RNA EDITING AND MODIFICATION IN DEVELOPMENT AND DISEASES

EDITED BY: Yanqiang Li, Jia Meng and Dongyu Zhao
PUBLISHED IN: Frontiers in Genetics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-83250-543-4
DOI 10.3389/978-2-83250-543-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

RNA EDITING AND MODIFICATION IN DEVELOPMENT AND DISEASES

Topic Editors:
Yanqiang Li, Boston Children's Hospital, Harvard Medical School, United States
Jia Meng, Xi'an Jiaotong-Liverpool University, China
Dongyu Zhao, Peking University, China

Citation: Li, Y., Meng, J., Zhao, D., eds. (2022). RNA Editing and Modification in Development and Diseases. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-543-4

Table of Contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 04 October 2022
DOI 10.3389/fgene.2022.1025445

OPEN ACCESS

EDITED BY
William C. Cho,
QEH, Hong Kong SAR, China

REVIEWED BY
Shizuka Uchida,
Aalborg University Copenhagen, Denmark

*CORRESPONDENCE
Yanqiang Li,
yanqiangli@childrens.harvard.edu
Jia Meng,
jiameng@xjtlu.edu.cn
Dongyu Zhao,
dongyuzhao@bjmu.edu.cn

SPECIALTY SECTION
This article was submitted to RNA, a section of the journal Frontiers in Genetics

RECEIVED 22 August 2022
ACCEPTED 12 September 2022
PUBLISHED 04 October 2022

CITATION
Li Y, Meng J and Zhao D (2022), Editorial: RNA editing and modification in development and diseases. *Front. Genet.* 13:1025445. doi: 10.3389/fgene.2022.1025445

Editorial: RNA editing and modification in development and diseases

Yanqiang Li[1,2]*, Jia Meng[3,4,5]* and Dongyu Zhao[6]*

[1]Basic and Translational Research Division, Department of Cardiology, Boston Children's Hospital, Boston, MA, United States, [2]Department of Pediatrics, Harvard Medical School, Boston, MA, United States, [3]Department of Biological Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China, [4]Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool, United Kingdom, [5]AI University Research Centre, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu, China, [6]Computational Biology and Bioinformatics at the School of Basic Medical Sciences, Peking University, Beijing, China

KEYWORDS

RNA modification, m6A, m5C, RNA binding protein, RNA editing

Editorial on the Research Topic
RNA editing and modification in development and diseases

RNA editing is a critical co-/post-transcriptional process perturbing the RNA in eukaryotic cells. RNA editing could occur in coding and non-coding regions, affecting the recoding of proteins, RNA splicing and stability. Adenosine to inosine (A-to-I) editing catalyzed by adenosine deaminases acting on RNA (ADAR) is the most dominant (~90%) RNA editing event in mammals (Athanasiadis et al., 2004). Despite extensive studies of RNA editing in human tissues, cancer development (Han et al., 2015) and the neural system (Hoopengardner et al., 2003), the study of RNA editing in other species apart from human and mouse were still limited. On the other side, the regulation of RNA editing in diseases such as cardiovascular, immune, and metabolism disorders needs to be investigated (Uchida and Jones, 2018). Besides RNA editing, general RNA modifications, including, m6A (Dominissini et al., 2012), pseudouridine (Carlile et al., 2014), m5C (Edelheit et al., 2013), m1A (Safra et al., 2017) and 2′-O-methylation (Dai et al., 2017; Elliott et al., 2019; Yi et al., 2021), have emerged as a critical layer for gene expression regulation, known as the epitranscriptome, attracting the best scientists globally.

To understand the role of m6A modification in lung cancer, Ma and Zhang conducted a consensus clustering analysis of 502 lung adenocarcinoma (LUAD) samples from TCGA based on the expression profiles of 20 m6A regulators. They found two m6A modification patterns with distinct overall survival (OS), activation of signaling pathways and tumor immunity. Furthermore, they identified 213 prognostic m6A-related genes, which were imported into LASSO-cox regression analysis. Next, they developed the m6A risk score and found that patients with low m6A risk score exhibited a prominent survival advantage in an independent dataset. Finally, they established a highly accurate nomogram containing independent prognostic

indicators. N6-methyladenosine (m6A) modification plays an important role in regulating the immunity microenvironment of breast cancer (BRCA). Zhang et al. established an m6A-related immune score (m6A-IS) to predict the immune infiltration and prognosis of BRCA accurately. The Low m6A-IS group is associated with enhanced antigen presentation and improved immune checkpoint expression, thus indicating sensitivity to immunotherapy. Furthermore, the m6A-IS can independently predict BRCA patients' response to immunotherapy.

Yang et al. attempted to map the m6A epitranscriptome of neuromyelitis optica spectrum disorder (NMOSD) patients compared to healthy controls. Towards this blood samples of NMOSD patients and healthy controls, the RNA were isolated and subjected to m6A-seq followed by bioinformatic analysis. Authors *via* bioinformatic analysis highlighted an extensive list of hyper and hypo m6A methylated transcripts and have attempted to connect it to NMOSD pathophysiology.

Ma et al. conducted an m6A-related lncRNA prognostic model using TCGA data. Twelve lncRNAs were included and validated in their new model. Also, the model was internally validated and showed good predictive values for ccRCC's survival.

Huang et al. studied breast cancer (BC)-specific m5C-related lncRNAs (m5C-lncRNAs) as potential biomarkers in breast cancer. They made a prognostic risk model, and analyzed the characteristics of tumor-infiltrating immune cells based on the subtypes of the risk model. The study used *in silico* data from TCGA database and for a validation experiment (sixteen pairs of fresh BC and paracarcinoma tissues) from their own patients. They found that these lncRNAs may serve as prognostic biomarkers in breasr cancer.

He et al. downloaded harmonized RNA-seq count data and clinical data for AML from several large study cohorts and analyzed the differential expression of a set of RNA binding proteins (RBPs) using the R package edgeR software together with statistical methods including univariate Cox regression analysis, LASSO-Cox regression analysis and multivariate Cox regression analysis. The authors established a prognostic model of 12-RBPs gene for AML and C-index and calibration diagrams were used to judge the accuracy of the model, and DCA was used to judge the net benefit. They have found that the net benefit and prediction accuracy of the prognostic model and the mixed model based on it was significantly higher than that of cytogenetics which was verified in one of the study cohorts where the data collected from and both of the selected gene set and the LASSO results have high credibility. By that, the authors concluded that their prognosis model of 12-RBPS gene is an optimized biomarker that can effectively stratify the risk of AML patients and the nomogram based on this prognostic model is a reliable method to predict the median survival time of patients.

Satir-Basaran et al. explored the role of ncRNA as vector for epigenetic inheritance *via* paternal germline. In their experimental set-up, the authors use four lines of mice (two susceptible to diet-induced obesity and diabetes and two resistant) and then perform phenotypic analyses in the males directly exposed to the diet, as well as in the following two generations of mice obtained *via* paternal transmission. To support a role for ncRNA, the authors rely on microinjection of RNA into fertilized oocytes. They used synthetic miR-19b-5p (found previously to induce obesity and diabetes), total sperm RNA, or RNA that was fractionated based on size (small, <200 nt and long, >200 nt) or subcellular location (DNA bound and free), which were obtained from males exposed to normal diet or diets with excess of fat. Overall, they conclude that ncRNA plays a role in the inter/trans-generational inheritance of the phenotypes and propose that the DNA-bound RNA may be a critical component.

Luo et al. investigated the expression of pyroptosis related genes in colon adenocarcinoma (COAD) and build a risk score model by utilizing the differential expressed genes. The model was found effective in predicting the overall survival and prognosis of COAD patients.

This review by Hao et al.gave an introduction about RNA editing, including RNA editing forms and organelles RNA editing occurred, then summarized the factors and mechanism involved in RNA editing in plant organelles, further reviewed RNA editing events identified in plant organelles through deep sequencing data, and finally discussed the functions of RNA editing in plant organelles. Pentatricopeptide repeat (PPR) proteins, one type of RNA-binding protein, are particularly prevalent in land plants. It has been revealed that PPR proteins are involved in the RNA processing of organellar genes and play a vital role in plant development and defense. In this manuscript, Qin et al. summarizes the recent progress in functional studies on PPR proteins, including plant fertility, chloroplast biogenesis, embryogenesis, stress responses, and plant development.

Li et al. performed univariate analysis to identify the prognosis-related RNA binding proteins (RBPs). They identified 11 RBPs as prognosis markers of HBV-related hepatocellular carcinoma (HCC). In addition, the authors claimed that their model has better predictive efficacy than the models using other clinical parameters.

Although lot of studies of RNA modification and editing in the development and diseases, there are several new aspects need further exploration. Because there are more than 140 kind of RNA modifications in RNA, new methods to detect novel RNA modification systemically are needed to develop. E.g Deep learning based method (Song et al., 2021) and new non-antibody technology will be good to find the RNA modification at base-resolution (Hu et al., 2022); On the otherside, with development of single cell technology, study RNA modification/editing at the single cell levels or even spatial levels will be helpful to understand the cellular development with a high resolution (Sapiro et al., 2019); In addition, intergration of the GWAS data of specific disease to see

the mutation the RNA binding proteins related to the RNA modification will be helpful to reveal the role of RNA modifications in disease (Nachmani et al., 2019).

To summarize, our unique topic covers RNA modifications such as m6A and m5C in disease, the role of RNA binding proteins and non-coding RNAs, and the editing progress of RNA editing in plants. These research studies are complemented by review articles that summarize our current understanding of RNA modifications in human disease and development. We hope our Research Topic will benefit the community of RNA modifications.

Author contributions

YL prepared the manuscript with extensive help from JM and DZ. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Athanasiadis, A., Rich, A., and Maas, S. (2004). Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol.* 2, e391. doi:10.1371/journal.pbio.0020391

Carlile, T. M., Rojas-Duran, M. F., Zinshteyn, B., Shin, H., Bartoli, K. M., and Gilbert, W. V. (2014). Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* 515, 143–146. doi:10.1038/nature13802

Dai, Q., Moshitch-Moshkovitz, S., Han, D., Kol, N., Amariglio, N., Rechavi, G., et al. (2017). Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat. Methods* 14, 695–698. doi:10.1038/nmeth.4294

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206. doi:10.1038/nature11112

Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O., and Sorek, R. (2013). Transcriptome-wide mapping of 5-methylcytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet.* 9, e1003602. doi:10.1371/journal.pgen.1003602

Elliott, B. A., Ho, H. T., Ranganathan, S. V., Vangaveti, S., Ilkayeva, O., Abou Assi, H., et al. (2019). Modification of messenger RNA by 2'-O-methylation regulates gene expression *in vivo*. *Nat. Commun.* 10, 3401. doi:10.1038/s41467-019-11375-7

Han, L., Diao, L., Yu, S., Xu, X., Li, J., Zhang, R., et al. (2015). The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell.* 28, 515–528. doi:10.1016/j.ccell.2015.08.013

Hoopengardner, B., Bhalla, T., Staber, C., and Reenan, R. (2003). Nervous system targets of RNA editing identified by comparative genomics. *Science* 301, 832–836. doi:10.1126/science.1086763

Hu, L., Liu, S., Peng, Y., Ge, R., Su, R., Senevirathne, C., et al. (2022). m(6)A RNA modifications are measured at single-base resolution across the mammalian transcriptome. *Nat. Biotechnol.* 40, 1210–1219. doi:10.1038/s41587-022-01243-z

Nachmani, D., Bothmer, A. H., Grisendi, S., Mele, A., Bothmer, D., Lee, J. D., et al. (2019). Germline NPM1 mutations lead to altered rRNA 2'-O-methylation and cause dyskeratosis congenita. *Nat. Genet.* 51, 1518–1529. doi:10.1038/s41588-019-0502-z

Safra, M., Sas-Chen, A., Nir, R., Winkler, R., Nachshon, A., Bar-Yaacov, D., et al. (2017). The m1A landscape on cytosolic and mitochondrial mRNA at single-base resolution. *Nature* 551, 251–255. doi:10.1038/nature24456

Sapiro, A. L., Shmueli, A., Henry, G. L., Li, Q., Shalit, T., Yaron, O., et al. (2019). Illuminating spatial A-to-I RNA editing signatures within the Drosophila brain. *Proc. Natl. Acad. Sci. U. S. A.* 116, 2318–2327. doi:10.1073/pnas.1811768116

Song, Z., Huang, D., Song, B., Chen, K., Song, Y., Liu, G., et al. (2021). Attention-based multi-label neural networks for integrated prediction and interpretation of twelve widely occurring RNA modifications. *Nat. Commun.* 12, 4011. doi:10.1038/s41467-021-24313-3

Uchida, S., and Jones, S. P. (2018). RNA editing: Unexplored opportunities in the cardiovascular system. *Circ. Res.* 122, 399–401. doi:10.1161/CIRCRESAHA.117.312512

Yi, Y., Li, Y., Meng, Q., Li, Q., Li, F., Lu, B., et al. (2021). A PRC2-independent function for EZH2 in regulating rRNA 2'-O methylation and IRES-dependent translation. *Nat. Cell. Biol.* 23, 341–354. doi:10.1038/s41556-021-00653-6

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 August 2021
doi: 10.3389/fgene.2021.707305

Systematic Analysis Identifies a Specific RNA-Binding Protein-Related Gene Model for Prognostication and Risk-Adjustment in HBV-Related Hepatocellular Carcinoma

*Maoshi Li[1,2], Zhongwei Liu[1,2], Jing Wang[1,2], Huimin Liu[1,2], Hongmei Gong[1,2], Shilian Li[1,2], Ming Jia[1,2] and Qing Mao[1,2]**

[1] Department of Infectious Diseases, Southwest Hospital, Army Medical University, Chongqing, China, [2] Chongqing Key Laboratory for Research of Infectious Diseases, Chongqing, China

OPEN ACCESS

Edited by:
Yanqiang Li,
Boston Children's Hospital, Harvard Medical School, United States

Reviewed by:
Guangyu Wang,
Houston Methodist Research Institute, United States
Xing Niu,
Shengjing Hospital of China Medical University, China

***Correspondence:**
Qing Mao
qingmao@tmmu.edu.cn

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 09 May 2021
Accepted: 28 June 2021
Published: 04 August 2021

Citation:
Li M, Liu Z, Wang J, Liu H, Gong H, Li S, Jia M and Mao Q (2021) Systematic Analysis Identifies a Specific RNA-Binding Protein-Related Gene Model for Prognostication and Risk-Adjustment in HBV-Related Hepatocellular Carcinoma. Front. Genet. 12:707305. doi: 10.3389/fgene.2021.707305

Objective: Increasing evidence shows that dysregulated RNA binding proteins (RBPs) modulate the progression of several malignancies. Nevertheless, their clinical implications of RBPs in HBV-related hepatocellular carcinoma (HCC) remain largely undefined. Here, this study systematically analyzed the associations of RBPs with HBV-related HCC prognosis.

Methods: Based on differentially expressed RBPs between HBV-related HCC and control specimens, prognosis-related RBPs were screened by univariate analyses. A LASSO model was then created. Kaplan-Meier curves, ROCs, multivariate analyses, subgroup analyses and external verification were separately applied to assess the efficacy of this model in predicting prognosis and recurrence of patients. A nomogram was created by incorporating the model and clinical indicators, which was verified by ROCs, calibration curves and decision curve analyses. By CIBERSORT algorithm, the association between the risk score and immune cell infiltrations was evaluated.

Results: Totally, 54 RBPs were distinctly correlated to prognosis of HBV-related HCC. An 11-RBP model was created, containing POLR2L, MRPS12, DYNLL1, ZFP36, PPIH, RARS, SRP14, DDX41, EIF2B4, and NOL12. This risk score sensitively and accurately predicted one-, three- and five-year overall survival, disease-free survival, and progression-free interval. Compared to other clinical parameters, this risk score had the best predictive efficacy. Also, the clinical generalizability of the model was externally verified in the GSE14520 dataset. The nomogram may predict patients' survival probabilities. Also, the risk score was related to the components in the immune microenvironment.

Conclusion: Collectively, RBPs may act as critical elements in the malignant progression of HBV-related HCC and possess potential implications on prognostication and therapy decision.

Keywords: hepatocellular carcinoma, HBV, RNA-binding protein, prognosis, signature, recurrence

INTRODUCTION

Hepatocellular carcinoma (HCC) is the most prevalent type of liver cancer and represents a common malignant neoplasm globally (Liu et al., 2019). Because of the high risk of recurrence and metastasis, the 5-year survival probabilities of advanced HCC are still undesirable. The epidemiology of HCC is affected by underlying liver diseases especially hepatitis B virus (HBV) (Chabrolles et al., 2020). It has been estimated that HBV infection is responsible for 50% of HCC worldwide (Li et al., 2019). HBV that integrates into cancer-relevant genes may drive hepatocarcinogenesis (Nakagawa et al., 2019). Nevertheless, the mechanism by which HBV infection contributes to HCC remain deficiently expounded (Torresi et al., 2019). The extensive decrease in HCC cases demands a broader range of universal HBV vaccination application and efficient therapy of HBV-relevant chronic hepatitis, which has a long way to go (Sagnelli et al., 2020).

