A signature of five 7-methylguanosine-related genes is a prognostic marker for lung squamous cell carcinoma.

Wensheng Liu
Ying Wang
Paola Ulivi
Simona Tavolari
Syed A A Rizvi

See next page for additional authors

Follow this and additional works at: https://mouseion.jax.org/stfb2023
Authors
Wensheng Liu, Ying Wang, Paola Ulivi, Simona Tavolari, Syed A A Rizvi, Enrico Capobianco, Alfons Navarro, and Yongjun Zhang
A signature of five 7-methylguanosine-related genes is a prognostic marker for lung squamous cell carcinoma

Wensheng Liu¹, Ying Wang², Paola Ulivi³, Simona Tavolari⁴, Syed A. A. Rizvi⁵, Enrico Capobianco⁶, Alfons Navarro⁷, Yongjun Zhang⁸

¹Critical Care Medicine Department, Zhejiang Cancer Hospital, Hangzhou, China; ²Gynecological Oncology Department, Zhejiang Cancer Hospital, Hangzhou, China; ³Biosciences Laboratory, IRCCS Istituto Ortopedico Rizzoli per lo Studio dei Tumori “Dino Amadori” (IRST), Meldola, Italy; ⁴Medical Oncology, IRCCS Azienda Ospedaliero-Università di Bologna, Bologna, Italy; ⁵College of Biomedical Sciences, Larkin University, Miami, FL, USA; ⁶Department of Computational Science, The Jackson Laboratory, Farmington, CT, USA; ⁷Molecular Oncology and Embryology Laboratory, Department of Surgery and Surgical Specializations, Human Anatomy Unit, Faculty of Medicine and Health Sciences, University of Barcelona (UB), Barcelona, Spain; ⁸Integration of Traditional Chinese and Western Medicine Department, Zhejiang Cancer Hospital, Hangzhou, China

Contributions: (I) Conception and design: W Liu, Y Zhang; (II) Administrative support: Y Wang; (III) Provision of study materials or patients: Y Wang; (IV) Collection and assembly of data: W Liu; (V) Data analysis and interpretation: Y Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yongjun Zhang, MM. Integration of Traditional Chinese and Western Medicine Department, Zhejiang Cancer Hospital, 1 Banshan Eastern Road, Hangzhou 310022, China. Email: zhangyj@zjcc.org.cn.

Background: N³-methylguanosine (m7G) is an important posttranscriptional modification affecting mRNA and tRNA functions and stability. The genes regulating the m7G process have been previously found involved in the carcinogenesis process. We aimed to analyze the role of m7G-related genes as potential prognostic markers for lung squamous cell carcinoma (LSCC).

Methods: Twenty-nine m7G-related genes were selected for the analysis in the LSCC cohort of the Cancer Genome Atlas (TCGA). Univariate, multivariate, and Kaplan-Meier analyses were used to evaluate the predictive value of risk model developed with m7G signature for overall survival (OS). The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses of differentially expressed genes (DEGs) were performed for high- and low-risk LSCC groups.

Results: We identified 17 differentially expressed m7G methylation-related genes in LSCC versus normal tissues. The expression of five m7G-related genes (EIF3D, LSM1, NCBP2, NUDT10, and NUDT11) was identified as an independent prognostic marker for OS in LSCC patients. A risk model with these five m7G-related genes predicted 2- and 3-year survival rates of 0.623 and 0.626, respectively. The risk score significantly correlated with OS: LSCC patients with a higher risk score had shorter OS (P<0.01) and it was associated with lower immune response (P<0.01).

Conclusions: We developed a novel m7G-related gene signature that can be of great utility to predict the prognosis for patients with LSCC.

