Estimating the prognosis of esophageal squamous cell carcinoma based on The Cancer Genome Atlas (TCGA) of m6A methylation-associated genes

Yu Pu, Xianfeng Lu, Xueqin Yang, Yi Yang, Dong Wang, Mengxia Li, Wei Guan, Mingfang Xu

Cancer Center of Daping Hospital, Army Medical University, Chongqing, China

Contributions: (I) Conception and design: M Xu, W Guan, D Wang, Y Pu, X Lu; (II) Administrative support: M Xu; (III) Provision of study materials or patients: X Yang, Y Pu, Y Yang; (IV) Collection and assembly of data: M Xu; (V) Data analysis and interpretation: M Xu, W Guan, M Li; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

These authors contributed equally to this work.

Correspondence to: Mingfang Xu; Wei Guan. Cancer Center of Daping Hospital, Army Medical University, Chongqing, China.
Email: xusiyi023@126.com; guanwei001@sina.com.

Background: N6-methyladenosine (m6A) mRNA modification is the most prevalent in certain tumors. However, its expression profile and prognostic value in human esophageal squamous cell carcinoma (ESCC) remains unknown.

Methods: Herein, we performed an extensive investigation of the m6A-associated gene expression profile and determined its significance in the prognosis of ESCC. We received the RNA expression profiles of 81 ESCC tissues and one normal esophageal tissue from The Cancer Genome Atlas (TCGA) database. Kaplan-Meier (KM) survival analysis was used to assess the predictability of m6A methylation-associated gene expression in ESCC prognosis. In addition, univariate and multivariate Cox regression, as well as least absolute shrinkage and selection operator (LASSO) regression models were employed for the establishment of prognostic signatures. Lastly, KM survival analysis, proportional hazard models, and receiver operating characteristic (ROC) curves were used to verify the prognostic value. Moreover, we also investigated the associations among the m6A prognostic signature, immune cell infiltration, and programmed cell death-ligand 1 (PD-L1) expression.

Results: We demonstrated that YTHDF3 [hazard ratio (HR): 0.910; 95% confidence interval (CI): 0.832–0.995; P=0.038], RBM15 (HR: 0.721; 95% CI: 0.549–0.948; P=0.019), KIAA1429 (HR: 0.801; 95% CI: 0.664–0.967; P=0.021), and ALKBH5 (HR: 0.948; 95% CI: 0.895–1.003; P=0.0.064) overexpression predicted better overall survival (OS) of ESCC patients. Furthermore, based on prognostic factors, the high-risk (H-R) cohort was found to have worse survival than the low-risk (L-R) cohort (P<0.001).

Conclusions: We revealed three m6A methylation-associated genes that were closely correlated with enhanced survival in ESCC patients. In addition, we generated an independent prognostic signature based on the expression of YTHDF3, RBM15, KIAA1429, and ALKBH5 genes. The results revealed significantly higher proportions of CD8+ T cells and higher expression of PD-L1 in the H-R group.

Keywords: N6-methyladenosine (m6A); esophageal squamous cell carcinoma (ESCC); prognosis; RNA methylation; immune cell infiltration

Submitted Sep 08, 2021. Accepted for publication Nov 25, 2021.
doi: 10.21037/jgo-21-686
View this article at: https://dx.doi.org/10.21037/jgo-21-686
Introduction

The esophageal squamous cell carcinoma (ESCC) subtype accounts for approximately 90% of all esophageal cancer (EC) cases worldwide. In 2012, there was a global estimate of 398,000 squamous cell carcinomas (SCCs) cases, with 79% occurring in South Eastern and Central Asia. Unfortunately, China accounted for 53% (210,000 cases) of all ESCC cases worldwide (1).

ESCC primarily appears in flat cells that line the upper two-thirds of the esophagus. It has multiple risk factors, including chronic irritation, inflammation, smoking and alcohol consumption, persistent ingestion of very hot drinks, high-temperature cooking, nutritional deficiencies (2-6), and so on. Thus far, there are no established prognostic factors for ESCC.