RNA-binding protein (RBP) represents a critical mediator of cancer phenotype (Müller-McNicoll and Neugebauer, 2013). RBP acts dynamically and multifunctionally on multiple levels of post-transcriptional gene expression, such as mRNA splicing, stability and translation (Hou et al., 2019). Over 1,500 human RBPs have been discovered, which possess 600 structure-specific RNA-binding domains (Schneider et al., 2019). Most of them are characterized by evolutionary conservatism and ubiquitous expression to maintain cellular homeostasis (Chabrolles et al., 2020). Genetic and proteomic data highlight that alterations in RBP expression display profound implications on HCC (Lin et al., 2019). For instance, RBP YTHDF2 facilitates cancer stem cell phenotype and metastasis in HCC through regulating OCT4 N^6-methyladenosine methylation (Zhang et al., 2020). RBP RPS3 leads to hepatocarcinogenesis through up-regulating SIRT1 at a post-transcriptional level (Zhao et al., 2019). RBP RBM3 induces proliferation of HCC cells *via* regulating circRNA-SCD production (Dong et al., 2019). Despite this, the functions of most RBPs in HBV-related HCC remain still unclear.

This study systematically dissected the prognosis-related RBPs of HBV-related HCC, and established and externally verified an RBP model for predicting prognosis and recurrence by applying least absolute shrinkage selection operator (LASSO). This model might serve as a potential prognostic stratification tool and offer several therapeutic targets for HBV-related HCC.

MATERIALS AND METHODS

Data Acquisition

RNA-seq transcriptome data and complete clinical information of 374 HCC and 50 control specimens were retrieved from the Cancer Genome Atlas (TCGA) database[1]. Among them, 108 HBV-related HCC were extracted from TCGA dataset, which were employed as the discovery set. Expression profiling and clinical features of 224 HBV-related HCC samples were obtained from the GSE14520 dataset in the Gene Expression Omnibus (GEO; [2]) repository. This GSE14520 dataset was based on the GPL571 and GPL3921 platforms. This dataset was used as the validation set. **Table 1** listed the clinical characteristics of HBV-related HCC patients. Each probe was transformed to the corresponding gene symbol. If multiple probes matched the same gene symbol, the average value was determined as the expression value of this gene. Based on published articles, a list of 1,542 RBPs was collected (**Supplementary Table 1**).

Differential Expression Analysis

Differences in expression levels of RBPs were analyzed between 108 HBV-related HCC and 50 control samples *via* limma package based on RNA-seq transcriptome data (Ritchie et al., 2015). Under the threshold of $|\text{fold-change}| > 2$ and adjusted $p < 0.001$, up- and down-regulated RBPs were screened for HBV-related HCC.

Functional Annotation Analysis

Gene ontology (GO) enrichment analysis primarily contains three categories: biological process (BP), cellular component (CC) and molecular function (MF). Furthermore, Kyoto Encyclopedia of Genes and Genomes (KEGG) can provide signaling pathways involved in RBPs. Here, GO, and KEGG enrichment analyses of differentially expressed RBPs were carried out *via* clusterProfiler package (Ashburner et al., 2000). Terms with false discovery rate (FDR) < 0.05 were significantly enriched by above RBPs. Functional association between HBV-related RBPs was analyzed through the STRING database[3] (Szklarczyk et al., 2017). A protein-protein interaction (PPI) network was conducted utilizing Cytoscape software (Doncheva et al., 2019).

Determining Candidate Prognosis-Related RBPs

Univariate Cox regression analysis was applied for analyzing the associations between overall survival (OS) of HBV-related HCC patients and differentially expressed RBPs utilizing survival package. Prognosis-related RBPs were determined with the threshold of $p < 0.05$. Then, key prognosis-related RBPs were screened through LASSO regression analysis with glmnet package (Engebretsen and Bohlin, 2019). Normalized regression coefficients of key prognosis-related RBPs were calculated based on multivariate regression analysis.

Establishment of an RBP-Related Prognostic Model

A risk score was established on the basis of the key prognosis-related RBPs for HBV-related HCC patients in the TCGA dataset.

Abbreviations: HCC, hepatocellular carcinoma; HBV, hepatitis B virus; RBP, RNA-binding protein; LASSO, least absolute shrinkage selection operator; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes; FDR, false discovery rate; OS, overall survival; DFS, disease-free survival; PFI, progression-free interval; ROC, receiver operating characteristic; AUC, area under the curve; DCA, decision curve analysis; GSEA, Gene Set Enrichment Analyses.

[1] https://portal.gdc.cancer.gov/

[2] https://www.ncbi.nlm.nih.gov/gds/

[3] http://string-db.org/

TABLE 1 | Clinical characteristics of HBV-related HCC patients.

Characteristics	TCGA (*n* = 108)	GSE14520 (*n* = 224)
Age		
<65	81	199
≥65	27	25
Gender		
Male	89	195
Female	19	29
Survival status		
Alive	20	86
Dead	88	138
TNM stage		
Stage I	72	96
Stage II	21	78
Stage III	11	50
Stage IV	2	0
Recurred status		
Yes	45	125
No	56	99

This study calculated the risk score of each patient, according to the formula: risk score = expression of RBP1 × regression coefficient of RBP1 + expression of RBP2 × regression coefficient of RBP2 + . . . + expression of RBPn × regression coefficient of RBPn. According to the median risk score, HBV-related HCC patients were assigned into high- and low-risk groups. Hierarchical clustering analysis was applied for showing the associations between expression patterns of above RBPs and clinical characteristics (stage, grade, gender, and age). Kaplan-Meier curve analyses were presented for investigating the differences in OS, disease-free survival (DFS) and progression-free interval (PFI) between two groups using Survival package. P values were determined with log-rank tests. Time-dependent receiver operating characteristic (ROC) curves under one-, three- and five-year OS, DFS, and PFI were generated by SurvivalROC package (Heagerty et al., 2000). Then, the area under the curve (AUC) was calculated to evaluate the predictive usefulness of the risk score.

Prognostic Model Verification

The associations between risk score, age, gender, grade, stage, and prognosis were evaluated utilizing univariate cox regression analysis. Prognostic factors with $p < 0.05$ were incorporated for multivariate cox regression analysis. Independent prognostic factors were then identified for HBV-related HCC. Time-independent ROCs of risk score, age, gender, grade, and stage were separately constructed and the predictive power was compared. In published literature, Fang and Chen (2020) proposed a two-m^{6}A RNA methylation regulator prognostic model (HNRNPA2B1 and RBM15) for HBV-related HCC prognosis. By ROCs, the predictive efficacy was compared with our prognostic model. For evaluating the clinical generalizability of the model, this model was verified in the GSE14520 dataset. With the same formula, the risk scores of patients were calculated in this dataset. Based on the median risk score, OS differences between high- and low-risk groups were assessed by Kaplan-Meier curves and log-rank tests. The predictive performance was verified by ROCs.

Subgroup Analysis

HBV-related HCC subjects in TCGA dataset were stratified into subgroups according to clinical characteristics, as follows: age ≥ 65 and < 65 subgroups, female and male subgroups, grade 1–2 and 3–4 subgroups and stage I-II and stage III-IV subgroups. In each subgroup, Kaplan-Meier curves of OS were conducted between high- and low-risk patients. Survival differences were estimated with log-rank tests.

Nomogram Construction

A nomogram by incorporating gender, age, grade, stage, and risk score was created for predicting one-, three-, and five-year survival probabilities of HBV-related HCC patients in the TCGA dataset. The accuracy in predicting prognosis was evaluated by ROC curves, calibration curves and decision curve analysis (DCA) (Vickers and Elkin, 2006). ROC curves were conducted for evaluating the predictive performance of this nomogram for one-, three- and five-year OS. By calibration curves, discrepancy between nomogram-estimated and actual one-, three-, and five-year survival duration was analyzed. DCA was applied for quantifying the clinical practical use with survival outcomes of a decision considered.

Gene Set Enrichment Analyses (GSEA)

HBV-related HCC specimens were stratified into high- and low-risk groups. By Gene Set Enrichment Analyses (GSEA)[4], potential mechanisms of this prognosis-related model were elucidated (Subramanian et al., 2005). GSEA was carried out for finding enriched KEGG pathways, with KEGG gene set "c2.cp.kegg.v7.0.symbols.gmt" as a reference. Pathways with nominal $p < 0.05$ were distinctly enriched.

Estimation of Immune Cell Infiltrations and HLA Expression

By applying CIBERSORT algorithm[5], the proportions of 22 immune cells in HBV-related HCC and control specimens were quantified based on their expression profiling (Newman et al., 2015). The proportions of all immune cells in each specimen were equal to 1. The LM22 signature was employed as a reference set. The permutations were set as 1,000. Samples with $p < 0.05$ were screened for further analyses. The correlations between risk score and infiltrations of immune cells and HLA family expression were assessed with Spearson correlation analysis.

[4]http://software.broadinstitute.org/gsea/index.jsp
[5]http://cibersort.stanford.edu/

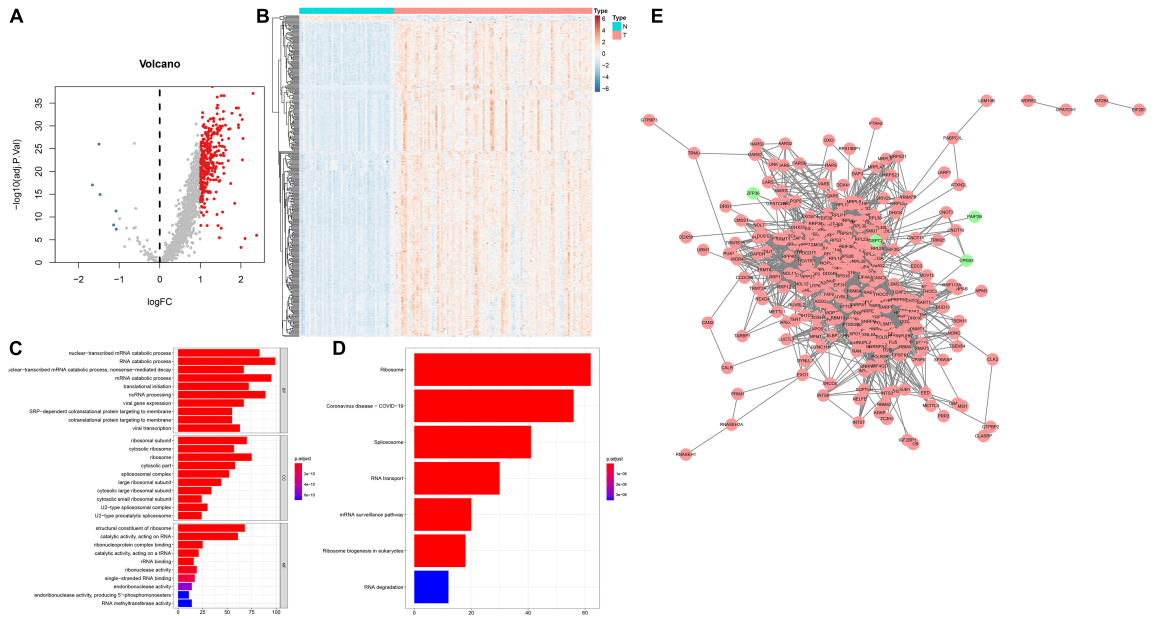

FIGURE 1 | Expression patterns and functions of RBPs in HBV-related HCC tissues. **(A)** Volcano plots of up- and down-regulated RBPs in HBV-related HCC than control specimens. Red dots represented up-regulated RBPs; blue dots represented down-regulated RBPs; gray dots represented no differentially expressed RBPs. **(B)** Heatmap visualizing expression values of up- (red) and down-regulated (blue) RBPs in HBV-related HCC and control specimens. **(C)** The first ten biological processes (BPs), cellular components (CCs) and molecular functions (MFs) involved in differentially expressed RBPs. **(D)** KEGG pathways enriched by above RBPs. The longer the band, the greater the number of enriched genes. The more the color tended to red, the smaller adjusted p-value. **(E)** A PPI network based on the differentially expressed RBPs. Red indicated up-regulated RBPs and green indicated down-regulated RBPs.

TABLE 2 | The first 20 up-regulated RBPs and 6 down-regulated RBPs in HBV-related HCC.

ID	logfold-change	Average expression	t	P	Adjusted p	B
PEG10	2.387979	2.638582	5.173505	6.75E-07	9.80E-07	4.947393
RNASEH2A	2.299935	4.022123	18.10863	1.07E-40	7.58E-38	82.02684
EEF1A2	2.10402	2.672998	4.815884	3.35E-06	4.71E-06	3.396926
TARBP1	2.052569	3.046862	13.85154	3.05E-29	6.00E-28	55.80843
PABPC1L	2.039105	3.485111	12.0097	3.86E-24	2.85E-23	44.13127
EZH2	2.031468	2.209527	15.56872	6.04E-34	3.73E-32	66.57365
RPS4Y1	2.002975	5.389658	3.65882	0.000343	0.000425	−1.01902
DDX39A	1.983421	5.158247	15.98784	4.45E-35	4.21E-33	69.16769
SMG5	1.942159	5.259853	14.82399	6.46E-32	2.41E-30	61.92894
BOP1	1.938054	5.235709	15.10358	1.11E-32	5.44E-31	63.6776
NELFE	1.937597	5.596385	16.63109	8.36E-37	1.19E-34	73.1174
SNRPE	1.883775	6.305086	16.68019	6.18E-37	1.19E-34	73.41722
MSI1	1.847698	1.586866	8.299861	3.97E-14	9.03E-14	21.2786
SNRPB	1.780655	7.311245	16.16782	1.46E-35	1.59E-33	70.27684
RPL22L1	1.747909	4.488432	10.002	1.26E-18	4.14E-18	31.52673
SPATS2	1.733092	2.806658	14.71432	1.29E-31	4.26E-30	61.24148
SF3B4	1.698588	5.667539	14.66409	1.77E-31	5.46E-30	60.92638
RBM3	1.697054	6.272179	13.86538	2.79E-29	5.58E-28	55.89584
PRIM1	1.688578	3.095013	11.94612	5.79E-24	4.17E-23	43.72825
EXO1	1.626191	1.287369	12.87215	1.57E-26	1.87E-25	49.60374
GSPT2	−1.0703	2.210642	−5.84768	2.69E-08	4.25E-08	8.079581
PAIP2B	−1.0761	2.263303	−7.60705	2.20E-12	4.47E-12	17.3157
ANG	−1.13722	9.092962	−6.28555	2.93E-09	4.93E-09	10.24674
AZGP1	−1.469	9.486897	−9.06554	4.04E-16	1.08E-15	25.81354
CPEB3	−1.4956	2.109004	−13.3884	5.81E-28	9.16E-27	52.87762
ZFP36	−1.65925	7.068678	−9.88183	2.67E-18	8.48E-18	30.78536

TABLE 3 | Prognosis-related RBPs in HBV-related HCC patients by univariate cox regression analysis.