Keywords: Non-small cell lung cancer (NSCLC); N³-methylguanosine (m7G); The Cancer Genome Atlas (TCGA); prognosis; biomarker

Submitted Sep 26, 2023. Accepted for publication Nov 09, 2023. Published online Nov 27, 2023.
doi: 10.21037/jtd-23-1504
View this article at: https://dx.doi.org/10.21037/jtd-23-1504
Introduction

Lung cancer is one of the most commonly diagnosed cancer and the leading cause of cancer-related deaths in the world, and it continues to be a significant health burden globally (1). Histologically, up to 85% of all lung cancer cases are diagnosed as non-small cell lung cancer (NSCLC), which can be further classified as lung adenocarcinoma (LUAD) or lung squamous cell carcinoma (LSCC) (2). Etiologically, tobacco smoke and various agents that can induce DNA damage and DNA repair deficiency, are the leading risk factors for lung cancer development (3). Tobacco smoke contains thousands of constituents, many of which can damage the genomic DNA and induce normal cell transformation and carcinogenesis (4,5). Clinically, LSCC occurs more frequently in males than in females and is closely associated with tobacco smoke (6). In addition, it is characterized by distinct gene alterations and clinical outcomes compared with those of LUAD (3,7). Advanced LSCC is associated with a very poor patient prognosis due to the lack of early detection biomarkers and treatment options (8,9). Thus, further investigation of NSCLC, including LSCC, could help to elucidate the underlying molecular mechanisms in order to discover novel biomarkers. Especially early detection and prediction of prognosis and treatment outcomes as well as novel strategies for the treatment of this very deadly disease could thus be developed.

The literature indicates, epigenetic alterations, such as RNA methylation, contribute to lung cancer development and progression (10-12). N7-methylguanosine (m7G) is an endogenous methylated nucleoside found in different RNA molecules; for example, when m7G occurs in RNA messenger (mRNA), it can regulate mRNA export, translation, and splicing (13), and when m7G occurs in RNA transfer (tRNA), it can change tRNA functions to affect mRNA translation and cell growth (14). As is well known, tRNA belongs to a class of noncoding RNAs and serves as a physical link between the amino acids and the ribosomes according to the matched codon in the mRNA molecules (15,16). To date, approximately 90 different modifications that occur in tRNA molecules have been reported (17,18). Although their complete functional implications remain to be determined, m7G is one of the most frequently found in tRNA. If it occurs at position 46 of tRNA, m7G will form a tertiary base pair with C13-G22 to stabilize the 3-dimensional (3D) core of the tRNA (19-21). In addition, m7G modification of tRNA is mediated by the METTL1-WDR4 complex (22), of which METTL1 is a writer of m7G in mRNA and various noncoding RNAs, such as tRNA (23). Gene mutations or altered functions of the enzyme can contribute to human disease; for example, a mutation in the WDR4 gene will impair tRNA m7G46 methylation and cause microcephalic primordial dwarfism (24,25). Recent studies also have shown that METTL1-mediated m7G editing is important in inhibition of lung cancer cell migration (26,27). Thus, in this study, we assessed the expression of m7G-related genes as a gene signature, which were selected in base of the regulation role of these genes in non-small cell lung cancer (28-30), to predict the prognosis of LSCC patients. We present this article in accordance with the TRIPOD reporting checklist (available at https://jtd.amegroups.com/article/view/10.21037/jtd-23-1504/rc).

Methods

Database searching and data downloading

In this study, we first searched the Cancer Genome Atlas (TCGA) database up to November 15, 2022 (https://portal.gdc.cancer.gov) to identify the differentially expressed genes (DEGs) in LSCC samples and downloaded the data of 502 LSCC patients and 49 normal adjacent lung tissues. We then searched the m7G-related genes and the complete clinicopathological information of the patients, including gender, age, tumor-node-metastasis (TNM) stage, tobacco...
smoking history, and survival data. We then included 29 m7G methylation-related genes to construct a risk model for LSCC (Table 1) after thoroughly searching the published literature (13) and the Gene Set Enrichment Analysis (GSEA) database (http://www.gsea-msigdb.org/gsea/login.jsp). The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

**Data analysis**

First, the differential expression of all m7G methylation-related genes between LSCC and adjacent normal tissues was determined, and then the data were imported into the limma R package (R Foundation for Statistical Computing, Vienna, Austria), according to a previous study (31). The genes with a log fold change (FC) >0.5 and an adjusted P value <0.001 were defined as DEGs.