N6-methyladenosine (m6A)-mediated reversible methylation is the most common form of mammalian mRNA modification. It is modulated by the m6A methyltransferase (writer), binding protein (reader), and demethylase (erasers). Aberrant m6A methylation is correlated with the development of malignant tumors (7). The role of m6A-mediated reversible methylation in cancer prognosis is a double edged sword. Some genes promote the invasion and/or proliferation of cancerous cells, which ultimately lead to poor prognosis after methylation or recognition (8,9), while others promote tumor development after loss of methylation status (10). Given its close relation to tumor formation, m6A methylation is often examined, particularly in terms of tumorigenesis and prospective target screening for cancer therapy. At present, there are no exhaustive studies on the importance of the m6A methylation-associated genes in ESCC.

PubMed The Cancer Genome Atlas (TCGA) is an effective tool to examining the role of m6A methylation-related genes in cancer (11,12). In view of the differences between ESCC and adenocarcinoma, this study, unlike a previous study (13), excluded esophageal adenocarcinoma, consisted of only patients with ESCC. Emerging evidences suggest that immune checkpoint inhibitors are highly effective in treating EC (14,15). The tumor immune microenvironment (TIME), such as, infiltrated CD8+ cytotoxic T lymphocyte and programmed cell death-ligand 1 (PD-L1) expression exert crucial impact on response to immune checkpoint inhibitor-based cancer therapy (16,17). Multiple m6A genes were confirmed to be related to the immune microenvironment (18-20). Qiu et al. unveiled the regulation of PD-L1 expression by m6A methylation-associated genes (21).

In this study, we retrieved the mRNA profiles and epidemiologic data of 81 ESCC samples from TCGA. We analyzed the relationship between m6A modification and ESCC progression, identified biological indicators for overall survival (OS) of ESCC patients, and generated an ESCC prognosis signature using the aforementioned m6A-modified mRNA, in order to assess the predictability of m6A modification in ESCC prognosis. Lastly, we investigated the associations among the m6A prognostic signature, immune cell infiltration, and PD-L1 (CD274) expression (Figure 1).

We present the following article in accordance with the TRIPOD reporting checklist (available at https://dx.doi.org/10.21037/jgo-21-686).

Methods

Ethical statement

This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Study population and transcriptome data

The mRNA expression profiles of 81 ESCC tumors and one normal tissue, as well as the clinical information including clinical stage and TNM staging of these patients were collected from the TCGA database, and were used to correlate m6A methylation-associated gene expressions with patient survival (https://portal.gdc.cancer.gov/). The subsequent analyses strictly followed TCGA guidelines.

Bioinformatic analysis

The differential expression (DE) of 15 well-acknowledged m6A RNA methylation modulators [six writers (METTL3, METTL14, WTCP, RBM15, ZC3H13, and KIAA1429), two erasers (FTO and ALKBH5), and seven readers (YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, and HNRNPC)] in ESCC and normal control samples were analyzed using the R software (version 4.0.4) (https://cran.r-project.org/src/base/R-4/R-4.0.4.tar.gz), and plotted in the form of a heatmap. Inter-gene associations were identified using the Pearson correlation analyses. Lastly, the Kaplan-Meier (KM) survival analysis in R was employed to identify the link, if any, between each relevant gene and OS of ESCC patients. P<0.05 was the significance threshold.
Principal component analysis (PCA) and survival analyses of subgroups

We used R’s “ConsensusClusterPlus” software package () to identify different subgroups of the 81 tumor samples, and verified the grouping results by PCA. KM survival analysis was conducted to analyze the survival curves of both subgroups.

