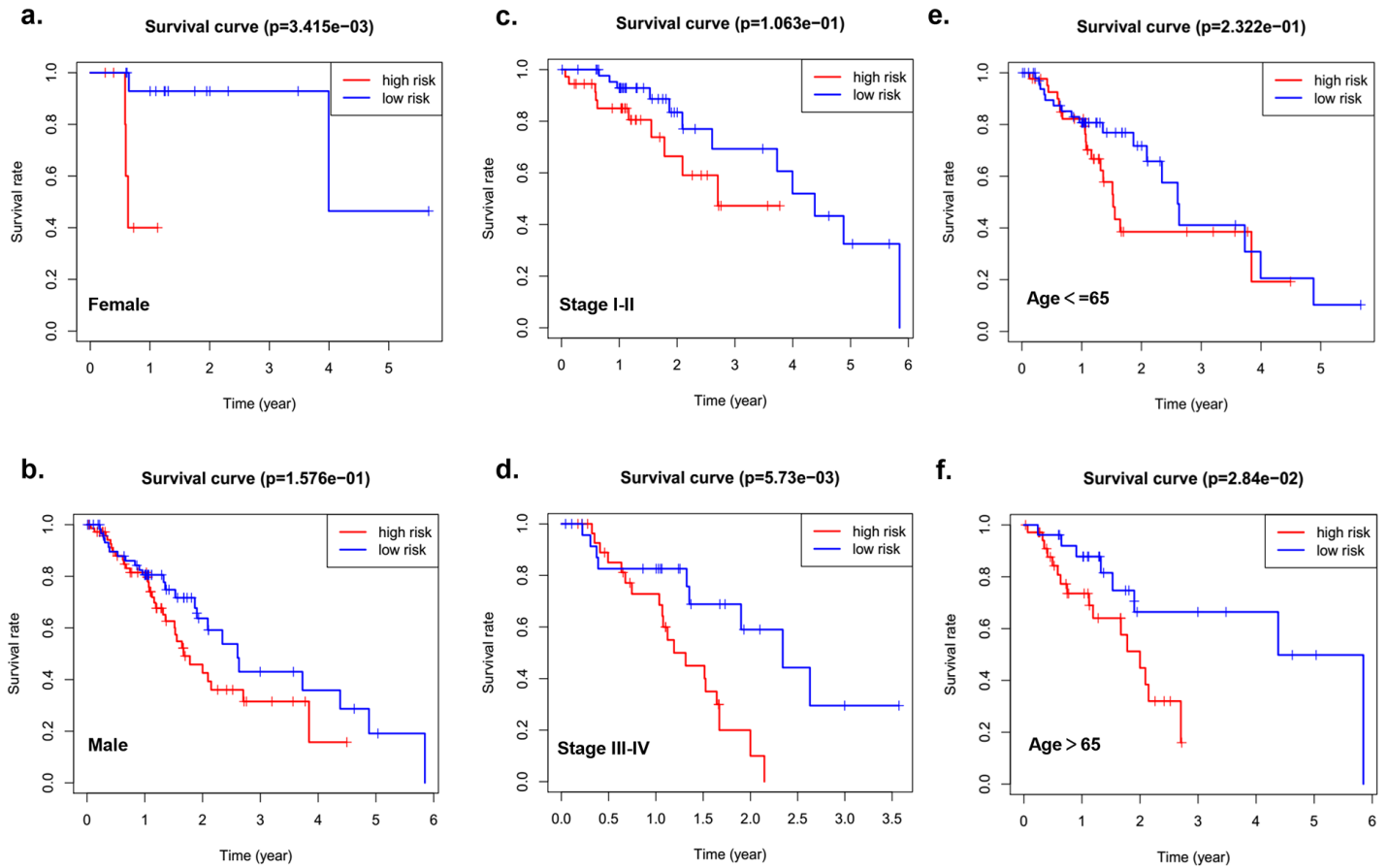

Figure 6

Subgroup analyses for survival in low-risk and high-risk group stratified by clinicopathological factors. a-b. Subgroup analyses for survival in low-risk and high-risk group stratified by gender; c-d. Subgroup analyses for survival in low-risk and high-risk group stratified by stage; e-f. Subgroup analyses for survival in low-risk and high-risk group stratified by age.
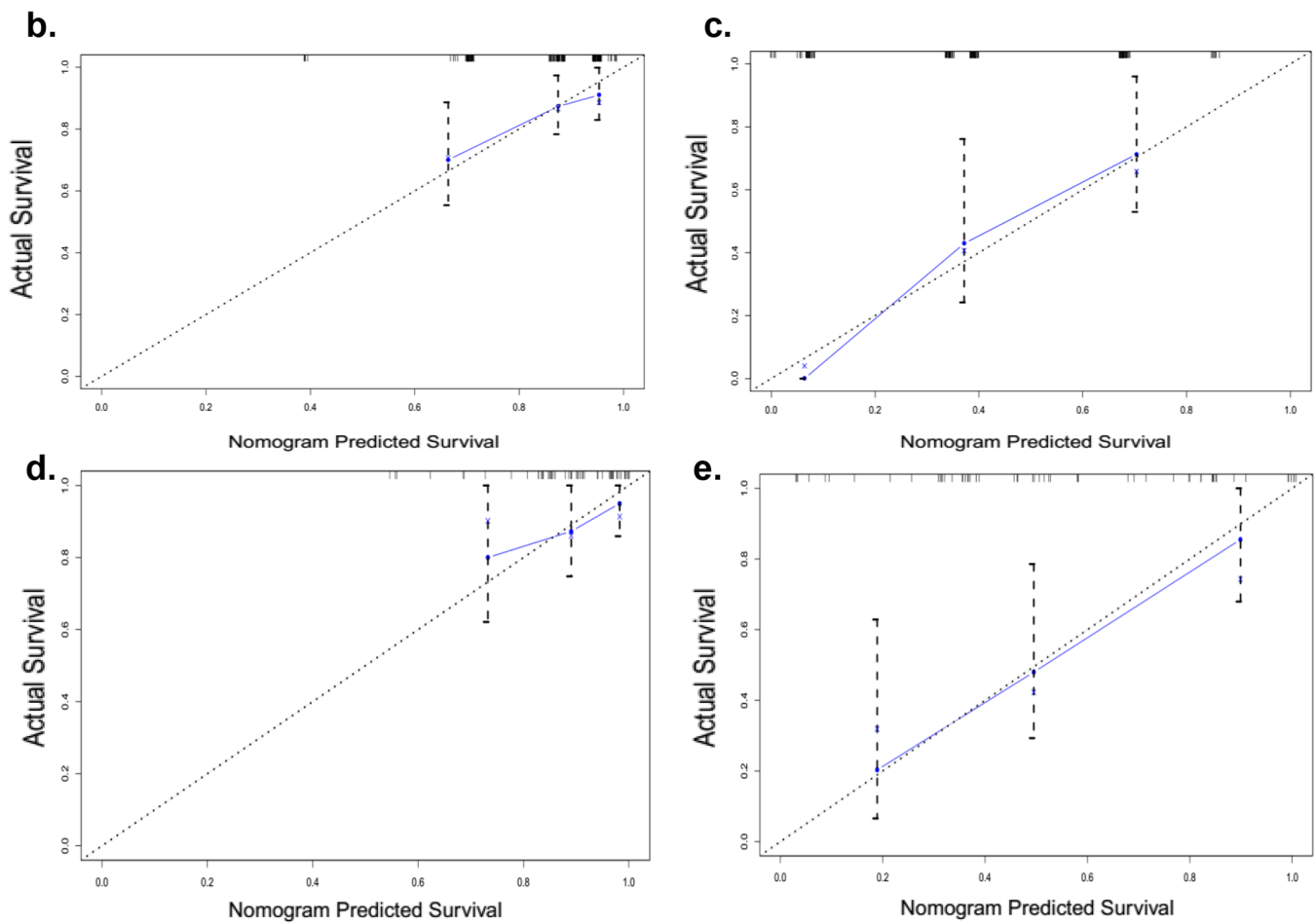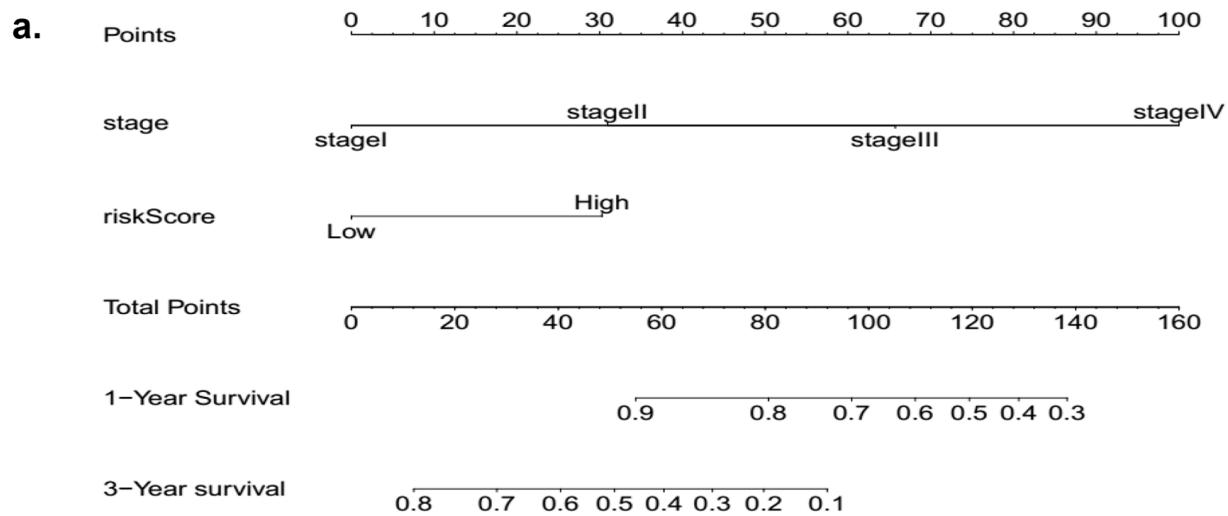
Figure 7

The clinical nomogram for predicting the prognosis in ESCA. a. The clinical nomogram of the training group (TCGA cohort) for predicting the OS in ESCA. b-c. Calibration curves present the concordance of 1-year (b) and 3-year survival (c) between the observation and the prediction in training group. d-e. Calibration curves present the concordance of 1-year (d) and 3-year survival (e) between the observation and the prediction in the GSE13898 ESCA cohort (testing group).
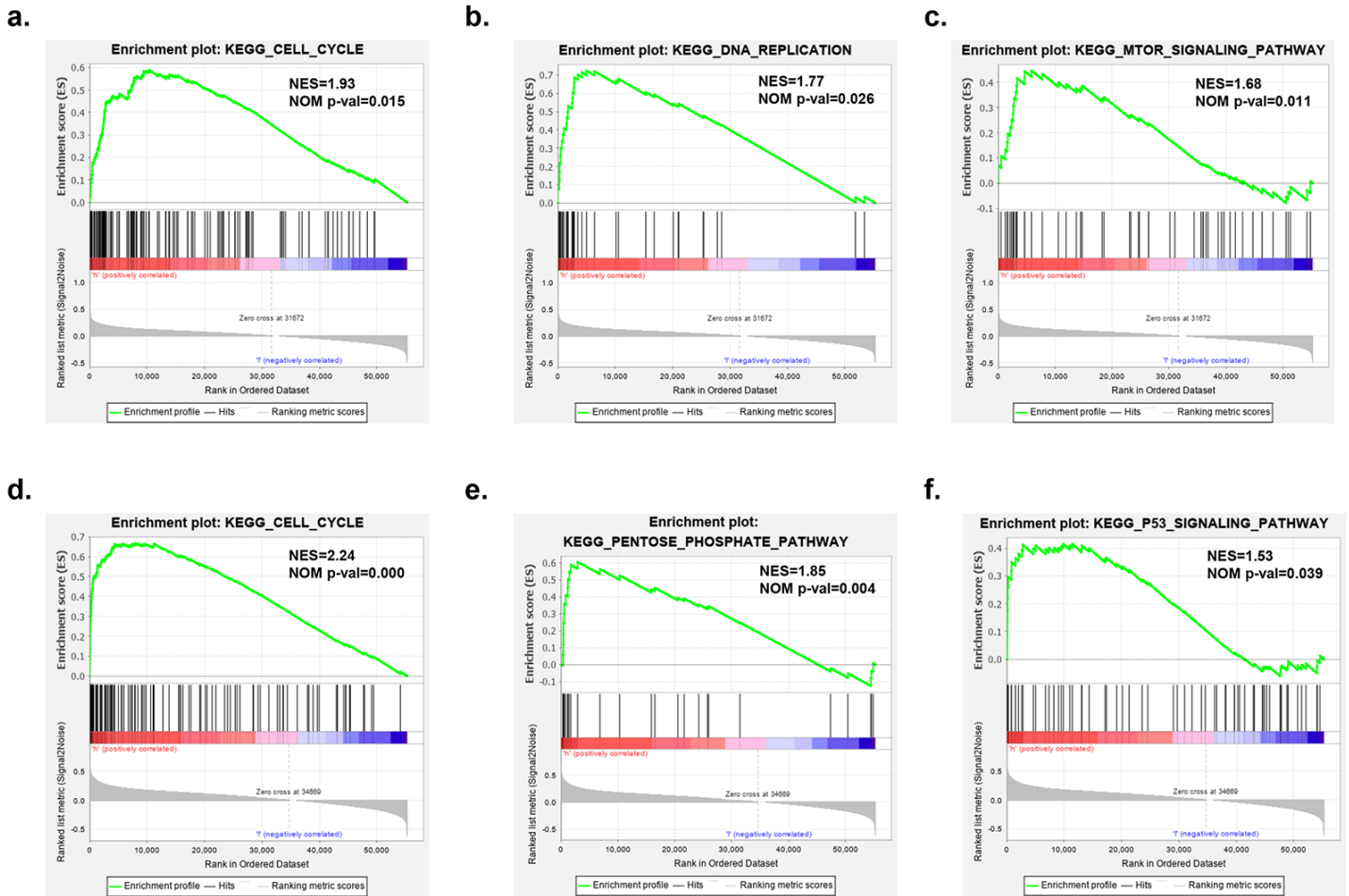
Figure 8

KEGG pathway enrichment plots from GSEA. GSEA results showing cell cycle (a), DNA replication (b), and mTOR signaling pathway (c) are differentially enriched in ALKBH5 high expression phenotype. Cell cycle (d), pentose phosphate pathway (e), and p53 signaling pathway (f) are differentially enriched in HNRNPC high expression phenotype. ES: enrichment score; NES: normalized ES; NOM p-val: normalized p-value.
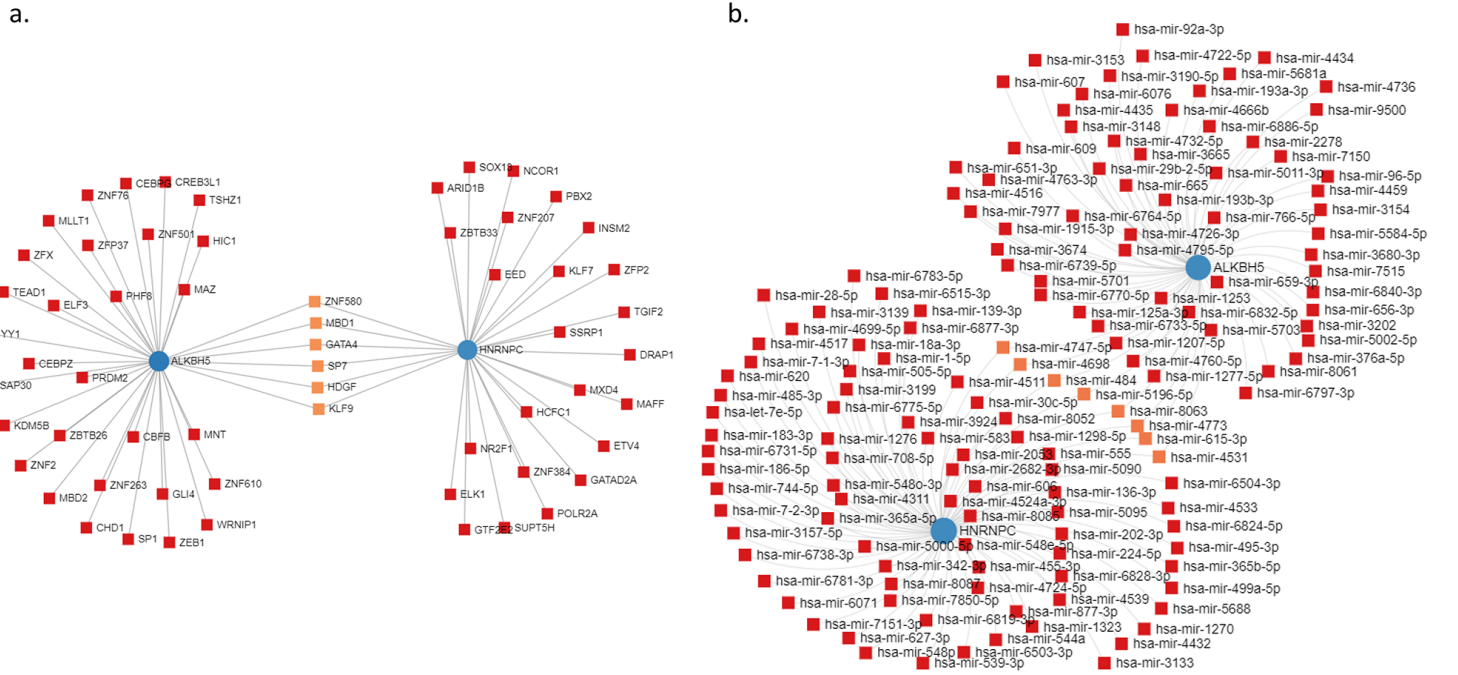
Figure 9

TFs-genes networks and miRNA-genes networks of AKLBH5 and HNRNPC a. The predicted networks of TFs and two m6A RNA methylation regulators (AKLBH5 and HNRNPC). The blue circles represent AKLBH5 and HNRNPC, the red and orange squares represent the predicted TFs; b. The predicted networks of miRNAs and two m6A RNA methylation regulators (AKLBH5 and HNRNPC). The blue circles represent AKLBH5 and HNRNPC; The red and orange squares represent the predicted miRNAs.

Supplementary Files

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

- figure.S1S7.docx