ID	HR	HR.95L	HR.95H	P	ID	HR	HR.95L	HR.95H	P
DDX41	4.4441	1.4451	13.667	0.0093	UTP6	2.6888	1.1424	6.3284	0.0235
NSUN5	2.6559	1.0087	6.9926	0.048	GPATCH4	2.1814	1.1159	4.2646	0.0226
ILF2	1.9392	1.0166	3.6988	0.0444	MRPS12	2.5059	1.3783	4.5562	0.0026
SNRPA	2.4521	1.1227	5.3556	0.0244	RNF113A	2.0476	1.052	3.9855	0.0349
SRP14	0.3565	0.132	0.963	0.0419	CNOT10	3.1297	1.5282	6.4095	0.0018
F3	3.027	1.2561	7.2945	0.0136	PHF5A	2.7005	1.1897	6.1301	0.0175
TCOF1	2.7229	1.3076	5.6701	0.0074	FTSJ1	2.3828	1.1602	4.894	0.0181
METTL5	2.3697	1.0282	5.4613	0.0428	NOP2	2.4581	1.0643	5.677	0.0352
DYNLL1	3.4144	1.3015	8.9577	0.0126	PPIH	2.6265	1.4203	4.8573	0.0021
NHP2	3.0865	1.3129	7.2564	0.0098	NPM1	2.1356	1.0892	4.1871	0.0272
RRP36	2.3214	1.0739	5.0183	0.0323	METTL1	1.9349	1.0773	3.4753	0.0272
PA2G4	2.8337	1.1739	6.8404	0.0205	PES1	2.1902	1.0737	4.4676	0.0311
RUVBL2	3.0262	1.3085	6.9991	0.0096	RNASEH1	3.0569	1.2504	7.4733	0.0143
POP4	2.1925	1.139	4.2203	0.0188	BUD13	2.2305	1.0776	4.6167	0.0307
EIF3B	2.4609	1.2188	4.9692	0.012	TAF9	2.3771	1.2161	4.6465	0.0113
THOC3	2.8281	1.337	5.9822	0.0065	HNRNPAB	2.8933	1.1947	7.0072	0.0186
HARS2	4.8941	2.1092	11.356	0.0002	RPP40	2.0237	1.0355	3.9551	0.0392
CD2BP2	2.5789	1.0669	6.2337	0.0354	MOV10	3.8285	1.6909	8.6686	0.0013
EIF2B4	4.2515	1.8898	9.5642	0.0005	MRPL33	2.385	1.2963	4.3881	0.0052
LARS	2.3723	1.0776	5.2223	0.0319	TRMT6	2.1637	1.0222	4.5801	0.0437
NOL12	1.9271	1.0658	3.4846	0.03	DKC1	2.1134	1.0423	4.285	0.038
NOP56	2.3817	1.317	4.3069	0.0041	EIF4A3	2.196	1.0119	4.7655	0.0466
WDR4	2.0355	1.0801	3.836	0.0279	POLR2L	2.5862	1.4397	4.6458	0.0015
PTGES3	2.9652	1.1175	7.8681	0.029	BRIX1	2.0967	1.1064	3.9735	0.0232
RARS	8.3614	2.9287	23.872	< 0.0001	ZFP36	1.6328	1.0219	2.6087	0.0403
GEMIN7	1.9608	1.0977	3.5025	0.0229	GAPDH	2.2464	1.2433	4.0587	0.0073
EIF3K	2.0916	1.1255	3.8871	0.0196	RBM24	1.3877	1.0278	1.8737	0.0324

RESULTS

Expression and Functions of RBPs in HBV-Related HCC

This study collected 1,542 RBPs and analyzed their expression in 108 HBV-related HCC and 50 control tissues. | Fold-change| > 2 and adjusted $p < 0.001$ were set as the screening thresholds. As a result, 340 RBPs were up-regulated in HBV-related HCC than control specimens (**Figures 1A,B**). The first 20 up-regulated RBPs were shown in **Table 2**. Meanwhile, there were six down-regulated RBPs (GSPT2, PAIP2B, ANG, AZGP1, CPEB3, and ZFP36) in HBV-related HCC compared to controls (**Figures 1A,B** and **Table 2**). These RBPs were primarily enriched in mRNA metabolic biological processes such as nuclear-transcribed mRNA catabolic process, RNA catabolic process, nuclear-transcribed mRNA catabolic process, mRNA catabolic process and ncRNA processing (**Figure 1C**). Our KEGG analysis demonstrated that above RBPs were distinctly related to key pathways including ribosome, spliceosome, RNA transport, mRNA surveillance pathway, ribosome biogenesis in eukaryotes and RNA degradation, confirming their roles on post-transcriptional gene regulation (**Figure 1D**). A PPI network was conducted for revealing the interactions between these HCC-related RBPs (**Figure 1E**).

Establishing a RBP Model for Predicting HBV-Related HCC Patients' Prognosis and Recurrence

By univariate analyses, we investigated which dysregulated RBPs were in relation to HBV-related HCC patients' survival outcomes. With the cutoff of $p < 0.05$, 54 RBPs were distinctly correlated to prognosis (**Table 3**). These prognosis-related RBPs were utilized for LASSO analysis. As a result, 10 key RBPs were screened for constructing a prognostic model (**Figures 2A,B**). The regression coefficients of above RBPs were as follows: POLR2L = 0.453206443450835; MRP S12 = 0.00334358058970182; DYNLL1 = 0.0556454917799665; ZFP36 = 0.215686692231002; PPIH = 0.300574508736105; RARS = 1.0223374037773; SRP14 = −1.08473362012663; DDX 41 = 0.00122492284215762; EIF2B4 = 0.348820298750318; NOL12 = 0.155516694418799. The risk score of each subject from TCGA dataset was determined according to expressions and regression coefficients of RBPs. Heatmap depicted the correlations between expression patterns of POLR2L, MRPS12, DYNLL1, ZFP36, PPIH, RARS, SRP14, DDX41, EIF2B4, and NOL12 and clinical characteristics of HBV-related HCC patients (**Figure 2C**). Patients from the TCGA dataset were assigned into high- and low-risk groups. As a result, high-risk subjects were indicative of unfavorable OS (p = 3.734e-06; **Figure 2D**),

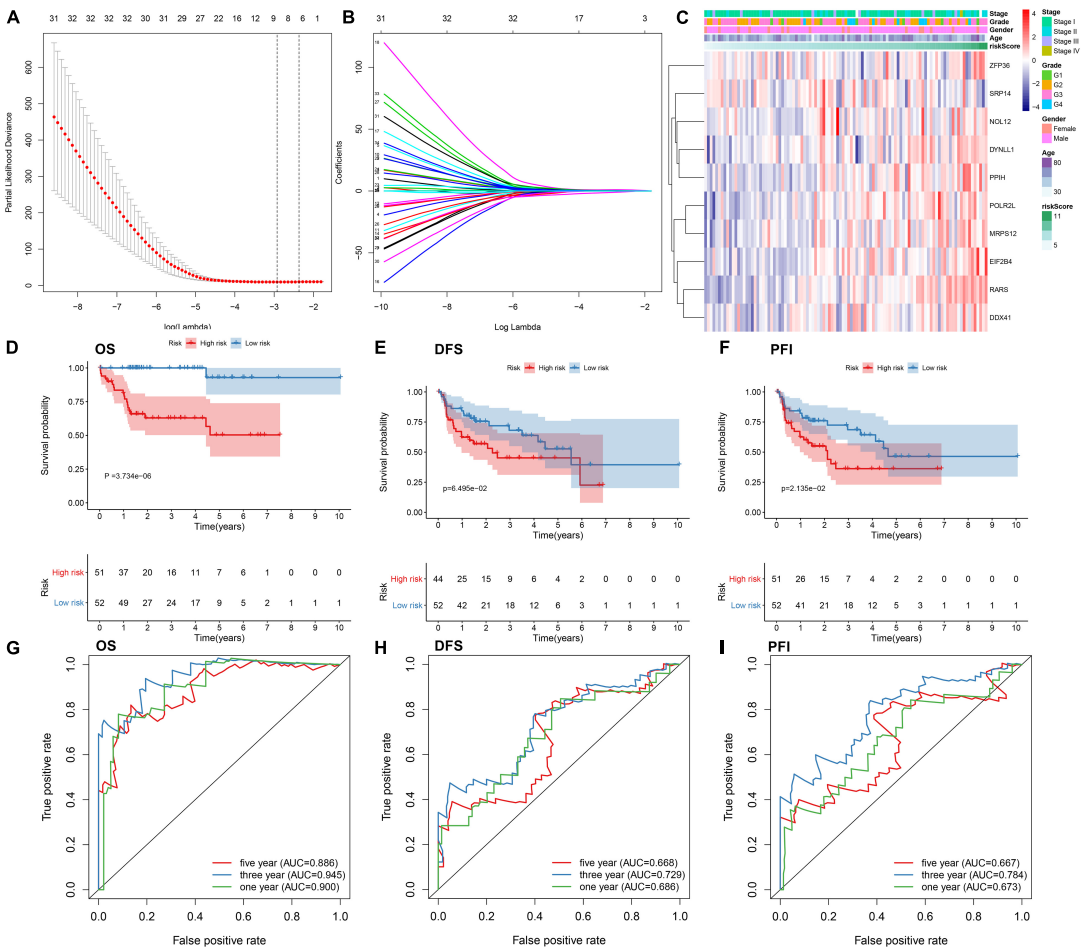

FIGURE 2 | Establishing an RBP prognostic model for HBV-related HCC patients in the TCGA dataset. **(A)** Relationships of λ and partial likelihood deviance. The vertical slash represents the optimal λ value. **(B)** LASSO regression coefficients of RBPs in this model. **(C)** Heatmap for the correlations between expression values of RBPs and clinical features. Kaplan-Meier curves of **(D)** OS, **(E)** DFS and **(F)** PFI between high- and low-risk patients. The ROCs of **(G)** OS, **(H)** DFS, and **(I)** PFI based on the risk score.

DFS (p = 6.495e-02; **Figure 2E**) and PFI (p = 2.135e-02; **Figure 2F**). The ROCs were plotted for evaluation of the predictive performance. The AUCs under one-, three-, and five-year OS were separately 0.900, 0.945 and 0.886, demonstrating that this model could be precisely predictive of OS probabilities (**Figure 2G**). Meanwhile, the AUCs under one-, three-, and five-year DFS were 0.686, 0.729 and 0.668 (**Figure 2H**) as well as the AUCs under one-, three-, and five-year PFI (**Figure 2I**) were 0.673, 0.784, and 0.667, which were indicative that this model possessed the well performance on predicting HBV-related HCC recurrence and progression.

The RBP Model Displays Independent and Well Predictive Power for HBV-Related HCC Patients

By univariate analyses, we investigated the associations between survival outcomes and risk score and clinical parameters in the TCGA dataset. In **Figure 3A**, risk score (p < 0.001) and stage (p = 0.009) were both risk factors of HBV-related HCC. Following multivariate analyses, the risk score was independently predictive of survival outcomes (p < 0.00l; **Figure 3B**). In comparison to other clinical parameters, this risk score exhibited the highest AUC of OS (0.940). This demonstrated that the risk score possessed more excellent predictive performance than other parameters (**Figure 3C**). Compared to the published prognostic model (HNRNPA2B1 and RBM15), higher AUC value was investigated in our model (**Figure 3D**). We further evaluated the clinical generalizability of this model. In the GSE14520 dataset, high-risk scores were also indicative of unfavorable survival outcomes (p = 2.999e-03; **Figure 3E**) and the AUC value was 0.656 (**Figure 3F**). Collectively, this model displayed the independent and well power in predicting prognosis.

Subgroup Analysis of This Prognostic RBP Model in HBV-Related HCC

For investigating the predictive sensitivity of this model in HBV-related HCC prognosis, survival analyses were carried out in

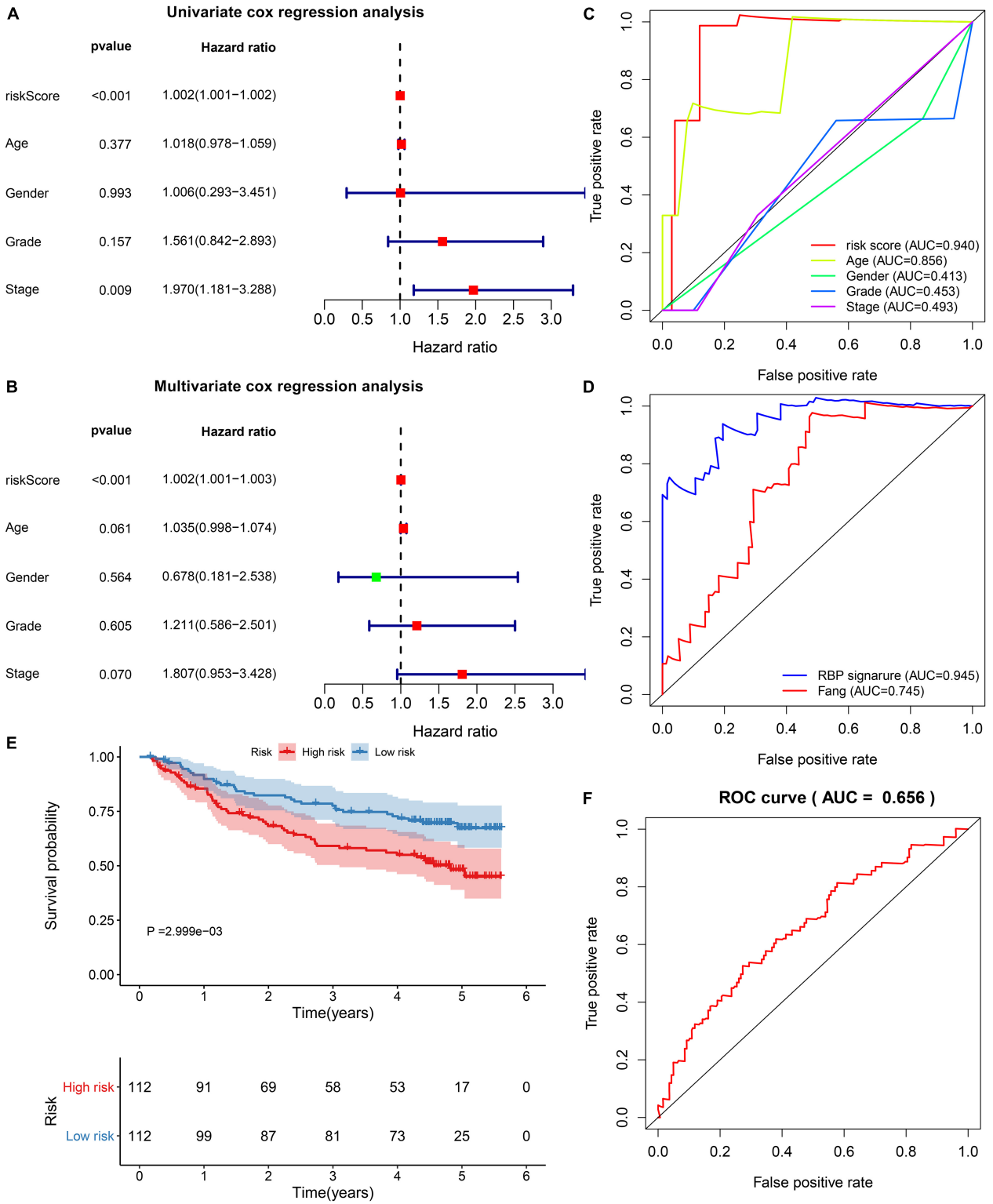

FIGURE 3 | Verification of the independency and accuracy of the prognostic model. **(A)** Univariate and **(B)** multivariate cox regression analyses of risk score, age, gender, grade, and stage with HBV-related HCC patients in the TCGA dataset. **(C)** The ROCs for comparing the AUCs of the risk score with age, gender, grade, and stage. **(D)** The ROCs for comparing the AUCs of the risk score with the model constructed by Fang et al. **(E)** Survival differences between high- and low-risk patients in the GSE14520 dataset. P value was estimated with log-rank test. **(F)** The ROC for evaluation of the predictive performance of the risk score.