Prognostic value of m6A methylation-associated genes

Univariate analysis was used to compare the 15 m6A methylation-associated genes and select ESSC relevant genes, according to the following condition: P<0.1. Next, the least absolute shrinkage and selection operator (LASSO) regression tool in the ‘glmnet’ package in R software (https://cran.rstudio.com/bin/macosx/contrib/4.1/glmnet_4.1-2.tgz) was used to identify the best prognostic factors for high-dimensional data (22). Subsequently, we selected four genes and calculated their risk scores, and separated the participants into high-risk (H-R) and low-risk (L-R) cohorts based on the median expression of m6A methylation-associated genes. KM survival analysis (log-rank tests) and receiver operating characteristic (ROC) curve were used to verify the prognostic value of the grouped clusters. Lastly, a heatmap of clinical factors and m6A methylation-associated genes were plotted for an overview of the link between relevant gene expression and clinical consequences. Furthermore, we used the CIBERSORT algorithm to calculate the proportion of 22 immune cells in ESCC tissues, and evaluated the associations among prognostic grouping, immune cell infiltration, and PD-L1 expression.

Statistical analysis

For all the above analyses, a P value less than 0.05 was regarded as statistically significant.
Results

Transcriptome data and patient characteristics

An initial search of the TCGA database led to the identification of 81 ESCC tissues and one adjacent normal tissue with transcriptome data. Meanwhile, the clinical data of corresponding patients were collected, including clinical stage, TNM staging, as well as survival time and survival status.

M6A methylation-associated gene profile in ESCC

A total of 15 m6A regulators were identified in this study. Figure 2A depicts a heatmap of the 15 m6A methylation-associated genes that were DE in ESCC. Since there was only one case in the control group, we did not perform a differential analysis between the tumor and normal tissues. Conclusions from our inter-gene correlation analysis are presented in Figure 2B. Based on our analysis, METTL14 was found to be strongly associated with multiple genes including RBM15, YTHDC1, and YTHDF2, with correlation coefficients of 0.58, 0.58, and 0.52, respectively.

Survival analysis of m6A methylation-associated genes

To evaluate the importance of the 15 m6A methylation-associated genes in patient survival, we conducted KM survival analysis in the R software. We demonstrated that elevated KIAA1429 and YTHDF3 levels were strongly correlated with poor survival (P<0.05) (Figure 3A,3B), whereas the remaining 13 genes showed no correlation with patient survival (P>0.05).

Furthermore, using the ConsensusClusterPlus package in R, we separated the 81 tumor samples into discrete subgroups based on their m6A methylation status, and determined cluster- and item-consensus (Figure 4A-4F). Interference between clusters was at a minimum when k=2 (Figure 4G,4H). Our results demonstrated the largest differences between clusters.

All participants were next stratified into two subgroups based on the most stable k value. Subgroup cluster 1 exhibited scarce gene expression, while subgroup cluster 2 displayed augmented expression (Figure 4G). In addition, the horizontal axis represented the principal component, whereas the vertical axis represented a different component. PCA showed individual subgroups in two clusters. OS analysis of DE genes revealed that subgroup 2 exhibited slightly longer survival duration; however, the difference was not statistically significant (P=0.086) (Figure 4H).

LASSO model generation

Fifteen m6A RNA methylation modulators were analyzed using univariate analysis and four candidate genes were identified at P<0.1 (Figure 5A). Next, we used the LASSO Cox regression model to identify the genes with the best predictive value. Based on our analysis, YTHDF3, RBM15, KIAA1429, and ALKBH5 were selected as indicators of ESCC prognosis (Figure 5B,5C). Subsequently, we calculated the risk scores of these genes for additional univariate and multivariate analyses.

The participants were then assigned to either a H-R or L-R cohort based on the combination model, with the median expression of the four candidate genes as the threshold. According to the KM survival curve, the H-R cohort exhibited markedly poor patient prognosis relative to the L-R cohort (P<0.001) (Figure 5D). Next, we analyzed the risk factor prognostic efficiency using the ROC curve. The area under the curve (AUC) was 66.3% (Figure 5E), suggesting a strong potential of m6A methylation-associated genes as biomarkers for ESCC prognosis.