**Survival analysis and construction of a risk prediction model**

First, we associated these 29 m7G methylation-related genes with the survival of LSCC patients using univariate Cox regression survival analysis. Second, we selected genes with a significant P value. The selections were instrumental to construct the risk prediction model by integrating the expression level of each gene and their corresponding coefficients. Third, we defined the risk prediction score as the risk score, and the predictive power of the risk score was used to predict the 1-, 2-, and 3-year survival rates of the patients using receiver operating characteristic (ROC) curve analysis. Finally, we further performed univariate and multivariate Cox regression analyses to identify the most significant independent risk factors for LSCC patients.

**Construction of a protein-protein interaction (PPI) network**

As the functions of any given gene are mediated through their coding proteins, a PPI network of the 29 m7G methylation-related genes using the igraph package in R was constructed. The PPI network shows the 29 m7G methylation-related genes as nodes, whereas each line connecting two nodes illustrates their biological relationship. Red lines indicate an upregulated correlation, whereas blue lines a downregulated correlation, with the color intensity representing the strength of the correlation.

**Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses**

Next, we performed GO and KEGG pathway enrichment analyses of the DEGs to identify their interactions and gene
pathways in LSCC development. In brief, based on the risk score (see above), we categorized all patients into a high- or low-risk category using the median risk score as the cut-off value and utilized the limma R package to identify all DEGs between the high- and low-risk categories. Genes were defined as DEGs when the false discovery rate (FDR) <0.05 and |log2FC| ≥1, and then these genes were analyzed for the GO terms and KEGG pathways using the clusterProfiler package in R, according to a previous study (32).

**Numeration of the immune infiltration score**

The immune response is important in NSCLC, especially in tobacco smoke-related carcinogenesis (33,34). Thus, we profiled the immune cell population and activation of immune-associated pathways in the tumor microenvironment, gene set signatures in different immune cells, and immune-associated pathways from the literature (28) and utilized the molecular signature database to explore their role in LSCC (https://www.gsea-msigdb.org/gsea/msigdb/). The enrichment score of each signature for each LSCC sample was then inferred based on the RNA-sequencing data and the single-sample GSEA (ssGSEA) by using the gene set variation analysis (GSVA) R package.

**Statistical analysis**

In this study, inlimma R Software package and GraphPad Prism 5.0 (GraphPad Software Inc., La Jolla, CA, USA) were used for statistical analysis. T-test was used for comparison among groups difference analysis, ANOVA was used for continuous variables, and categorical variables were analyzed using Chi-square tests. P value less than 0.05 is defined as significance.

**Results**

**Identification of differentially expressed m7G methylation-related genes in LSCC**

In this study, the flow chart is shown in Figure 1, we first searched the TCGA database for DEGs in 502 LSCC samples versus 49 normal tissue samples and found 17 m7G methylation-related DEGs in LSCC tissues versus normal tissues using the following cut-off values: |log2FC| >0.5, FDR <0.05, and P<0.01 (Figure 2A). Of these, 14 genes (NUDT10, NUDT3, NUDT11, SNUPN, AGO2, WDR4, DCPS, NCBP1, METTL1, LSM1, NSUN2, LARP1, NCBP2, and EIF3D) were upregulated in LSCC tissues, whereas 3 genes (NCBP2L, IFT5, and EIF4E3) were downregulated.

We then performed the PPI network analysis and evaluated the relationship among 29 m7G methylation-related genes (Figure 2B).

**Classification and prognosis prediction of LSCC patients according to the expression of m7G methylation-related genes**

We analyzed the consensus clustering of the investigated 29 m7G-related genes in LSCC tissue samples by selecting k=2 to divide these 502 LSCC patients into high- and low-expression clusters (Figure 3A). The relationships between the gene expression profile and the clinical features (age ≤60 vs. >60 years, TNM stage, history of tobacco smoking, and survival data according to the high- and low-expression clusters) were presented in a heatmap (Figure 3B). Although we were unable to find significant differences in the clinical features between these two clusters, we did find a significant difference in the overall survival (OS) of patients between the two clusters (P=0.0062; Figure 3C), indicating the usefulness of this gene signature in predicting the prognosis of LSCC patients.