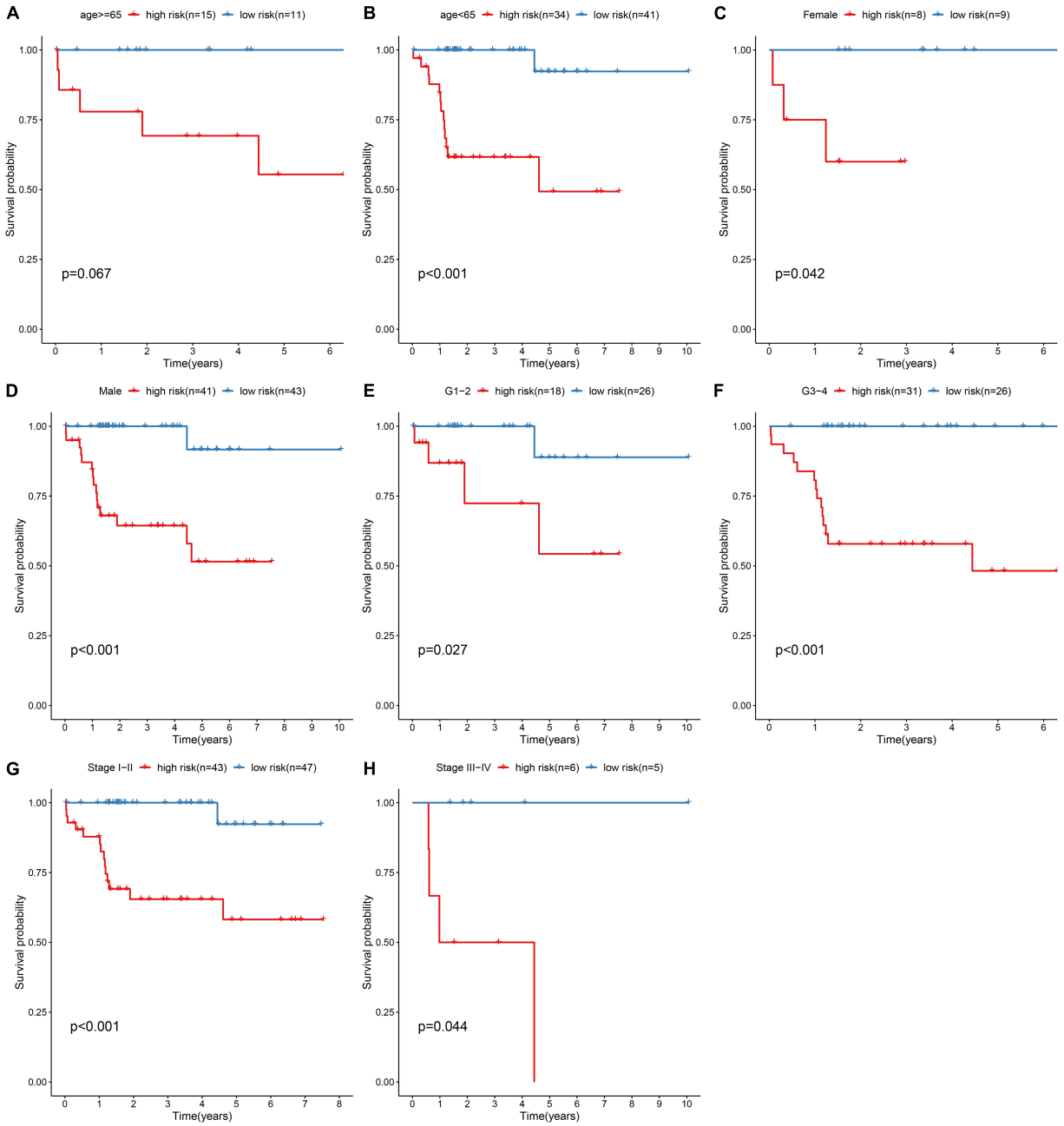

FIGURE 4 | Subgroup analysis for assessing the sensibility of the risk score to predict HBV-related HCC patients' prognosis. Kaplan-Meier curves for survival outcomes of high- and low-risk patients in **(A)** age ≥ 65 and **(B)** < 65 subgroups, **(C)** female and **(D)** male subgroups, **(E)** grade 1–2 (G1-2) and **(F)** grade 3–4 (G3-4) subgroups, **(G)** stage I–II and **(H)** stage III–IV subgroups. *P* values were estimated with log-rank tests.

different subgroups. Our data demonstrated that high-risk scores were distinctly predictive of poorer survival outcomes than low-risk score in age ≥ 65 ($p = 0.067$; **Figure 4A**) and < 65 ($p < 0.001$; **Figure 4B**) subgroups, female ($p = 0.042$; **Figure 4C**) and male ($p < 0.001$; **Figure 4D**) subgroups, grade 1–2 ($p = 0.027$; **Figure 4E**) and grade 3–4 ($p < 0.001$; **Figure 4F**) subgroups and stage I-II ($p < 0.001$; **Figure 4G**) and stage III-IV ($p = 0.044$; **Figure 4H**) subgroups.

Constructing a Prognostic Nomogram for HBV-Related HCC

To facilitate personalized treatment, we constructed a nomogram by incorporating gender, age, grade, stage, and risk score in TCGA dataset. This nomogram was utilized for estimating one-, three-, and five-year survival probabilities (**Figure 5A**). ROC curves demonstrated the well performance on predicting one-, three- and five-year (**Figure 5B**). As shown in our

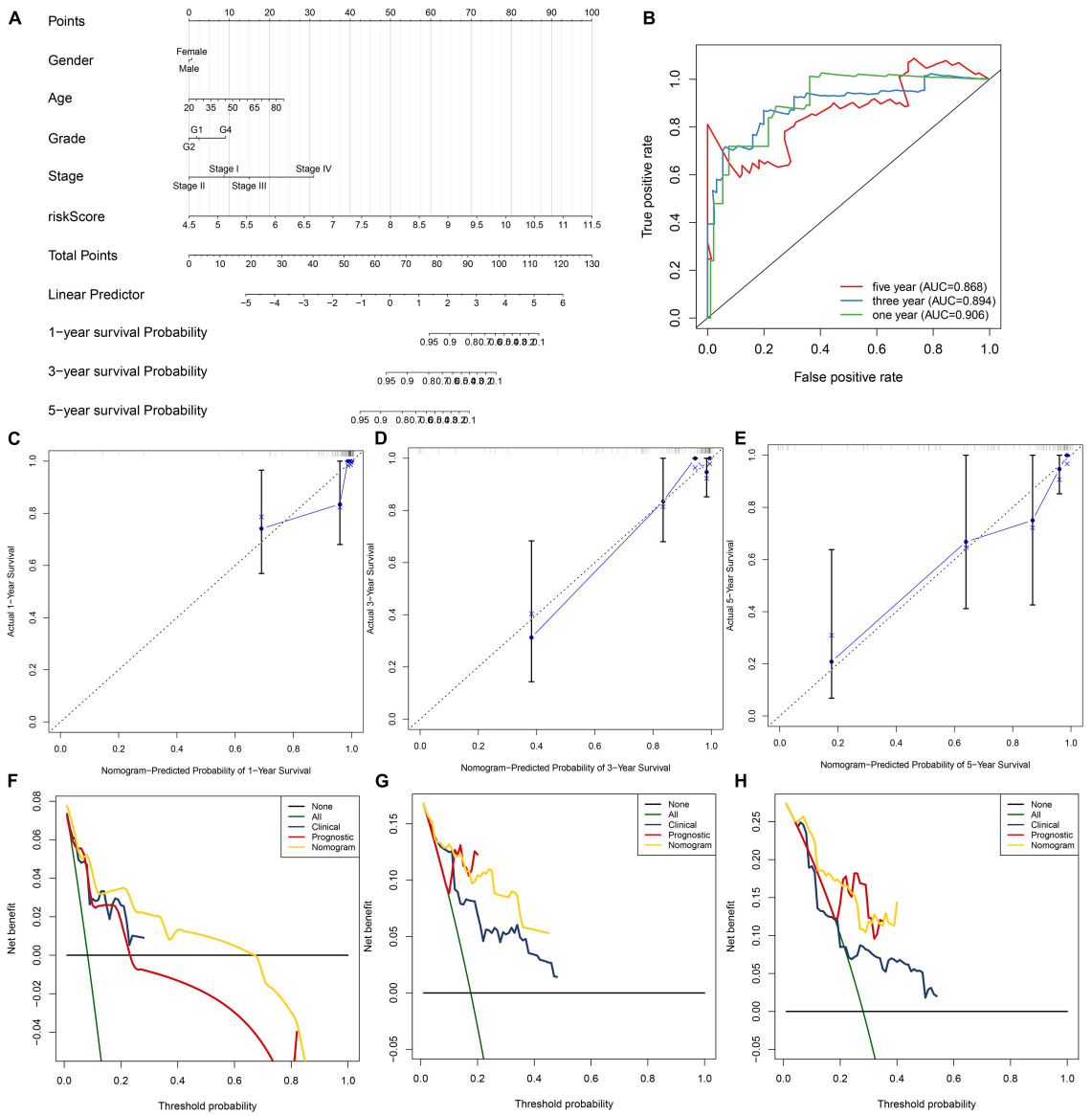

FIGURE 5 | Establishment of the nomogram for HBV-related HCC patients' prognosis. **(A)** The nomogram that contained gender, age, grade, stage, and risk score for estimating one-, three- and five-year survival probabilities. **(B)** The ROC for evaluation of the predictive performance of this nomogram. **(C–E)** The calibration curves for investigating the discrepancy between nomogram-estimated and actual one-, three- and five-year survival duration. **(F–H)** The DCA for calculating the clinical net benefit of the nomogram, clinical factor, and prognostic factor in comparison to all or none strategies.

calibration plots, nomogram-estimated one-, three- and five-year survival probabilities were close to actual survival consequences (**Figures 5C–E**). Meanwhile, DCA showed that this nomogram exhibited the best net benefit for one-, three- and five-year survival duration (**Figures 5F–H**). Hence, the nomogram model could assist clinical management and decision.

Activated Pathways in High-Risk HBV-Related HCC

For observing potential pathways involved in unfavorable survival outcomes, GSEA was carried out. We found that endocytosis (**Figure 6A**), RNA degradation (**Figure 6B**), spliceosome (**Figure 6C**) and ubiquitin-mediated proteolysis (**Figure 6D**) were distinctly activated in high-risk HBV-related HCC specimens.

The Risk Score Is Associated With Immune Microenvironment of HBV-Related HCC

CIBERSORT algorithm was applied for inferring the proportions of 22 immune cells in HBV-related HCC tissues, including B cells naïve, B cells memory, plasma cells, T cells CD8, T

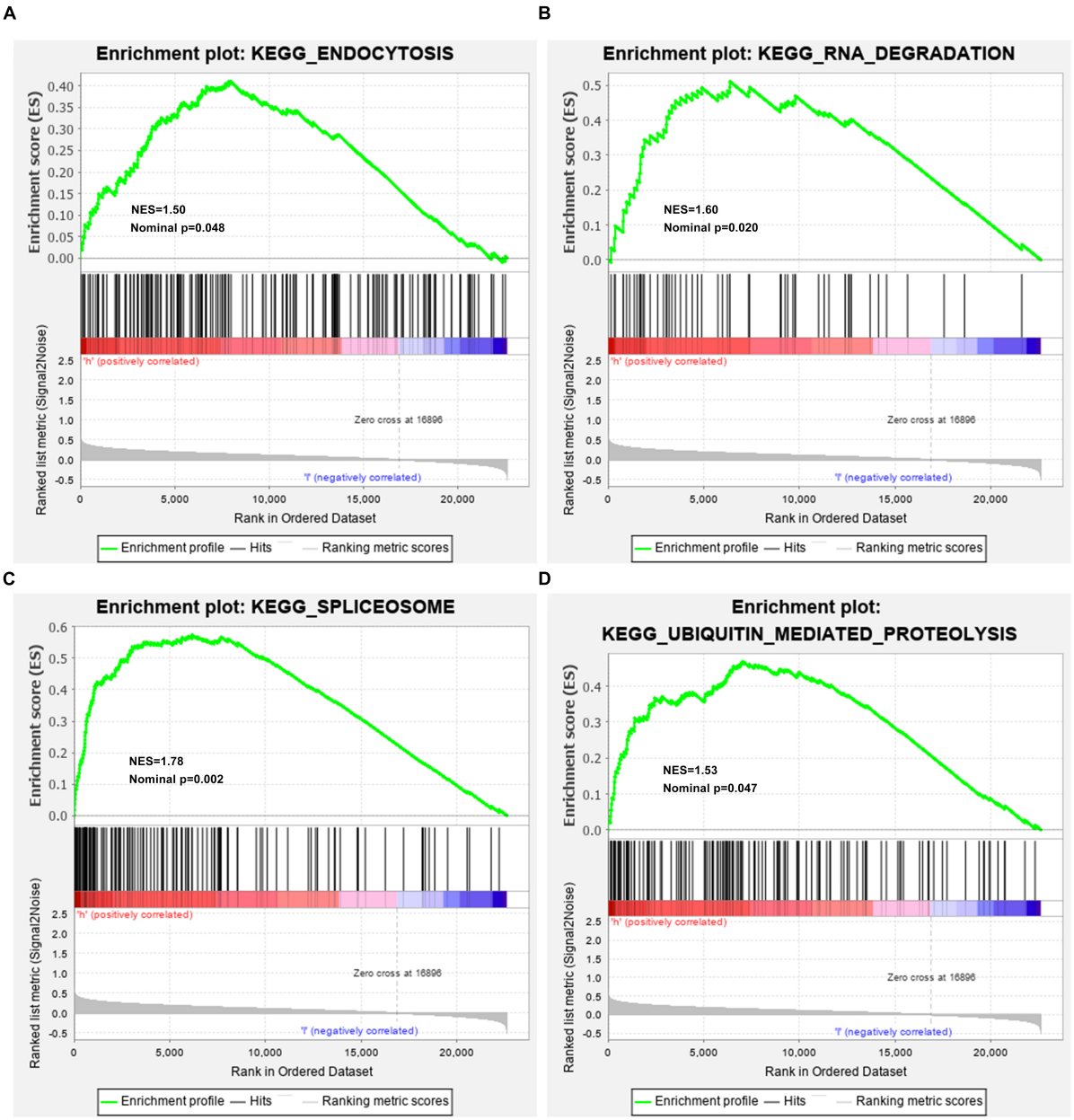

FIGURE 6 | Activated KEGG pathways in high-risk HBV-related HCC samples from the TCGA dataset by GSEA. **(A)** Endocytosis; **(B)** RNA degradation; **(C)** spliceosome; **(D)** ubiquitin-mediated proteolysis.

cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), T cells gamma delta, NK cells resting, NK cells activated, monocytes, macrophages M0, macrophages M1, macrophages M2, dendritic cells resting, dendritic cells activated, mast cells resting, mast cells activated, eosinophils and neutrophils (**Figure 7A**). There was the heterogeneity in the immune microenvironment among subjects. The close crosstalk between immune cells was found, as shown in **Figure 7B**. Also, the risk score was associated with mast cells resting, NK cells resting, neutrophils, mast cells activated, macrophages M0, Tregs and eosinophils. Moreover, the risk score displayed the significant correlations to HLA expression in HBV-related HCC specimens (**Figure 7C**). The data indicated that the risk score was related to the immune microenvironment of HBV-related HCC.

DISCUSSION

In this study, an RBP-related gene model was created, which could robustly predict prognosis and recurrence of HBV-related HCC individuals. High risk scores were indicative of undesirable

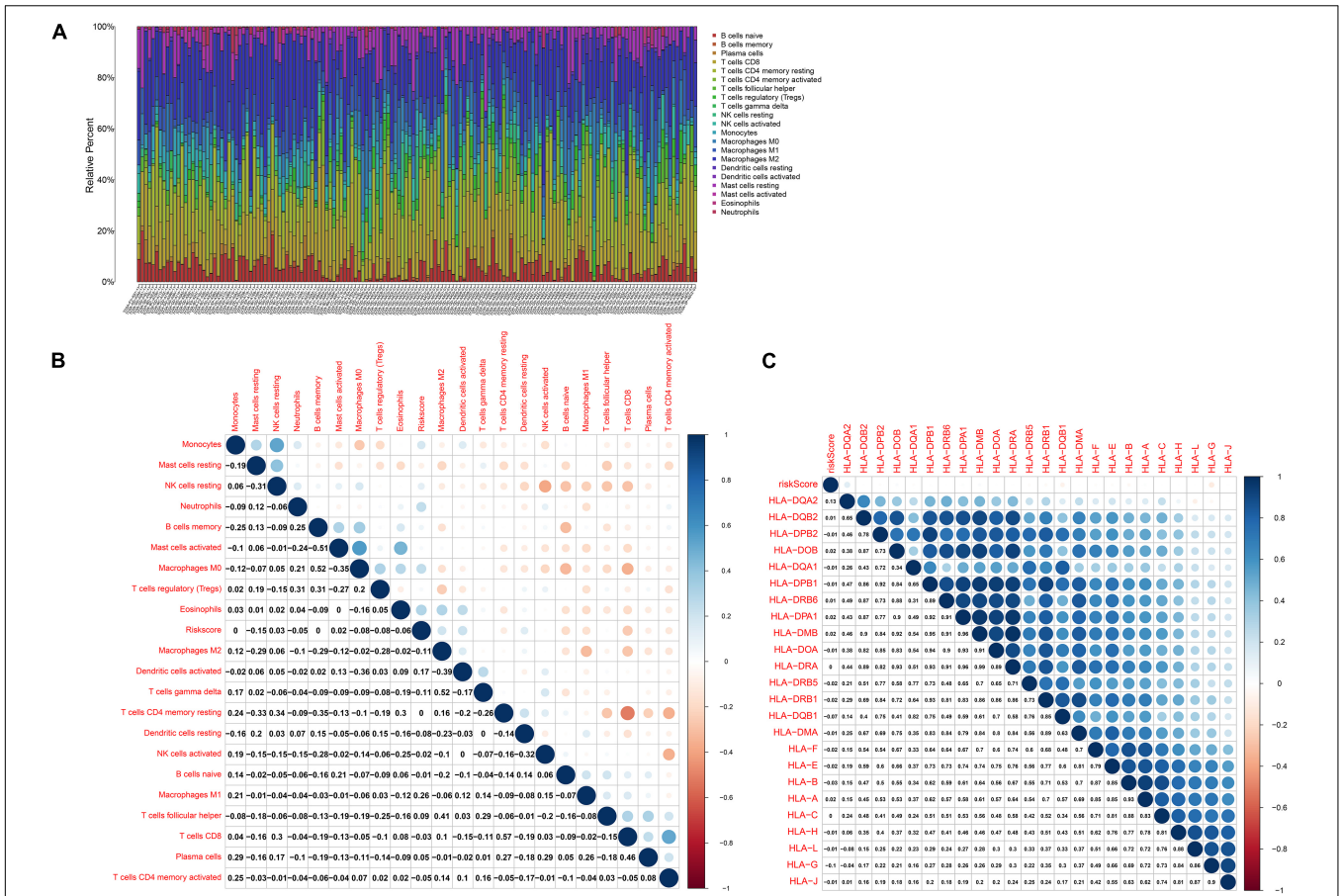

FIGURE 7 | The risk score is related to the immune microenvironment of HBV-related HCC in the TCGA dataset by CIBERSORT. **(A)** The proportions of immune cells in HBV-related HCC samples. **(B)** Heatmap of the correlations between the risk score and immune cells. **(C)** Heatmap visualizing the correlations between the risk score and HLA expression. Correlation coefficients were marked in the box.

survival outcomes. Our data confirmed RBPs as critical elements in the malignant progression of HBV-related HCC. The gene model acted as a key clinical implication in prognostication and therapy decision.