Prognostic value of the four m6A methylation-associated genes

Using univariate analysis, we found that the N and clinical stages, as well as the risk score of m6A methylation-associated genes accurately estimated patient prognosis (P<0.05), whereas the T and M stages did not (P>0.05) (Figure 6A). Since the age and clinical grade data were missing, and all female patients were still alive at the end of the statistical time, we chose not to analyze the significance of age, grade, and gender on patient survival. In addition, multiple regression analysis revealed that the TNM stages, and particularly, the risk scores of the m6A methylation-associated genes, were discrete indicators of ESCC prognosis (P<0.05) (Figure 6B). Moreover, we showed that YTHDF3, RBM15, KIAA1429, and ALKBH5 levels exerted a protective function and were elevated in L-R patients (Figure 6C).

We next used CIBERSORT to investigate the association between the m6A prognostic grouping and proportions of 22 immune cell types. The results showed significantly low proportions of M0 macrophages and resting CD4+ memory...
T cells and high proportions of CD8⁺ T cells in the H-R group, relative to the L-R group (Figure 7A-7C). We also investigated the association between the m6A prognostic grouping and PD-L1 expression. The results showed significantly higher expression of PD-L1 in the H-R group, relative to L-R (Figure 8).

**Discussion**

A study that stratified the global incidence of EC by histological subtypes revealed an estimated 398,000 ESCC cases in 2012, which was far greater than the incidence of esophageal adenocarcinoma [52,000] (1). In the Chinese population, ESCC is the predominant histological
 subtype of EC (23,24), which is a highly heterogeneous disease without effective therapy or available prognostic biomarkers (25). At present, the exact molecular mechanism of ESCC is not fully understood. As a result, there are limited targeted therapies, and therefore, poor-prognosis related to ESCC. Increasing evidences (8,26,27) suggest a strong contribution of m6A modification in tumor proliferation, differentiation, tumorigenesis, invasion, and metastasis. Additionally, m6A methylation can regulate both oncogenic and antioncogenic functions. Being a highly prevalent form of RNA modification, m6A is modulated by multiple regulatory proteins and, in turn, regulates numerous cellular functions (28-30). Among its modulators are methyltransferases (‘writers’ like METTL3, METTL14, WTAP, RBM15, ZC3H13, and KIAA1429), RNA binding proteins (‘readers’ like YTHDF1, YTHDF2, YTHDF3, YTHDC1, YTHDC2, and HNRNPC), and demethylases (‘erasers’ like FTO and ALKBH5) (31-34).

Multiple studies have performed exhaustive analyses of the relationship between m6A regulator expression and epidemiological data of cancer patients from TCGA. Among the cancer types examined were colorectal cancer (35), hepatocellular carcinoma (36), and lung adenocarcinoma (37). At present, there are no studies on the correlation between m6A regulator expression and ESCC prognosis. Hence, we investigated the link between m6A regulator expression and ESCC prognosis in 81 ESCC tumor samples and one normal esophageal tissue.

Using the cBioPortal database, Zhao et al. demonstrated that the elevated expressions of HNRNPC, YTHDC2, WTAP, VIRMA, IGF2BP3, and HNRNPA2B1 genes were strongly correlated with poor prognosis in EC patients (13). However, this study showed no association between m6A regulators and OS of ESCC patients. In another study involving adenocarcinoma, it was revealed that ALKBH5, YTHDF2, and METTL14 were closely linked to enhanced OS, whereas augmented levels of HNRNPC and WTAP were associated with poor OS. However, in this study, we discovered YTHDF3, RBM15, and KIAA1429 levels to be strongly related to better OS in ESCC patients. Conversely, our data showed that the levels of HNRNPC, YTHDC2, WTAP, and HNRNPA2B1 have no relation to ESCC prognosis. One possible reason for this discrepancy may be that the same m6A methylation regulator may play different roles in different tumors (30). Interestingly, it may also play different roles in histologically different tumors within the same organ.

In our study, we generated two clusters of the TCGA-ESCC cohort based on the individual m6A methylation-associated gene profiles. We demonstrated a slight difference in OS between the two subgroups; however, this was not statistically significant, likely due to the small sample size and short follow-up time. Based on our analysis, we also generated a prognostic signature, involving YTHDF3, RBM15, KIAA1429, and ALKBH5 expression profiles, which showed an excellent performance in estimating ESCC prognosis.

In this study, we investigated the association between prognostic grouping (based on four-m6A methylation-related genes risk signature), infiltration of immune cells, and PD-L1 expression. Relative to L-R ESCC patients, we observed obviously low proportions of M0 macrophages and
Figure 4 Identification and analysis of two subgroups of tumor samples that exhibited distinct m6A expression. (A-C) Consensus clustering matrix for k=2, 3, and 4. (D-F) Tracking plot for k=2 to 9. (G) PCA of the two subgroups. (H) KM survival plots of the two subgroups. m6A, N6-methyladenosine; PCA, principal component analysis; KM, Kaplan-Meier; CDF, cumulative distribution function.
resting CD4⁺ memory T cells, as well as high proportions of CD8⁺ T cells, and PD-L1 expression in tumor tissues of H-R ESCC patients. Previous studies also reported that m6A methylation-related genes play a critical role in the anti-tumor immune response (38,39). Paradoxically, the PALACE-1 study (40) indicated that the percentage of transcription factor 1 positive CD8⁺ T cells was significantly higher in pathological complete remission (pCR) group specimens, compared to the non-pCR group after co-treatment with pembrolizumab and chemoradiotherapy. There is increasing evidence that the PD-L1 expression (41-43) is associated with the efficacy of immune checkpoint inhibitors in EC. Based on these studies, we speculate that the elevated risk, identified by the m6A regulatory gene risk signature, may be due to alterations in the immune microenvironment as well as upregulated PD-L1 expression in tumor tissues. Introducing immune checkpoint inhibitors may reverse the poor prognosis associated with the H-R ESCC patients.

Only one normal esophageal tissue in TCGA database was the deficiency of this study, because it is important to have a proper number of controls to determine the physiologically normal methylation status, even though the purpose of this study was to detect the effect of m6A methylation related genes on the prognosis of patients with ESCC, rather than the pathogenic factors of the genes themselves. We also examined the expression of prognostic signature genes in postoperative pathological ESCC specimens to verify its correlation with prognosis. However, due to the long survival time of postoperative patients with EC, the median survival time could not be determined. We plan to continue our investigation during outpatient follow-ups.
Figure 6 Forest plot and heatmap of the prognostic signatures and clinical risk factors. (A) Forest plot of univariate Cox regression analysis in ESCC. (B) Forest plot of multivariate Cox regression analysis in ESCC. (C) Heatmap of four m6A methylation-related genes and clinical risk factors in ESCC. * P<0.05. ESCC, esophageal squamous cell carcinoma; m6A, N6-methyladenosine.
Conclusions

In conclusion, we demonstrated that elevated levels of YTHDF3, RBM15, and KIAA1429 genes are strongly linked to enhanced OS in ESCC patients. Moreover, using LASSO regression analysis, we generated a prognostic signature involving the gene expressions of YTHDF3, RBM15, KIAA1429, and ALKBH5, and demonstrated this to be a discrete indicator of ESCC prognosis and a biomarker of immune checkpoint inhibitors. We speculate that the introducing of immune checkpoint inhibitors could reverse the adverse outcomes of H-R group identified by this prognostic signature. We believe that the conclusions of this study will add insight into the growing literature of prospective ESCC therapies.

Acknowledgments

Funding: This work was supported in part by the Chongqing Science and Health Joint Medical Research Project (No. 2021MSXM009) and the Fund of Key Specialized Construction in Chongqing, China.

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://dx.doi.org/10.21037/jgo-21-686

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/jgo-21-686). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. This study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).
Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References