Univariate Cox regression analysis was used for primary screening of the survival-related genes. The results revealed that the expression of five m7G methylation-related genes (EIF3D, LSM1, NCBP2, NUDT10 and NUDT11) was significantly associated with the OS of the LSCC patients (Figure 4A). We then performed the least absolute shrinkage and selection operator (LASSO) Cox regression analysis to construct a gene signature using these five genes with the optimum λ value (Figure 4B). Subsequently, multivariate Cox regression analysis was performed to calculate the risk score for each LSCC case. The risk model formula for the risk score was as follows: risk score = (−0.005762 × EIF3D exp.) + (−0.007321 × LSM1 exp.) + (−0.005766 × NCBP2 exp.) + (−0.01864 × NUDT10 exp.) + (−0.012726 × NUDT11 exp.). We then classified these patients into high- and low-expression clusters based on the cut-off value of the median risk score (Figure 4C) and performed principal component analysis (PCA; Figure 4D). We found that the LSCC patients with a high-risk score had a significantly shorter OS than those with a low-risk score (P=0.002; Figure 4E). Moreover, our ROC analysis revealed that such a risk score was able to predict the 1-, 2-, and 3-year survival rates of LSCC patients after surgery, with area under the curve (AUC) values of 0.542, 0.623, and 0.626, respectively (Figure 4F).
Identification of the signature of five m7G methylation-related genes as an independent prognostic factor for LSCC patients

Using univariate Cox regression analysis we stated that the higher the risk score was, the worse the prognosis of LSCC [hazard ratio (HR) =1.5503; 95% confidence interval (CI): 1.1585–2.0745; P=0.0032, Figure 5A]. To confirm the predictive value of this signature of five m7G methylation-related genes as an independent prognostic factor for LSCC patients, we performed a multivariate Cox regression analyses of the risk score and clinical features of the patients. We found that the risk score was an independent prognostic factor for the LSCC patients, specifically, the higher the risk score was, the worse the prognosis of LSCC (HR =1.57; 95% CI: 1.1702–2.1064; P=0.0026, Figure 5B). The pathological stage and the smoking status also emerged as independent factors for OS.

Furthermore, we associated this risk score with the clinicopathological features of the patients and the expression of the individual genes and found that the patients with a high-risk score exhibited significantly lower expression of NUDT10, NUDT11, LSM1, EIF3D, and NCBP2 (Figure 5C).

Figure 1 Flow chart of data collection and analysis. LSCC, lung squamous cell carcinoma; TCGA, The Cancer Genome Atlas; m7G, N7-methylguanosine; DEGs, differentially expressed genes; LASSO, least absolute shrinkage and selection operator.
**Figure 2** Expression of 29 m7G-related genes in LSCC tissues and their interactions. (A) Heatmap plot showing the expression levels of 29 m7G-related genes in LSCC and normal tissues. The red bars indicate high expression, whereas the green bars refer low expression. (B) PPI network among 29 m7G-related gene-coded proteins. The red bars indicate upregulation, whereas the blue bars refer to downregulation, with the depth of color indicating the strength of the relevance. ***, P<0.001. N, normal tissue; T, tumour tissue; LSCC, lung squamous cell carcinoma; PPI, protein-protein interaction; m7G, N°-methylguanosine.

**Association of the risk score with other biological features of LSCC**

According to the risk score, we classified the patients and imported the data into the limma R package to search for the DEGs between the high- and low-risk groups using FDR <0.05 and |log₂ FC| ≥1. We obtained a total of 496 DEGs (332 upregulated genes and 164 downregulated genes in the high-risk group of patients; Table S1). We then performed GO and KEGG pathway analyses using the clusterProfiler package in R. Our data showed that the DEGs based on the risk model of this signature of five m7G methylation-related genes were mainly correlated with humoral immune response, cellular calcium ion homeostasis, and cytokine-cytokine receptor interaction (Figure 6A,6B).

**Comparison of the immune activity between subgroups**

To evaluate the changes in the immune profile according to the risk score of each patient, we enumerated the abundance of different tumor-infiltrating immune cells in the tumor mass using the whole-exome expression data (28) and found that the LSCC patients with high-risk scores had a low immune inflammatory microenvironment, which was shown by significantly lower levels of tumor-infiltrating immune cells such as T- and B-lymphocytes, dendritic cells, macrophages, and neutrophils (Figure 7A). After that, we obtained the enrichment scores for 13 immune-associated pathways, such as the cytotoxic activity, antigen presentation, inflammation-promoting, and interferon pathways, to compare their activities between the high- and low-risk cohorts of patients. As shown in Figure 7B, the high-risk-scored tumors had significantly lower activations of the pathways related to immune checkpoint activation, cytotoxic activity, antigen presentation, inflammatory response, or type II interferon response.

**Discussion**

LSCC treatment depends on many factors including the tumor stage, resectability, performance status, and genomic alterations; advanced unresectable LSCC tissues are treated with chemotherapy, radiation therapy, epidermal growth factor receptor (EGFR)-targeted therapy, antiangiogenetic therapy, and/or immune therapy; however, although the decades-long development of vascular endothelial growth factor receptor (VEGFR) inhibition and recent immunotherapy have been shown to improve the survival
and quality of life of LSCC patients, the effectiveness of LSCC treatment has plateaued for several decades (35,36). At present, predictive biomarkers have been reported in LUAD (37,38), few studies was done in LSCC biomarkers. Thus, it is necessary to search for novel biomarkers could help medical oncologists to successfully predict LSCC treatment outcomes and prognoses. In the current study, we evaluated the prognostic significance of m7G methylation-related gene expression in LSCC tissues and established a signature of five m7G methylation-related genes (EIF3D, LSM1, NCBP2, NUDT10, and NUDT11) as a risk model to predict the OS of LSCC patients. We found that a high-risk score indicated a worse prognosis for LSCC patients. We also demonstrated that LSCC tissues from patients with

Figure 3  LSCC classification according to the expression level of different m7G-related genes. (A) Illustration of patients according to high and low expression of the m7G-related genes. A total of 502 LSCC patients were divided into high and low expression of the m7G-related genes according to the consensus clustering matrix (k=2). (B) Heat map and the clinicopathological characteristics of the two clusters of different m7G-related gene expression levels. The red bars indicate high expression, whereas the green bars refer low expression. (C) Kaplan–Meier OS curves stratified by the two clusters (P=0.0062). LSCC, lung squamous cell carcinoma; m7G, N7-methylguanosine; OS, overall survival.
Figure 4 Construction of the risk signature using the TCGA cohort of patients. (A) Univariate Cox regression analysis. The OS of patients with each m7G-related gene and the 5-gene signature was analyzed for their prognostic significance for LSCC patients. (B) The cross-validation using LASSO regression analysis. (C) Identification of the cut-off value of the risk score. LSCC patients were divided into high- or low-risk expression groups using the cut-off value of the median risk score. (D) PCA. Based on the expression levels of the five genes, each patient was accordingly divided into a high- or low-risk group. (E) Kaplan-Meier curves. The survival rates of the LSCC patients were analyzed by using the Kaplan-Meier survival curves and the log rank test stratified by the high- and low-risk scores. (F) ROC curves. The analysis was performed to demonstrate the predictive efficiency of the risk score for LSCC patients. PCA, principal component analysis; AUC, area under the curve; TCGA, The Cancer Genome Atlas; OS, overall survival; m7G, N7-methylguanosine; LSCC, lung squamous cell carcinoma; LASSO, the least absolute shrinkage and selection operator; ROC, receiver operating characteristic.
A high-risk score were enriched with altered gene pathways that were related to the immune response and inflammatory response.

To the best of our knowledge, this is one of only a few studies of m7G methylation in lung cancer available in the literature (23) that explores the usefulness of m7G methylation-related genes as a predictive marker in LSCC. In the current study, we analyzed 29 m7G methylation-related genes in LSCC tissues versus normal specimens and identified five m7G methylation-related genes as a risk model to estimate the survival of LSCC patients. Specifically, eukaryotic translation initiation factor 3 (EIF3) is necessary for the initiation of protein synthesis in cells and consists of subunits of EIF3A-M; if its subunits are altered, oncogene expression is upregulated, and tumor transformation occurs (39-41). For example, EIF3D, as the core subunit of EIF3, plays an oncogenic role in NSCLC (42), and the percentage of apoptotic cells in renal cell carcinoma cells was increased after knockdown of EIF3D expression (43). The U6 small nuclear RNA (snRNA)-associated Sm-like protein LSM1 functions to intimately connect with RNA processing and degradation (44-47). In addition, the downregulation of LSM1 expression is associated with breast cancer progression (48,49). Moreover, nuclear cap-binding proteins 2 (NCBP2) encodes a subunit of the nuclear cap-binding complex to directly contact the 5’-cap through the RNA recognition motif (50) by binding to the 7-methylguanosine cap.