Alterations in RBP expression may lead to carcinogenesis. Here, we identified 340 dysregulated RBPs in HBV-related HCC. These RBPs were distinctly related to mRNA metabolic biological processes such as nuclear-transcribed mRNA catabolic process, RNA catabolic process, nuclear-transcribed mRNA catabolic process, mRNA catabolic process and ncRNA processing as well as key pathways including ribosome, spliceosome, RNA transport, mRNA surveillance pathway, ribosome biogenesis in eukaryotes and RNA degradation. Thus, these RBPs acted as key regulators on post-transcriptional gene expression.

By LASSO analysis, we created an RBP prognostic model, which contained POLR2L, MRPS12, DYNLL1, ZFP36, PPIH, RARS, SRP14, DDX41, EIF2B4, and NOL12. Following external verification, this model possessed higher accuracy and sensitivity on prognostication in comparison to other clinical parameters. The biological implications of above RBPs in this model have been reported in previous research. Liu et al. (2020) found that POLR2L displayed a correlation to survival duration and alternative splicing in lung squamous cell carcinoma patients. MRPS12 functioned as an oncogene and a prognostic candidate in ovarian carcinoma (Qiu et al., 2021). DYNLL1 hypomethylation and upregulation was characterized by stage- and grade-dependent manners and correlated to unfavorable survival outcomes in HCC (Berkel and Cacan, 2020). ZFP36 down-regulation was detected in HCC tissues and served as a tumor suppressor (Kröhler et al., 2019). PPIH was highly expressed in stomach adenocarcinoma and down-regulated PPIH suppressed cellular migratory and invasive behaviors (Li et al., 2021). Also, PPIH up-regulation contributed to undesirable survival outcomes. DDX41 displayed a correlation to tumor stage and grade in HCC (Qi et al., 2020). NOL12 was in relation to kidney renal clear cell carcinoma prognosis (Xiang Y. et al., 2020). Nevertheless, biological roles and clinical implications of these RBPs may require in-depth exploration in HBV-related HCC.

Our results showed that endocytosis, RNA degradation, spliceosome and ubiquitin-mediated proteolysis pathways were markedly activated in high-risk HBV-related HCC specimens, indicating that these pathways might modulate HCC progress

and metastasis. RBPs may regulate the stability of mRNAs encoding immune-related proteins, thereby ornamenting the immune microenvironment (Käfer et al., 2019). For instance, RBP ZFP36, as an inflammatory regulator, restrained T cell activation and anti-viral immunity (Moore et al., 2018). RBP YTHDF3 restrained interferon-dependent anti-viral response through increasing FOXO3 translation (Zhang et al., 2019). Suppressing YTHDF1 in dendritic cells induced durable neoantigen-specific immunity and enhanced the efficacy of anti-PD-L1 therapy (Han et al., 2019). PCBP1 served as an intracellular immune checkpoint toward maintaining T cell functions (Ansa-Addo et al., 2020). Recently, the roles of RBPs have been investigated thoroughly in HCC immune microenvironment of HCC. Xiang J et al. (2020) reported that RBP SIRT7 enhanced the efficacies of anti-PD-L1 therapy through MEF2D in HCC cells. Here, our data demonstrated the close associations of the risk score with immune cells in HCC tissues such as mast cells resting, NK cells resting, neutrophils, mast cells activated, macrophages M0, Tregs and eosinophils. More studies will be carried out for verifying the roles of the risk score on ornamenting immune microenvironment in HCC.

CONCLUSION

Collectively, this study established an RBP model for predicting OS and recurrence of HBV-related HCC individuals. Following external verification, this model possessed the well predictive efficacy and acted as a robust and specific prognostic indicator. Thus, our findings might assist guide clinical decision and personalized therapies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

QM conceived and designed the study. ML, ZL, and JW conducted most of the experiments and data analysis, and wrote the manuscript. HL, HG, SL, and MJ participated in collecting data and helped to draft the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This work was funded by National Science and Technology Major Project Specialized for Infectious Diseases Prevention and Treatment (2017ZX10203201-006 and 2017ZX10202201-004-010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.707305/full#supplementary-material

Supplementary Table 1 | A list of 1,542 RBPs.

REFERENCES

Ansa-Addo, E. A., Huang, H. C., Riesenberg, B., Iamsawat, S., Borucki, D., Nelson, M. H., et al. (2020). RNA binding protein PCBP1 is an intracellular immune checkpoint for shaping T cell responses in cancer immunity. *Sci. Adv.* 6:eaaz3865. doi: 10.1126/sciadv.aaz3865

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Berkel, C., and Cacan, E. (2020). DYNLL1 is hypomethylated and upregulated in a tumor stage- and grade-dependent manner and associated with increased mortality in hepatocellular carcinoma. *Exp. Mol. Pathol.* 117:104567. doi: 10.1016/j.yexmp.2020.104567

Chabrolles, H., Auclair, H., Vegna, S., Lahlali, T., Pons, C., Michelet, M., et al. (2020). Hepatitis B virus Core protein nuclear interactome identifies SRSF10 as a host RNA-binding protein restricting HBV RNA production. *PLoS Pathog.* 16:e1008593. doi: 10.1371/journal.ppat.1008593

Doncheva, N. T., Morris, J. H., Gorodkin, J., and Jensen, L. J. (2019). Cytoscape StringApp: network analysis and visualization of proteomics data. *J. Proteome Res.* 18, 623–632. doi: 10.1021/acs.jproteome.8b00702

Dong, W., Dai, Z. H., Liu, F. C., Guo, X. G., Ge, C. M., Ding, J., et al. (2019). The RNA-binding protein RBM3 promotes cell proliferation in hepatocellular carcinoma by regulating circular RNA SCD-circRNA 2 production. *EBioMedicine* 45, 155–167. doi: 10.1016/j.ebiom.2019.06.030

Engebretsen, S., and Bohlin, J. (2019). Statistical predictions with glmnet. *Clin. Epigenetics* 11:123. doi: 10.1186/s13148-019-0730-1

Fang, Q., and Chen, H. (2020). The significance of m6A RNA methylation regulators in predicting the prognosis and clinical course of HBV-related hepatocellular carcinoma. *Mol. Med.* 26:60. doi: 10.1186/s10020-020-00185-z

Han, D., Liu, J., Chen, C., Dong, L., Liu, Y., Chang, R., et al. (2019). Anti-tumour immunity controlled through mRNA m(6)A methylation and YTHDF1 in dendritic cells. *Nature* 566, 270–274. doi: 10.1038/s41586-019-0916-x

Heagerty, P. J., Lumley, T., and Pepe, M. S. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* 56, 337–344. doi: 10.1111/j.0006-341x.2000.00337.x

Hou, J., Zhang, H., Liu, J., Zhao, Z., Wang, J., Lu, Z., et al. (2019). YTHDF2 reduction fuels inflammation and vascular abnormalization in hepatocellular carcinoma. *Mol. Cancer* 18:163. doi: 10.1186/s12943-019-1082-3

Käfer, R., Schmidtke, L., Schrick, K., Montermann, E., Bros, M., Kleinert, H., et al. (2019). The RNA-Binding Protein KSRP Modulates Cytokine Expression of CD4(+) T Cells. *J. Immunol. Res.* 2019:4726532. doi: 10.1155/2019/4726532

Kröhler, T., Kessler, S. M., Hosseini, K., List, M., Barghash, A., Patial, S., et al. (2019). The mRNA-binding Protein TTP/ZFP36 in Hepatocarcinogenesis and Hepatocellular Carcinoma. *Cancers* 11:1754. doi: 10.3390/cancers11111754

Li, A., Wu, J., Zhai, A., Qian, J., Wang, X., Qaria, M. A., et al. (2019). HBV triggers APOBEC2 expression through miR-122 regulation and affects the proliferation of liver cancer cells. *Int. J. Oncol.* 55, 1137–1148. doi: 10.3892/ijo.2019.4870

Li, J., Zhou, W., Wei, J., Xiao, X., An, T., Wu, W., et al. (2021). Prognostic Value and Biological Functions of RNA Binding Proteins in Stomach Adenocarcinoma. *Onco. Targets Ther.* 14, 1689–1705. doi: 10.2147/ott.S297973

Lin, Y., Liang, R., Qiu, Y., Lv, Y., Zhang, J., Qin, G., et al. (2019). Expression and gene regulation network of RBM8A in hepatocellular carcinoma based on data mining. *Aging* 11, 423–447. doi: 10.18632/aging.101749

Liu, Y., Jia, W., Li, J., Zhu, H., and Yu, J. (2020). Identification of Survival-Associated Alternative Splicing Signatures in Lung Squamous Cell Carcinoma. *Front. Oncol.* 10:587343. doi: 10.3389/fonc.2020.587343

Liu, Z., Jiang, Y., Yuan, H., Fang, Q., Cai, N., Suo, C., et al. (2019). The trends in incidence of primary liver cancer caused by specific etiologies: results from

the Global Burden of Disease Study 2016 and implications for liver cancer prevention. *J. Hepatol.* 70, 674–683. doi: 10.1016/j.jhep.2018.12.001

Moore, M. J., Blachere, N. E., Fak, J. J., Park, C. Y., Sawicka, K., Parveen, S., et al. (2018). ZFP36 RNA-binding proteins restrain T cell activation and anti-viral immunity. *Elife* 7:e33057. doi: 10.7554/eLife.33057

Müller-McNicoll, M., and Neugebauer, K. M. (2013). How cells get the message: dynamic assembly and function of mRNA-protein complexes. *Nat. Rev. Genet.* 14, 275–287. doi: 10.1038/nrg3434

Nakagawa, H., Fujita, M., and Fujimoto, A. (2019). Genome sequencing analysis of liver cancer for precision medicine. *Semin. Cancer Biol.* 55, 120–127. doi: 10.1016/j.semcancer.2018.03.004

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi: 10.1038/nmeth.3337

Qi, Z., Yan, F., Chen, D., Xing, W., Li, Q., Zeng, W., et al. (2020). Identification of prognostic biomarkers and correlations with immune infiltrates among cGAS-STING in hepatocellular carcinoma. *Biosci. Rep.* 40:BSR20202603. doi: 10.1042/bsr20202603

Qiu, X., Guo, D., Du, J., Bai, Y., and Wang, F. (2021). A novel biomarker, MRPS12 functions as a potential oncogene in ovarian cancer and is a promising prognostic candidate. *Medicine* 100:e24898. doi: 10.1097/md.0000000000024898

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007

Sagnelli, E., Macera, M., Russo, A., Coppola, N., and Sagnelli, C. (2020). Epidemiological and etiological variations in hepatocellular carcinoma. *Infection* 48, 7–17. doi: 10.1007/s15010-019-01345-y

Schneider, T., Hung, L. H., Aziz, M., Wilmen, A., Thaum, S., Wagner, J., et al. (2019). Combinatorial recognition of clustered RNA elements by the multidomain RNA-binding protein IMP3. *Nat. Commun.* 10:2266. doi: 10.1038/s41467-019-09769-8

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* 45, D362–D368. doi: 10.1093/nar/gkw937

Torresi, J., Tran, B. M., Christiansen, D., Earnest-Silveira, L., Schwab, R. H. M., and Vincan, E. (2019). HBV-related hepatocarcinogenesis: the role of signalling pathways and innovative ex vivo research models. *BMC Cancer* 19:707. doi: 10.1186/s12885-019-5916-6

Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med. Decis. Making* 26, 565–574. doi: 10.1177/0272989x06295361

Xiang, J., Zhang, N., Sun, H., Su, L., Zhang, C., Xu, H., et al. (2020). Disruption of SIRT7 Increases the Efficacy of Checkpoint Inhibitor via MEF2D Regulation of Programmed Cell Death 1 Ligand 1 in Hepatocellular Carcinoma Cells. *Gastroenterology* 158, 664–678.e24. doi: 10.1053/j.gastro.2019.10.025

Xiang, Y., Zhou, S., Hao, J., Zhong, C., Ma, Q., Sun, Z., et al. (2020). Development and validation of a prognostic model for kidney renal clear cell carcinoma based on RNA binding protein expression. *Aging* 12, 25356–25372. doi: 10.18632/aging.104137

Zhang, C., Huang, S., Zhuang, H., Ruan, S., Zhou, Z., Huang, K., et al. (2020). YTHDF2 promotes the liver cancer stem cell phenotype and cancer metastasis by regulating OCT4 expression via m6A RNA methylation. *Oncogene* 39, 4507–4518. doi: 10.1038/s41388-020-1303-7

Zhang, Y., Wang, X., Zhang, X., Wang, J., Ma, Y., Zhang, L., et al. (2019). RNA-binding protein YTHDF3 suppresses interferon-dependent antiviral responses by promoting FOXO3 translation. *Proc. Natl. Acad. Sci. U. S. A.* 116, 976–981. doi: 10.1073/pnas.1812536116

Zhao, L., Cao, J., Hu, K., Wang, P., Li, G., He, X., et al. (2019). RNA-binding protein RPS3 contributes to hepatocarcinogenesis by post-transcriptionally up-regulating SIRT1. *Nucleic Acids Res.* 47, 2011–2028. doi: 10.1093/nar/gky1209

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fgene.2021.715840

Systematic Construction and Validation of an RNA-Binding Protein-Associated Prognostic Model for Acute Myeloid Leukemia

Hongwei Luo[1†], Yingchun Zhang[2†], Nan Hu[3], Yancheng He[4] and Chengcheng He[2*]*

[1] People's Hospital of Mianzhu, Deyang, China, [2] People's Hospital of Zhongjiang, Deyang, China, [3] School of Clinical Medicine, Southwest Medical University, Luzhou, China, [4] School of Medicine and Rehabilitation, Jiangyang City Construction College, Luzhou, China

Background: The abnormal expression of RNA-binding proteins (RBPs) in various malignant tumors is closely related to the occurrence and development of tumors. However, the role of RBPs in acute myeloid leukemia (AML) is unclear.

Methods: We downloaded harmonized RNA-seq count data and clinical data for AML from UCSC Xena, including The Cancer Genome Atlas (TCGA), The Genotype-Tissue Expression (GTEx), and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) cohorts. R package *edgeR* was used for differential expression analysis of 337 whole-blood data and 173 AML data. The prognostic value of these RBPs was systematically investigated by using univariate Cox regression analysis, least absolute shrinkage and selection operator (LASSO)–Cox regression analysis, and multivariate Cox regression analysis. C-index and calibration diagram were used to judge the accuracy of the model, and decision curve analysis (DCA) was used to judge the net benefit. The biological pathways involved were revealed by gene set enrichment analysis (GSEA). The Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and the protein–protein interaction (PPI) network performed lateral verification on the selected gene set and LASSO results.

Results: A prognostic model of 12-RBP signature was established. In addition, the net benefit and prediction accuracy of the prognostic model and the mixed model based on it were significantly higher than that of cytogenetics. It is verified in the TARGET cohort and shows good prediction effect. Both the selection of our gene set and the LASSO results have high credibility. Most of these pathways are involved in the development of the disease, and they also accumulate in leukemia and RNA-related pathways.

Conclusion: The prognosis model of the 12-RBP signature found in this study is an optimized biomarker that can effectively stratify the risk of AML patients. Nomogram based on this prognostic model is a reliable method to predict the median survival time of patients. This study expands our current understanding of the role of RBPs in the occurrence of AML and may lay the foundation for future treatment of the disease.

Keywords: acute myeloid leukemia, RNA-binding proteins, prognostic signature, bioinformatics, prognostic model

Abbreviations: TCGA, the cancer genome atlas; GTEx, the genotype-tissue expression; TARGET, therapeutically applicable research to generate effective treatments; WBC, white blood cell count; BM, bone marrow; PB, peripheral blood; OS, overall survival; GO, gene ontology; KEGG, kyoto encyclopedia of genes and genomes; PPI, protein–protein interaction.

OPEN ACCESS

Edited by:
Jia Meng,
Xi'an Jiaotong-Liverpool University, China

Reviewed by:
Fengjuan Fan,
Huazhong University of Science and Technology, China
Xu Ye,
The Second Affiliated Hospital of Guangzhou Medical University, China

Correspondence:
Chengcheng He
hcc19861109@163.com
Yancheng He
516577060@qq.com

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *27 May 2021*
Accepted: *26 August 2021*
Published: *24 September 2021*

Citation:
Luo H, Zhang Y, Hu N, He Y and He C (2021) Systematic Construction and Validation of an RNA-Binding Protein-Associated Prognostic Model for Acute Myeloid Leukemia. Front. Genet. 12:715840. doi: 10.3389/fgene.2021.715840

INTRODUCTION

Acute myeloid leukemia (AML) is a cancer of myeloid blood cells, characterized by clonal dilatation of myeloid precursors at different stages of differentiation, resulting in dysgenesis of normal blood cells, and bone marrow failure (Board, 2019). It is the most common subtype of leukemia, with genetic diversity, a worldwide incidence of 3/100,000 per year, poor prognosis, and high mortality (Zhou and Chng, 2014; Döhner et al., 2015). Although most AML patients achieve their first complete remission after induction therapy, relapse is the main reason for the high mortality of patients (Ravandi, 2013; Cornelissen et al., 2015). Postremission therapy (PRT) is an important means to prevent recurrence. According to the recommendations of the European LeukemiaNet (ELN), patients at adverse risk should opt for allogeneic transplantation, and those at favorable molecular risk should undergo intensive chemotherapy (Döhner et al., 2017). Thus, prognostic assessment of the patient is critical to the development of appropriate treatment decisions and follow-up strategies. Cytogenetics is an important prognostic factor for AML and is the basis of current risk classifications for the disease (Döhner et al., 2010; Grimwade et al., 2010). Many cytogenetic abnormalities are known to be associated with poor prognosis and a higher risk of relapse after treatment (Slovak et al., 2000). However, some patients still relapse in the absence of adverse risk factors (Röllig et al., 2011). Therefore, in order to improve the prognosis assessment of AML patients, biomarkers must be optimized. Current sequencing work has revealed extensive genomic heterogeneity of AML and provided valuable information on diagnosis and prognosis, and enabling the optimization of biomarkers.

RNA-binding proteins (RBPs) are proteins that interact with RNA through RNA-binding domains. As important coordinators for maintaining genomic integrity, RBPs are widely expressed in cells and play a core and conservative role in gene regulation (Gerstberger et al., 2014; Nishida et al., 2017). RBPs are involved in regulating all aspects of RNA metabolism and function, including RNA biogenesis, maturation, transport, cellular localization, and stability (Masuda and Kuwano, 2019). When the nuclear RNA emerges from the RNA polymerase, the RNA transcript is immediately covered by the RNA-binding protein, exerting its functions and ultimately affecting the expression of each gene (Campos-Melo et al., 2014). Given the importance of RBPs in regulating life processes, it is not surprising that some aberrant, deregulated RBPs are closely associated with the onset and progression of disease.

Because of their important role, RBPs have been widely studied in recent years. RBPs have been found to play a critical role in tumor development, and hundreds of RBPs are clearly dysregulated in cancer (Wang et al., 2018). In fact, previous studies have linked known cancer drivers to RBP disorders, including AML. Some RBP-encoding genes promote the development of cancer cells. For example, TRIM21 promotes the transformation of breast cancer cells from epithelium to stroma (Jin et al., 2019); FOXK2 promotes colorectal cancer metastasis by upregulating ZEB1 and EGFR expression (Du et al., 2019). Wang E. et al. (2019) revealed 21 RBP candidates upregulated in AML that are critical for maintaining RNA splicing and survival of AML. Mutations in RPS14, SRBP2, SF3B1, and U2AF1 can lead to myelodysplastic abnormalities, hematopoietic dysfunction, AML, and other blood-related diseases (Ebert et al., 2008; Komeno et al., 2015; Shirai et al., 2015; Mortera-Blanco et al., 2017; de Rooij et al., 2019).

Taken together, these studies suggest that RBPs is closely related to the occurrence and development of human tumors. RBP-encoding genes have been used to build prognostic models of cancer but are still lacking in AML. For example, Li et al. (2020) used eight-RBP gene to predict the prognosis of patients with lung adenocarcinoma. Therefore, in AML, systematic use of high-throughput transcriptome data to identify the expression profile of RBP-encoding genes in normal and tumor tissues is a necessary step to understand its role in the pathogenesis, which not only contributes to the understanding of the pathogenesis, but also has a guiding role in the prognosis.

MATERIALS AND METHODS

Data Collection

UCSC Toil RNA-seq Recompute[1] processing more than 20,000 unaligned RNA samples from The Cancer Genome Atlas (TCGA),[2] Therapeutically Applicable Research to Generate Effective Treatments (TARGET),[3] and Genotype-Tissue Expression (GTEx)[4] datasets resulted in a combined cohort free of computational batch effects between different repository. In this study, the RNA-seq expression profiles and corresponding clinical data of AML patients in the TCGA and TARGET were retrieved, respectively, from the UCSC Toil RNA-seq Recompute. The AML patients from TCGA-LAML project were chosen as the training cohort to establish the risk classification system based on the RBP signatures and to construct predictive model. An independent dataset (TARGET[5]) was employed for its external validation. The case selection criteria for data extraction were patients diagnosed with AML and available clinical information such as survival status and overall survival time, age, gender, FAB classification, and cytogenetic risk stratification.

For RBP-encoding genes, we obtained a reliable correlation gene summarized by Gerstberger et al. (2014)[6]. The summation of all RBP-encoding genes was used to further identify AML-related features. For clinical characteristics, R package *tableone* was used to use chi-square test for classified data, and analysis of Kruskal test was used for continuous variables, which were represented by median (Yoshida et al., 2020).

[1]https://toil.xenahubs.net

[2]https://portal.gdc.cancer.gov/

[3]https://ocg.cancer.gov/programs/target

[4]https://gtexportal.org/home/

[5]https://ocg.cancer.gov/programs/target/projects/acute-myeloid-leukemia

[6]https://www.nature.com/articles/nrg3813#supplementary-information

Identification of Differentially Expressed mRNA in Acute Myeloid Leukemia

Differential expression analysis was performed between 337 whole blood RNA-seq data of GTEx and 173 AML RNA-seq data of samples of TCGA using the R package *edgeR* (Robinson et al., 2010), and the screening criteria were |log2 (fold change)| $\geq$ 1.5 and FDR <0.05. The final results were visualized by using *ggplot* to plot the volcano for the differences between RBP-encoding genes and remaining genes (Wickham, 2016).

Construction of the Prognostic RNA-Binding Protein-Encoding Gene Signature

Univariate Cox regression analysis and Least absolute shrinkage and selection operator (LASSO)-Cox regression analysis were performed to identify the prognosis-related RBP-encoding genes and construct the prognostic gene signature. We used the TCGA data set as the training cohort and the TARGET data set as the validation cohort.

Least absolute shrinkage and selection operator is a popular method that avoids overfitting by incorporating the best performance parameters to produce a simpler and more easily interpreted model, which is widely used in Cox proportional hazard regression model for high-dimensional data survival analysis (Simon et al., 2011). The R package *survival* was used for Univariate Cox regression analysis, and the RBP-encoding genes with differential expression of $p < 0.01$ were screened and incorporated into the LASSO regression model (Therneau and Grambsch, 2000). The LASSO regression was analyzed with R package *glmnet*, and the prognosis model of RBP-encoding genes was generated (Friedman et al., 2010). In the LASSO regression, the setting parameters are cross-verified, and the partial likelihood deviation satisfies the minimum criterion. The risk score was constructed based on the expression of prognostic RBP-encoding genes. The risk score for each sample was calculated as the following formula:

$$\textit{Risk score} = \sum_i (\textit{Coef}_i \times \textit{Exp}_i)$$

$\textit{Exp}_i$ is the relative expression of the gene in the patient signature, and $\textit{Coef}_i$ was the LASSO coefficient of the gene. The median risk score in the training cohort was used as the cutoff value for the AML cohort dichotomy. In both the training and validation sets, patients were divided into high- and low-risk groups based upon resultant risk score values, respectively. Kaplan–Meier (KM) survival curve and time-dependent ROC curve estimates were then performed for each cohort to assess the predictive power of the prognostic model of RBPs.

Model Construction and Validation

To determine whether genes can be used independent of clinical information as a prognostic indicator for AML patients, univariate and multivariate Cox regression analyses were performed.

Univariate Cox screening index was used, which included the clinical characteristics common to TCGA cohort and the TARGET cohort, as well as the high- and low-risk index of base 12 gene construction, to include the index with $p < 0.05$ into the multivariate Cox regression model. Then stepwise regression was employed to further select the best model. Forest plots provided a visualization of the hazard ratio (HR) and 95% confidence intervals.

The nomogram is a kind of visual regression model, which sets the scoring standard according to the regression coefficient of independent variables, through which we can calculate and predict the patient outcome by comparing the situation of the patient (Iasonos et al., 2008). In our study, a combined model of all independent prognostic characteristics screened by regression analysis was used to establish a nomogram to assess median survival in AML patients. The calibration diagram and C index are used to evaluate the predicted results of the regression model. The decision curve analysis (DCA) quantifies the net benefit under different threshold probabilities to determine the clinical practicability of nomograms and to find the model that predicts the maximum net benefit, so it is widely used (Vickers and Elkin, 2006). The C index and DCA are used to compare the prediction accuracy between individual components and composite models.

Protein–Protein Interaction, Gene Ontology, and Kyoto Encyclopedia of Genes and Genomes Analyses Were Used to Further Verify the Results

After differential analysis, differentially expressed RBP-encoding genes were extracted for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to determine whether they were enriched in RBP-related pathways. We constructed a protein–protein interaction (PPI) network for the different RBP-encoding genes, identified the core module, and judged the distribution of genes screened by LASSO in the module. In the GO analysis and KEGG pathway analysis, respectively, enrichGO function and enrichKEGG function in the R package *clusterProfiler* are used (Yu et al., 2012). Finally, the bubble chart showing both was used (Yu et al., 2012). Both *p*-values and *q*-values <0.05 were considered statistically significant. We input the differentially expressed RBP-encoding genes into the STRING database[7] to obtain the protein–protein interaction network, and the results were visualized using Cytoscape 3.7.2 software (Franceschini et al., 2012). The molecular complex detection (MCODE) plug-in was used to identify the core module, and the distribution of the 12-RBP signature in the core module was detected, and then the detected module was displayed.

Gene Set Enrichment Analyses

Gene set enrichment analysis (GSEA) is used to determine whether a defined set of genes has statistically significant and consistent differences between two biological states (Subramanian et al., 2005). Using the R package *clusterProfiler* (Yu et al., 2012), the potential biological pathways between high- and low-risk groups were recognized. The enriched pathways

[7]http://www.string-db.org/

in each phenotype were sequenced using nominal *p*-values and normalized enrichment scores (NES), and we selected a subset of these pathways for display.

RESULTS

Patient Characteristics

As shown in **Table 1**, a total of 151 TCGA cohort samples and 228 TARGET cohort samples were divided into high- or low-risk groups by 12 prognostic RBP-encoding genes, and the data distribution of different clinical characteristics and statistical tests was calculated. In the TCGA cohort, age, BM blasts, and cytogenetic risk stratification were considered statistically significant, while only cytogenetic risk stratification was observed in the TARGET cohort (**Table 1**).

Differentially Expressed RNA-Binding Protein-Encoding Genes Between Acute Myeloid Leukemia and Normal Tissue

A total of 6,278 genes with significant differences in expression were identified between TCGA—AML and GTEx—blood samples, which were tested by quasi-likelihood *F*-tests and generalized linear models (glms) in *edgeR*. A total of 322 differential RBP-encoding genes met the criteria, of which 221 were upregulated and 101 were downregulated (**Supplementary Table 1**). **Figure 1A** shows a volcanic map showing this differential distribution.

Construction and Validation of Prognostic RNA-Binding Protein Signature Model

First, a total of 33 RBP-encoding genes with significant prognostic correlation were screened for TCGA (training cohort) by univariate Cox regression analysis (**Supplementary Table 2**). LASSO regression analysis was then performed to further screen genes to avoid overfitting and generate a simpler, more easily interpreted model (**Figures 1B,C**). Finally, 12 prognostic RBP-encoding genes were selected, including LARP1B, TRNT1, SMN2, MRPL28, TRIM21, RPS19BP1, XPO6, TSR2, ISG20, HELZ2, EXOSC4, and EIF2AK4. In the training set (TCGA) and the validation set (TARGET), both differential expression analyses yield basically the same results of the 12 gene expression trends between AML and normal samples (**Supplementary Figure 1**). The risk score = −0.01322 × expression of LAR P1B−0.113572 × expression of TRNT1−0.16828 × expression of SMN2+0.01175 × expression of MRPL28+0.02097 × expression of TRIM21+0.06365 × expression of RPS19BP1+0.08799 × expression of XPO6+0.09969 × expression of TSR2+0.10 714 × expression of ISG20+0.15339 × expression of HELZ2+0.18212×expression of EXOSC4+0.29918×expression of EIF2AK4. The truncated value of the risk score of the patient was divided into high-risk group and low-risk group according to the median value of TCGA (training cohort) risk score of −0.085.

In the high- and low-risk groups with truncated boundaries, the survival analysis showed significant differences in both the training and validation cohort data sets (**Figures 2A,B**).

TABLE 1 | Correlation of clinicopathologic characteristics and the 12-gene signature in acute myeloid leukemia (AML).

	Training cohort (TCGA)				Validation cohort (TARGET)			
Characteristics	**Total patients**	**High risk**	**Low risk**	***p*-value**	**Total patients**	**High risk**	**Low risk**	***p*-value**
Gender				0.679				0.598
Female	70	33	37		116	55	61	
Male	81	42	39		112	58	54	
Age (median)	57	63.0	51.0	**<0.001**	10	10.0	9.0	0.359
FAB Category				**0.002**				0.288
M0	13	5	8		8	6	2	
M1	35	20	15		27	12	15	
M2	35	17	18		52	20	32	
M3	13	2	11		58	26	32	
M4	33	12	21		40	25	15	
M5	17	14	3		4	2	2	
M6	2	2	0		10	6	4	
M7	3	3	0		15	7	8	
WBC (median)	15.5	12.0	20.0	0.682	36.8	31.1	49.5	0.086
BM Blasts (median)	32	17.0	39.5	**0.021**	73.2	77.5	72.0	0.546
PB Blasts (median)	71	67.0	72.0	0.254	60	56.5	62.0	0.390
Cytogenetic risk stratification				**0.001**				**<0.001**
Favorable	29	6	23		78	24	54	
Intermediate	89	47	42		118	69	49	
Adverse	31	21	10		22	14	8	

WBC, white blood cell count; BM, bone marrow; PB, peripheral blood; TCGA, the cancer genome atlas; TARGET, therapeutically applicable research to generate effective treatments. The bold value means P < 0.05 with statistically significant.

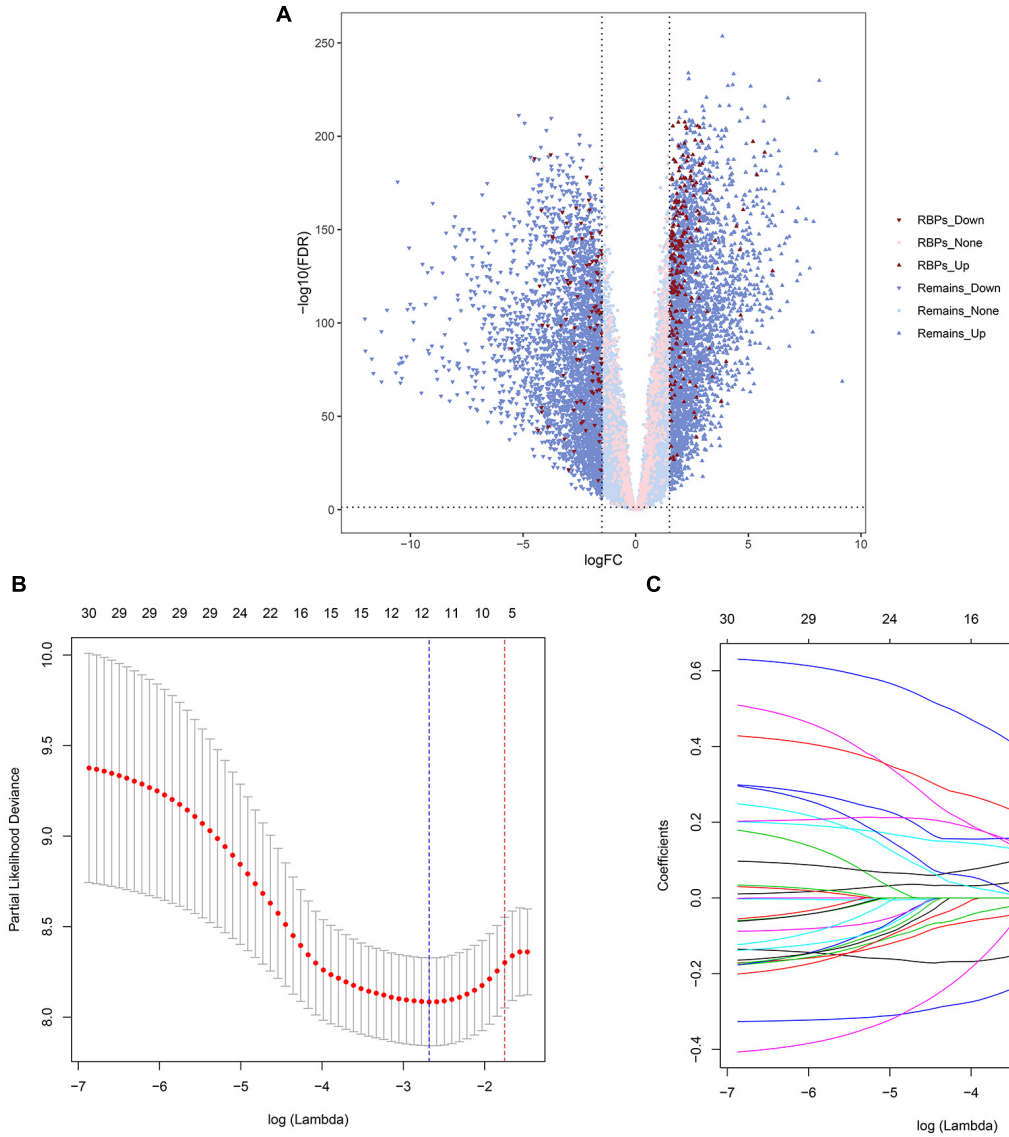

FIGURE 1 | Volcano map and least absolute shrinkage and selection operator (LASSO) regression analysis of The Cancer Genome Atlas (TCGA) cohort. The volcano map shows the differential expression of genes. The up and down arrows represent upregulated and downregulated genes, respectively; red and pink represent genes with and without differential expression of RNA-binding protein (RBP)-encoding genes, respectively **(A)**. The selection of LASSO regression truncation value after cross-validation, and the blue dotted line represents the lambda corresponding to the lowest error mean. The red dotted line represents the maximum lambda corresponding to the error mean within one standard deviation of the minimum **(B)**. Prognosis-related gene selection in the LASSO–Cox regression **(C)**.

A heat map shows the prognosis of the 12-RBP signature in the distribution of the high- and low-risk groups; one of the first module contains three protect genes (regression coefficient is less than zero), and their expression in the low-risk group is higher than in the high risk group. The second module contains nine risk genes (regression coefficient is greater than zero); their expression in the low-risk group is lower than in the high-risk group (**Figures 2C,D**). The areas under the ROC curve of 1-, 3-, and 5-year overall survival (OS) rates for patient risk score in the training cohort were 0.724, 0.683, and 0.650 (**Figure 2E**), and 0.724, 0.683, and 0.650, respectively, in the validation cohort (**Figure 2F**). Taken together, these results suggest that the risk score based on 12-RBP signature is a good predictor of patient prognosis.

The Combination Model Has Good Predictive Effect

Univariate Cox was used to analyze the risk indicators constructed based on the 12-RBP signature and the clinical features including gender, age, FAB category, WBC, BM blasts, PB blasts, and cytogenetic risk stratification. It was found that age, cytogenetic risk stratification, and risk indicators were of significant prognostic value. After that, the stepwise regression

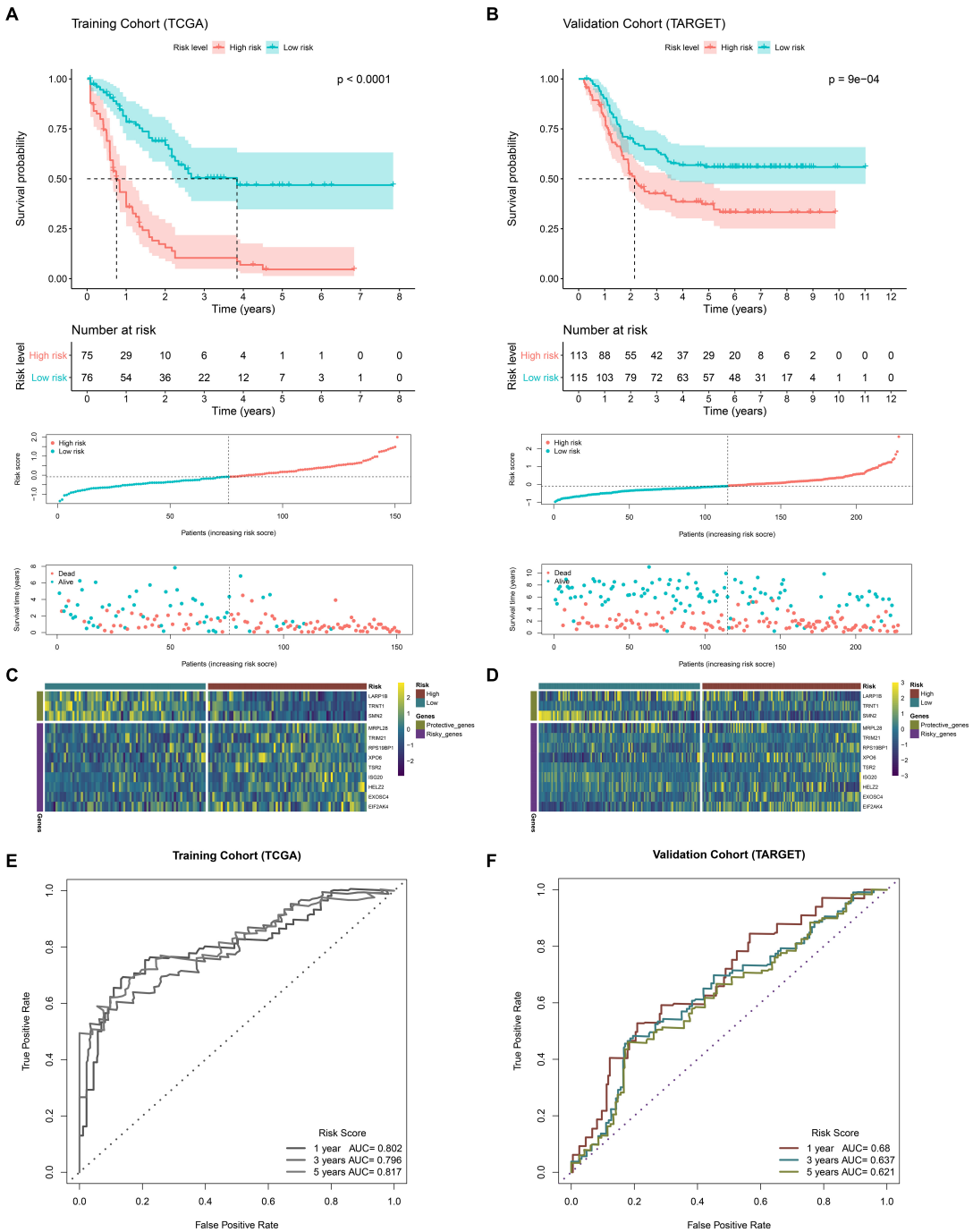

FIGURE 2 | Correlation analysis based on the 12-RBP signature. Survival analysis of high and low risks separated by the median value of the TCGA cohort risk score in the TCGA **(A)** and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) cohort **(B)**. The heatmap shows the expression of the 12-RBP signature at high and low risks. The green module indicates protective genes with regression coefficient less than zero, and the purple module indicates risk genes with regression coefficient greater than zero **(C,D)**. ROC curve of risk score at 1, 3, and 5 years **(E,F)**.

optimization model was further used, and cytogenetic risk stratification and risk indicators were included in the multivariate Cox regression. It was found that these two factors had a great influence on the prognosis of patients in the two data sets, and they were both independent prognostic factors of OS (**Figure 3**).

A nomogram was constructed using the two previously screened indicators, and the median survival time was used to demonstrate the prognosis of patients (**Figure 4A**). In order to verify the accuracy of the model, we calculated the C index and drew the calibration curve. In the training cohort

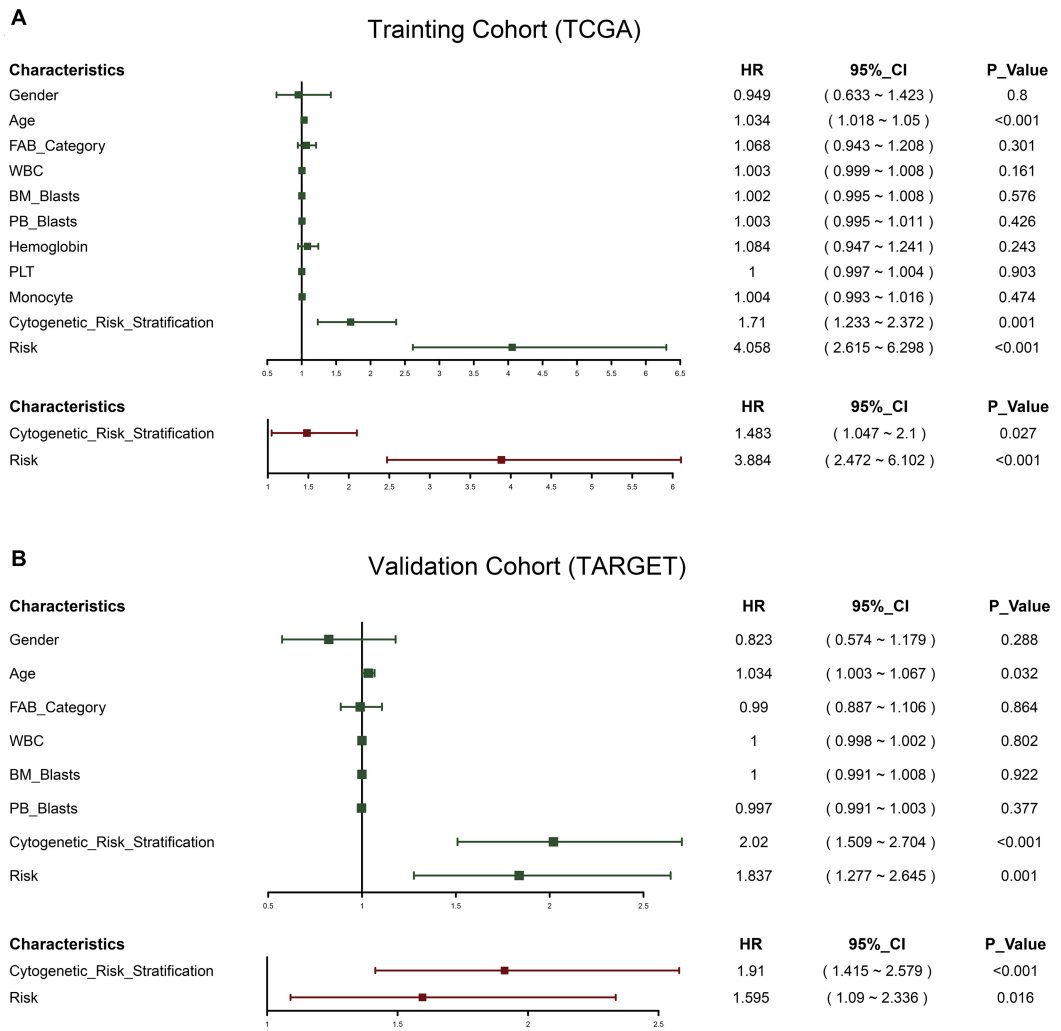

FIGURE 3 | Forest plot of the univariate and multivariate Cox regression analysis in acute myeloid leukemia (AML). Green represents univariate Cox, and red represents multivariate Cox. TCGA cohort **(A)**; TARGET cohort **(B)**.

(TCGA), the C index of the cytogenetic model, the prognostic model, and the combined model were 0.590, 0.675, and 0.699 (**Table 2**), respectively. According to the results of the C index, the prognostic model constructed by us has a better prediction effect than the cytogenetic model currently used in clinical application to judge the prognosis of patients, and the prediction accuracy of the combined model is significantly higher than that of the single index model. In addition, the calibration diagram also shows that the nomogram performs well (**Figure 4B**). According to DCA, both the prognostic model and the combined model had a higher clinical net benefit rate than the cytogenetic model at 1, 2, and 3 years (**Figures 4C–E**). In **Supplementary Figure 2**, the prognostic model has a good degree of recognition within clinically relevant subgroups by survival analysis. Therefore, the model we analyzed is effective for the heterogeneous disease of AML.

To sum up, the prognostic model constructed by us may improve the prediction accuracy of traditional cytogenetic model and bring some net clinical benefits, which is helpful for clinical management.

The Selection and Subsequent Analysis of RNA-Binding Protein-Encoding Gene Data Were Verified to Be Accurate

In differential expression analysis, 322 RBP-encoding genes were changed between AML and normal samples. According to the results of GO and KEGG analysis, it was found that the differentially expressed RBP-encoding genes were enriched in the processes of synthesis, regulation, transport, and translation of RNA, indicating the reliability of the source of these RBP-encoding genes (**Figure 5** and **Supplementary Table 3**). In order to study the interaction between the differentially expressed RBP-encoding genes, we created a PPI network, which demonstrated a total of 3,038 edges and 322 nodes (**Figure 6**). The MODE plugin was used to identify a total of 11 core modules

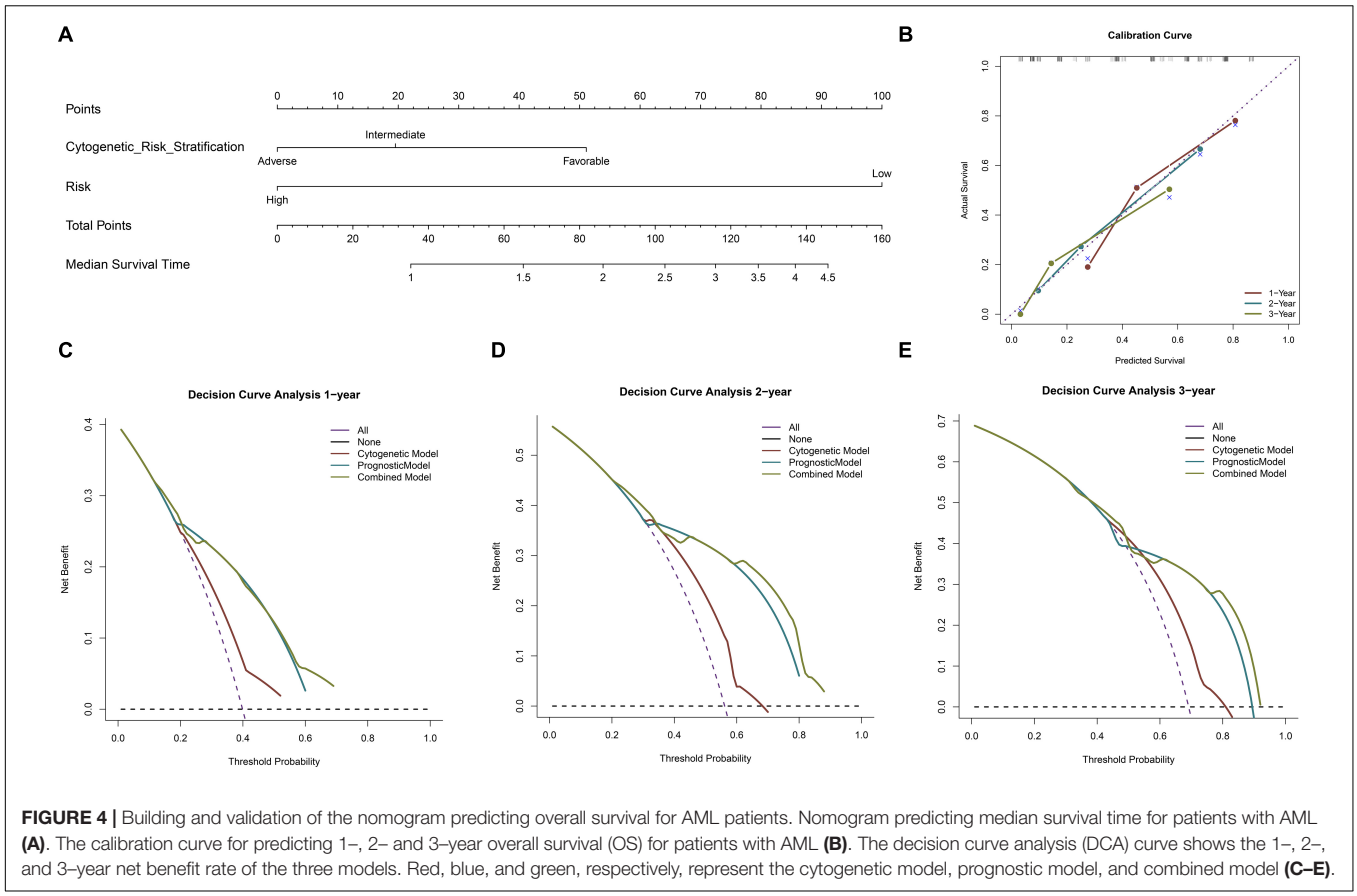

FIGURE 4 | Building and validation of the nomogram predicting overall survival for AML patients. Nomogram predicting median survival time for patients with AML **(A)**. The calibration curve for predicting 1–, 2– and 3–year overall survival (OS) for patients with AML **(B)**. The decision curve analysis (DCA) curve shows the 1–, 2–, and 3–year net benefit rate of the three models. Red, blue, and green, respectively, represent the cytogenetic model, prognostic model, and combined model **(C–E)**.

TABLE 2 | Comparison of the cytogenetic model, prognostic model, and combined model.

	Training cohort (TCGA)			Validation cohort (TARGET)		
Models	**C-index**	**95% CI**	***p*-value**	**C-index**	**95% CI**	***p*-value**
Cytogenetic model	0.59	(0.531–0.65)	–	0.62	(0.575–0.665)	–
Prognostic model	0.675	(0.629–0.721)	<0.001	0.576	(0.53–0.622)	<0.001
Combined model	0.699	(0.646–0.751)	<0.001	0.646	(0.595–0.696)	<0.001

(**Supplementary Table 4**). We searched the distribution of 12-RBP signatures in the core modules and found that five of them were targeted in the five core modules of 1, 3, 5, 8, and 11. These five core modules are associated with the processes and biogenesis of ribosome, spliceosome, RNA transport, and degradation (**Supplementary Figure 3** and **Supplementary Table 5**). In a word, evidence from pathway enrichment analysis indicating that the prognostic genes we screened had a high degree of reliability.

Gene Set Enrichment Analyses

Gene set enrichment analyses analyzed 120 significantly enriched KEGG pathways between the high- and low-risk groups (**Supplementary Table 6**). The most enriched pathways are involved in the development of the disease, and they also accumulate into pathways associated with leukemia. The majority of the ABC transporters, hematopoietic cell lineage, NF-kappa B signaling pathway, chemokine signaling pathway, Toll-like receptor signaling pathway, VEGF signaling pathway, and JAK-STAT signaling pathway were enriched in the high-risk group (**Figure 7**). These biological processes are highly associated with promoting cell survival and inhibiting cell death, which play an important role in AML.

DISCUSSION

RNA-binding proteins are an important class of evolutionarily conserved proteins involved in regulating all aspects of RNA metabolism and functions (Nishida et al., 2017; Masuda and Kuwano, 2019). Disorders of these genes *in vitro* have been shown to cause a variety of diseases, including AML. Due to the important characteristics of RBPs, the use of RBPs to evaluate the prognosis of some cancers has achieved good results (Wang K. et al., 2019; Li et al., 2020). AML has a high recurrence rate due to the avoidance of drugs by leukemic stem cells (LSCs). PRT mode

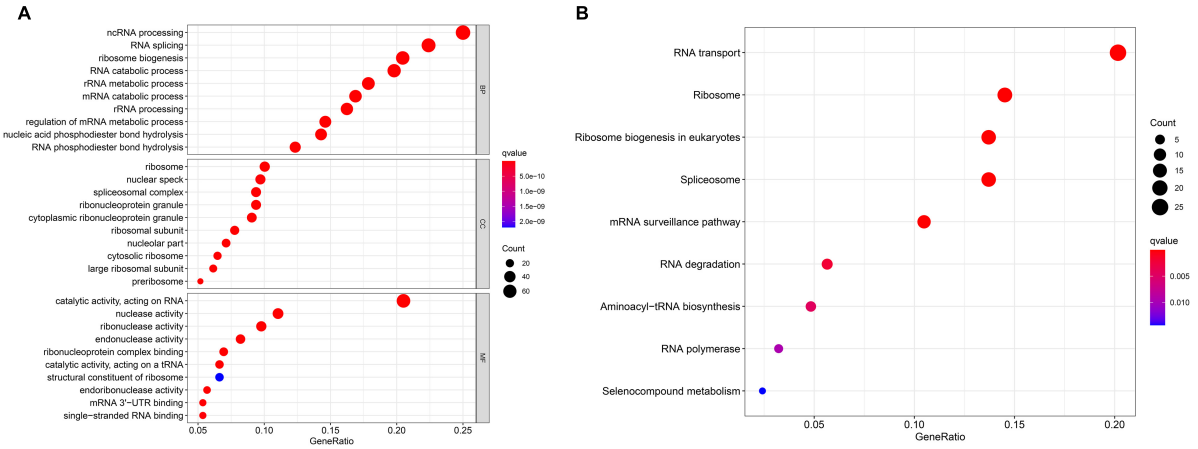

FIGURE 5 | Gene Ontology (GO) enrichment analysis **(A)** and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway **(B)** of differentially expressed RBP-encoding genes.

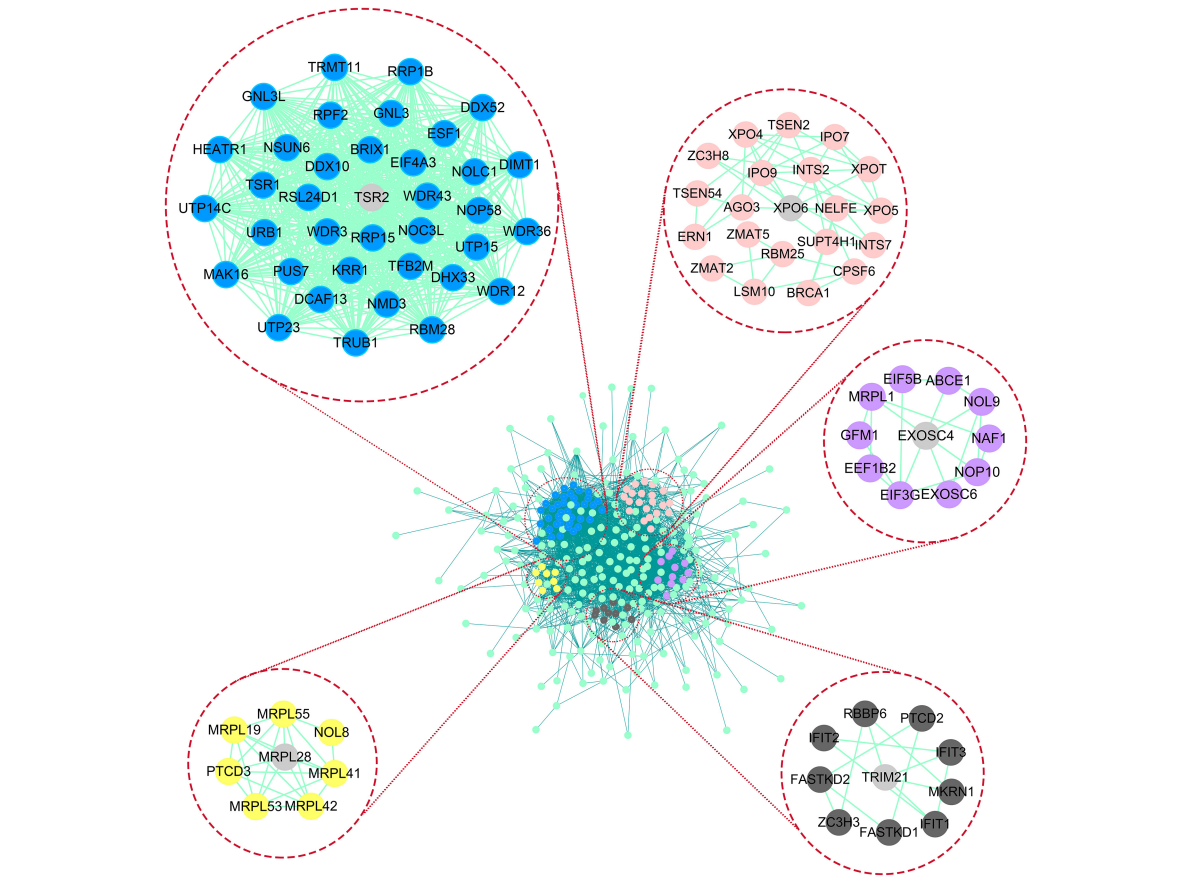

FIGURE 6 | Protein–protein interaction (PPI) network and modules analysis. In the PPI network of differential genes, the matching of 12-RBP signature and core module in LASSO results was demonstrated, among which five core modules contained prognostic gene, and the gray gene represented the matching gene.

selection is the main means to prevent recurrence, but it depends on reliable prognostic markers to determine the prognosis of patients. Therefore, our systematic estimation of RBP-encoding gene changes in AML may be an important way to improve the prognosis assessment of patients with AML and may shed light on the underlying biological mechanisms.

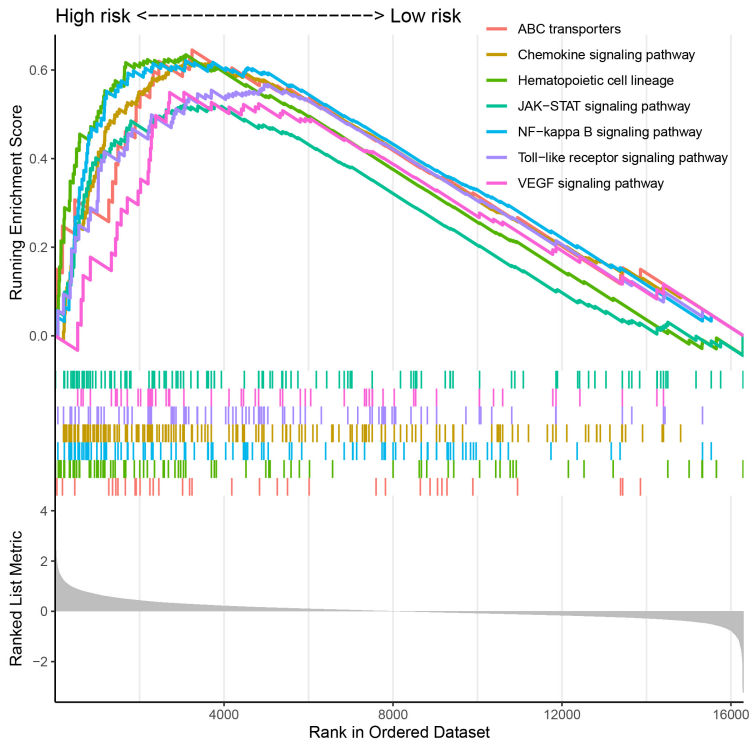

FIGURE 7 | The significantly enriched KEGG pathways by Gene Set Enrichment Analyses (GSEA).

In this study, a new and efficient prognostic feature based on the 12-RBP signature was determined based on the TCGA dataset (the training cohort), and its effectiveness was also verified in the TARGET dataset (the independent external validation sets). The group defined as high risk had poor prognosis, which was consistent in validation cohorts. Moreover, this prognostic feature is independent of other clinical factors, showing a stably high prognostic value for AML. In addition, it was considered that both the cytogenetic risk stratification and the risk indicator were significant prognostic values in AML by Cox regression. Combining the prognostic feature with the cytogenetic risk stratification, a combined model was constructed, which has higher prediction accuracy and clinical net benefit than the single model and provides a potential theoretical basis for clinical application.

In our study, 12-RBP signatures were identified and constructed the prognostic model. The expressions of LARP1B, TRNT1, and SMN2 were correlated with favorable outcomes. On the contrary, the expressions of MRPL28, TRIM21, RPS19BP1, TSR2, XPO6, ISG20, HELZ2, EXOSC4, and EIF2AK4 were involved in adverse outcomes. According to previous studies, most of the 12 RBP-encoding genes in our model are strongly cancer and other diseases. Some genes can be used as biomarkers for prognosis and diagnosis of diseases. LARP1B is a member of the evolutionary conserved family of La-related proteins (LARP) involved in RNA transcription, translation, and B-cell differentiation, which has been shown to drive tumorigenesis (Stavraka and Blagden, 2015; Lagou et al., 2018). TRNT1, as an enzyme necessary for the synthesis of the 3′-terminal CCA sequence in tRNA molecules, can lead to developmental delay, sideroblastic anemia, periodic fever, retinitis pigmentosa, B-cell immunodeficiency, and other diseases when it is abnormal (Frans et al., 2017; Slade et al., 2020). Spinal muscular atrophy (SMA) results from the absence or mutation of SMN1, plus the inability of SMN2 to compensate for the loss of SMN1 due to exon seven jumping (Lorson et al., 1999). MRPL28 encodes mitochondrial ribosomal protein, and its low expression can reduce the mitochondrial activity of pancreatic tumor cells and increase glycolysis (Chen et al., 2009). TRIM21, as a E3 ubiquitin ligase with multiple domains, regulates ubiquitination and proteasomal degradation and is responsible for the control of cell protein expression. It has been shown to regulate the cell cycle, cell proliferation and differentiation of cancer, and is a prognostic marker for hepatocellular carcinoma, breast cancer, pancreatic cancer, and lymphoma (Brauner et al., 2015; Ding et al., 2015; Nguyen and Irby, 2017; Zhou et al., 2018). As the first reported direct regulator of SIRT1 that modulates p53-mediated growth regulation, RPS19BP1 (also known as AROS) promotes survival in a panel of human cancer cell lines (Kim et al., 2007; Knight et al., 2013). TSR2 can induce apoptosis of laryngeal cancer cells by inhibiting NF-κB signaling pathway (He et al., 2018). In addition, TSR2 variation is associated with Diamond–Blackfan anemia (DBA) (Clinton and Gazda, 2009). As a nucleocytoplasmic transporter, XPO6 expression is closely related to poor prognosis of patients and is a potential prognostic biomarker for prostate cancer (Hao et al., 2016).

ISG20 is involved in small nucleolar RNA maturation and ribosomal biogenesis, and controls RNA stability (Espert et al., 2006). ISG20 expression has been shown to be elevated in the course of infection and a potential biomarker for several types of cancer (Rajkumar et al., 2011; Van Tong et al., 2018; Xu et al., 2020). Katano-Toki et al. (2013) found that HELZ2 synergistic with Thrap3 enhances PPAR enzyme-mediated gene activation and plays an important role in the terminal differentiation of adipocytes. EXOSC4, one of the noncatalytic members of RNA exosome complex, is involved in RNA degradation and has been reported to promote colorectal cancer (Pan et al., 2018). EIF2AK4 induces subunit phosphorylation of the translation initiation factor eIF2, which plays an important role in oncogenesis (Koromilas, 2015). For example, EIF2AK4 participates in the EIF2AK4–EIF2alpha–ATF4 pathway, which is crucial for maintaining metabolic homeostasis in tumor cells and a potential target for tumor therapy (Ye et al., 2010; Singleton and Harris, 2012). Although the role of the 12 RBP-encoding gene expression in the pathogenesis of AML remains unclear, these genes mediate important biological processes, and their abnormalities can lead to disease. What is more, most of these genes are instructive for the prognosis in different diseases, indicating the important role of these genes.

CONCLUSION

The 12-RBP signature prognostic model is an optimized biomarker for predicting the prognosis of AML patients and can be used to select postremission treatment (PRT) for AML patients, thereby reducing recurrence rates. Nomogram, a prognostic model, predicts median survival time. This study expands our current understanding of the role of RBPs in the occurrence of AML and may lay the foundation for future treatment. However, the current study also has some limitations, and we need to conduct experimental verification of the screened important genes to further verify our results.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://xenabrowser.net/datapages/.

AUTHOR CONTRIBUTIONS

HL and YZ conceived and designed the study. NH and YH performed the data analysis. CH and HL wrote the manuscript. YH and CH revised the manuscript. All authors read and approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.715840/full#supplementary-material

REFERENCES

Board, P. A. T. E. (2019). "Adult acute myeloid leukemia treatment (PDQR®))," in *PDQ Cancer Information Summaries [Internet]*, (Bethesda, MY: National Cancer Institute (US). Available online at: https://www.ncbi.nlm.nih.gov/books/NBK65996/

Brauner, S., Zhou, W., Backlin, C., Green, T. M., Folkersen, L., Ivanchenko, M., et al. (2015). Reduced expression of TRIM21/Ro52 predicts poor prognosis in diffuse large B-cell lymphoma patients with and without rheumatic disease. *J. Int. Med.* 278, 323–332. doi: 10.1111/joim.12375

Campos-Melo, D., Droppelmann, C. F., