Development and validation of m6A RNA methylation regulators-based signature in lung adenocarcinoma

Wei Guo1, Qi-Lin Huai2, Si-Jin Sun1, Lei Guo3, Xue-Min Xue3, Peng Song1, Jian-Ming Ying3, Yi-Bo Gao1, Shu-Geng Gao1, Jie He1

1Department of Thoracic Surgery, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China; 2Department of Graduate School, Zunyi Medical University, Zunyi, Guizhou 563000, China; 3Department of Pathology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China.

To the Editor: Primary lung cancer is the most commonly diagnosed type of malignant tumor and is the leading cause of cancer death worldwide. Non-small cell lung cancer (NSCLC) makes up 80% to 85% of the overall incidents of primary lung cancer and is classified into two distinct histological subtypes: lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). LUAD constitutes nearly 30% to 35% of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1] As a prevalent mRNA of all primary lung cancer cases, and the recurrence of it seems to be increasing in most countries.[1]

Accumulative evidence has shown that m6A methylation can regulate gene expression in different physical processes, such as stem cell differentiation and pluripotency, circadian rhythm, embryogenesis, and DNA damage response, and has an important impact on the regulation of tumorigenesis and development.[4] In this study, we aim to develop an m6A RNA methylation regulators-premised prognostic signature for LUAD patients.

In this study, we used a wealth of bioinformatics techniques to construct the signature of m6A methylation regulators in the STRING database. Red, yellow, and blue nodes in the plot represent the writer, reader, and eraser, respectively. The six hub genes were RBM15, WTAP, METTL3, and METTL14 (writers), m6A-binding proteins IGF2BP1, IGF2BP2, IGF2BP3, YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, HNRNPC, and RBMX (readers), and the demethylases FTO and ALKBH5 (erasers).[5] In this study, we used a wealth of bioinformatics techniques to construct the signature of m6A methylation regulators in the STRING database. Red, yellow, and blue nodes in the plot represent the writer, reader, and eraser, respectively. The six hub genes were RBM15, WTAP, METTL3, and METTL14 (writers), m6A-binding proteins IGF2BP1, IGF2BP2, IGF2BP3, YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, HNRNPC, and RBMX (readers), and the demethylases FTO and ALKBH5 (erasers).[5]

Next, by performing univariate Cox regression analysis, eight genes including KIAA1429, RBM15, RBM15B, IGF2BP2, IGF2BP3, HNRNP2B1, YTHDF2, and HNRNPC were selected since their $P$ value was $<0.1$ [Figure 1A]. The simplest model ranging from $\lambda \pm 1\ SE$ was chosen [Figure 1B and 1C]. The coefficients for each factor were calculated by the Lasso-penalized Cox model. The risk score for each case was computed using the following formula: risk score = 0.0789 $\times$ ExpIGF2BP3 + 0.0755 $\times$ ExpHNRNP2B1 + 0.1493 $\times$ ExpHNRNPC. After model construction, all cases
Figure 1: Construction of the m6A prognostic signature. (A) Univariate Cox regression of 20 m6A RNA methylation regulators. (B) Cross-validation to determine \( \lambda \) based on C-index. The left vertical line indicates the \( \lambda \) values with maximum C-index and the right vertical line indicates the largest \( \lambda \) value within one standard deviation of maximum C-index. (C) The coefficient of each factor with selected \( \lambda \) value. The red line denotes the value selected for \( \lambda \). (D) The distributions of risk score, survival status, and gene expression pattern between low-risk and high-risk groups. (E) Kaplan–Meier curve of OS between low-risk and high-risk groups. (F) ROC curve was applied to evaluate the prediction performance of the m6A prognostic signature. AUC: Area under the curve; LUAD: Lung adenocarcinoma; m6A: N6-methyladenosine; OS: Overall survival; ROC: Receiver operating characteristic; TCGA: The Cancer Genome Atlas.
were stratified in low-risk and high-risk groups according to the median risk score. The distributions of risk score, survival status, and gene expression pattern between the two groups were shown in Figure 1D. Kaplan–Meier curve with log-rank test indicated that overall survival (OS) was significantly worse \((P < 0.05)\) in the high-risk than those in the low-risk group [Figure 1F]. The area under the curve (AUC) of the m6A prognostic signature for 2, 3, 5-year-survival were 0.628, 0.607, and 0.603 in TCGA-LUAD, respectively [Figure 1F]. In addition, univariate Cox regression demonstrated a significant association between OS and T stage, N stage, M stage, pathologic stage, and risk score \((P < 0.05)\). Further multivariate Cox regression analysis suggested that T stage and risk score were two independent unfavorable prognostic factors for LUAD patients \((P < 0.05)\) [Supplementary Figure 4, http://links.lww.com/CM9/A564]. The expression levels of the three regulators (IGF2BP3, HNRNPC, and HNRNPA2B1) in the prognostic signature all showed a negative correlation with the OS of LUAD patients in the GEPIA bioinformatics analysis platform [Supplementary Figure 5, http://links.lww.com/CM9/A564] except for HNRNPA2B1, the prognostic potential of the other two regulators was further confirmed in the Kaplan–Meier lung cancer cohort [Supplementary Figure 6, http://links.lww.com/CM9/A564]. Four independent datasets were used to assess the prognostic value of the m6A prognostic signature. Significantly higher OS was observed in the low-risk group compared with the high-risk group in three-quarters of the datasets \((P < 0.05)\). Model performance using AUC under the receiver operating characteristic (ROC) curve revealed that AUC values were >0.6 in three cohorts [Supplementary Figure 7, http://links.lww.com/CM9/A564]. To further improve prognostic ability, an integrated model combining the m6A prognostic signature and tumor node metastasis staging was developed. And the results demonstrated that the predictive accuracy of the integrated model was higher than that of the m6A signature or tumor node metastasis staging alone in all three datasets. The ROC curve also shows the high accuracy of model prediction [Supplementary Figure 8, http://links.lww.com/CM9/A564]. In addition, a PPI network was constructed to understand the most closely related genes with m6A signature, and Kyoto Encyclopedia of Genes and Genomes (KEGG) and gene ontology (GO) functional enrichment analysis were performed [Supplementary Figure 9, http://links.lww.com/CM9/A564]. The complete functional annotation list is provided in Supplementary Table 3, http://links.lww.com/CM9/A564. Finally, we used the Hallmark gene sets to perform GSEA on m6A prognostic signature, and the results showed that many gene sets in the samples with high-expression of the signature were positively enriched. The five most significantly related biological pathways are sequentially shown in Supplementary Figure 10, http://links.lww.com/CM9/A564.