**Figure 5** Association of five m7G methylation-related gene signature with the clinicopathological features and prognosis of LSCC patients. (A) Univariate regression analysis. The analysis was performed to associate the prognostic significance of the risk score and other clinicopathological parameters. (B) Multivariate regression analysis. The analysis was used to associate the prognostic significance of the risk score and other clinicopathological parameters. (C) The expression levels of these five genes and their correlation with the risk score and clinicopathological parameters were plotted into the heat map. The red bars indicate high expression, whereas the green bars refer low expression. m7G, N7-methylguanosine; LSCC, lung squamous cell carcinoma.
Figure 6 The signature of five m7G methylation-related genes as a prognostic model. (A) Bubble graph of the GO terms. A bigger bubble indicates that more genes are enriched. The increase in the blue color strength indicates a more obvious difference. The q-value is the adjusted P value. (B) Bar plot graph of the KEGG-enriched gene pathways. A longer bar indicates that more genes are enriched. The increase in the blue color intensity indicates that there is a more obvious difference. BP, biological process; CC, cell component; MF, molecular function; m7G, N\(^7\)-methylguanosine; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

Figure 7 Differential enrichments of 16 immune cell types and 13 immune-related pathways between the high- and low-risk LSCC patients from the TCGA dataset using the signature of five m7G-related genes. The LSCC patients were divided into a high- or low-risk LSCC cohort of the TCGA dataset using the signature of five m7G-related genes, and then the infiltrations of 16 immune cell types and activities of 13 immune-related pathways were compared between these two groups of patients. *, P<0.05; **, P<0.01; ***, P<0.001. aDCs, activated dendritic cell; NK, natural killer; pDCs, plasmacytoid dendritic cell; TIL, tumor infiltrating lymphocyte; APC, Antigen presenting cell; CCR, cytokine-cytokine receptor; HLA, human leukocyte antigen; MHC, major histocompatibility complex; IFN, interferon; LSCC, lung squamous cell carcinoma; TCGA, The Cancer Genome Atlas; m7G, N\(^7\)-methylguanosine.
and is then added to the emerging 5’-end of nascent RNA to protect against 5’–3’ exonucleolysis (51). A previous study has shown that NCBP2 can mediate gene interactions to modify neurodevelopmental defects of the 3q29 deletion (52). Additionally, nudix hydroxylases (NUDTs) belong to the versatile, widely distributed, housecleaning enzyme family to catalyze the hydrolysis of a wide range of nucleoside pyrophosphates linked to amino acids (53,54). Furthermore, previous researches have reported that NUDT10 expression is associated with malignant behaviors of gastric cancer by promoting tumor cell invasion and a poor prognosis (55,56), whereas NUDT11 has been shown to be associated with longer survival period in liver cancer (57) and bladder cancer (58).

In our current study, we revealed that the expression of these five m7G methylation-related genes was associated with a poor prognosis of LSCC patients. However, further investigation is needed to understand their role in LSCC development and progression.

As PPI measurements have increased, more and more PPI network-based protein function prediction methods have been proposed and are generally superior to the homology-based prediction methods (59). We analyzed the PPI among 29 m7G methylation-related genes, and found that LARP1, DCP2 and GEMIN5 have the strongest positive relevance (the lines among three genes are the reddest). Although the genes regulating the m7G process have been found involved in the carcinogenesis process, the potential modulation between tumor immunity and m7G methylation-related genes remains elusive. Based on the DEGs between different risk groups, we performed GO analyses and discovered that humoral immune responses were enriched. To further explore the correlation between the risk score and immune status, we profiling immune cell populations, immune-associated pathways, and the role of gene set signatures with ssGSEA. Interestingly, we found that the high-risk groups have higher fractions of immune activity. The reason maybe was excessive immune activation promotes immune invasion, and this phenomenon has been found in hepatocellular carcinoma (60,61).

Despite our current data being interesting and potentially useful for determining the prognosis of LSCC patients, this study does have its limitations. For example, we only analyzed the data from the public TCGA database for construction of the risk model for LSCC. The expression of these genes at the protein level needs to be determined to clarify their association with LSCC prognosis. In addition, our current study had no opportunity to validate the expression, functions, and mechanisms of these genes using our own cohort of tissue specimens and cell lines.

**Conclusions**

The data from the current study demonstrated that detection of the signature of five m7G methylation-related genes could be useful for predicting the prognosis of LSCC patients. Moreover, this five-gene signature is an independent prognostic predictor for the survival of LSCC patients. Our future study will provide validation of our current data using a collection of LSCC samples.

**Acknowledgments**

*Funding:* This study was supported in part by a grant from the Zhejiang Provincial Traditional Chinese Medicine Foundation (No. 2019ZA020).

**Footnote**

*Reporting Checklist:* The authors have completed the TRIPOD reporting checklist. Available at [https://jtd.amegroups.com/article/view/10.21037/jtd-23-1504/rc](https://jtd.amegroups.com/article/view/10.21037/jtd-23-1504/rc)


*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at [https://jtd.amegroups.com/article/view/10.21037/jtd-23-1504/coif](https://jtd.amegroups.com/article/view/10.21037/jtd-23-1504/coif)). 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. The 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


54. Malik A, Hept MA, Purcell EB. Sound the (Smaller)


<table>
<thead>
<tr>
<th>Gene</th>
<th>LogFC</th>
<th>Fold Change</th>
<th>False Discovery Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNTFR</td>
<td>4.33E-08</td>
<td>0.6506501</td>
<td>0.8266877</td>
</tr>
<tr>
<td>GMNC</td>
<td>0.6885246</td>
<td>1.1213881</td>
<td>2.6314422</td>
</tr>
<tr>
<td>LPA</td>
<td>0.0016579</td>
<td>0.0058169</td>
<td>1.0765475</td>
</tr>
<tr>
<td>CBLN2</td>
<td>4.82E-07</td>
<td>0.0279398</td>
<td>5.1670529</td>
</tr>
<tr>
<td>LILRA4</td>
<td>0.2955612</td>
<td>4.2454206</td>
<td>0.0449301</td>
</tr>
<tr>
<td>MSLNL</td>
<td>1.4621609</td>
<td>0.0711109</td>
<td>0.0330461</td>
</tr>
<tr>
<td>MSN</td>
<td>0.4474395</td>
<td>5.44E-08</td>
<td>7.8776831</td>
</tr>
<tr>
<td>LTA</td>
<td>-1.700126</td>
<td>0.1148553</td>
<td>2.3018394</td>
</tr>
<tr>
<td>GB1</td>
<td>2.1650636</td>
<td>0.028764</td>
<td>3.2148269</td>
</tr>
<tr>
<td>SCML4</td>
<td>0.8873543</td>
<td>0.018105</td>
<td>1.2397788</td>
</tr>
<tr>
<td>MMP28</td>
<td>6.72E-08</td>
<td>2.2119027</td>
<td>17.652384</td>
</tr>
<tr>
<td>KRT33A</td>
<td>2.05E-05</td>
<td>0.0162217</td>
<td>4.21E-06</td>
</tr>
<tr>
<td>SLC34A2</td>
<td>0.0001009</td>
<td>1.02683</td>
<td>0.8835729</td>
</tr>
<tr>
<td>PF4</td>
<td>0.0216537</td>
<td>7.11E-08</td>
<td>1.6654578</td>
</tr>
<tr>
<td>FAM129C</td>
<td>9.0565855</td>
<td>0.006716</td>
<td>1.44E-06</td>
</tr>
<tr>
<td>ANKRD63</td>
<td>0.744994</td>
<td>1.6147458</td>
<td>0.002366</td>
</tr>
<tr>
<td>FXYD2</td>
<td>-1.474245</td>
<td>1.0641945</td>
<td>3.80E-07</td>
</tr>
<tr>
<td>H2BFWT</td>
<td>18.10752</td>
<td>0.0001441</td>
<td>0.8933575</td>
</tr>
<tr>
<td>BEST2</td>
<td>2.51E-09</td>
<td>P</td>
<td>-1.272362</td>
</tr>
</tbody>
</table>

**Supplementary Table:**

The DEGs between the high- and low-risk groups.