In the present study, using TCGA-LUAD cohorts, we constructed a novel m6A RNA methylation regulators-based prognostic signature and successfully validated it in three independent validation datasets, indicating that this prognostic model is highly robust for prognosis prediction. Our prognostic signature revealed that the expression of IGF2BP3, HNRNPA2B1, and HNRNPC was positively related to the prognosis of LUAD. According to previous studies, IGF2BP3 is a powerful post-transcriptional oncogene that can enhance tumor growth, drug-resistance, and metastasis. In lung cancer, studies have reported that IGF2BP3 promoted lung tumorigenesis by reducing p53 stability. Currently, very limited details are available for the part played by HNRNPC2B1 and HNRNPC in tumorigenesis of lung cancer. HNRNPC2B1 could regulate the epithelial–mesenchymal transition in pancreatic cancer cells via the extracellular regulated protein kinases/smooth muscle α-actin signaling pathway. As for HNRNPC, some scholars have shown that HNRNPC might induce human lung cancer cell progression and metastasis through the activation of the interferon-α-Janus kinase–signal transducer and activator of transcription 1 signaling pathway. However, in general, related studies including other m6A RNA methylation regulators are very limited, and its role in LUAD is still unclear.

To sum up, this research has profiled the substantially altered m6A RNA methylation regulators between LUAD and ordinary controls, which could be vital to the development of LUAD. Moreover, a strong three-gene signature was constructed and validated in three different independent LUAD cohorts. The m6A prognostic signature might be a suitable prediction indicator in LUAD for predicting the long-term survival and identifying high-risk patients. Prospective, case-control design, and further research on key modulators in human samples must be conducted to verify its effectiveness in prediction and examine its practical value in tailor-made therapy and personalized treatment management.

Acknowledgements

The authors thank the patients and investigators who participated in TCGA and GEO for providing data.

Funding

This study was supported by grants from the National Key R&D Program of China (Nos. 2017YFC1311000 and 2018YFC1312100), the CAMS Initiative for Innovative Medicine (Nos. 2017-I2M-1-005, 2017-I2M-2-003, and 2019-I2M-2-002), the Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (Nos. 2018PT32033 and 2017PT32017), and the Innovation team development project of Ministry of Education (No. IRT_17R10).

Conflicts of interest

None.

References
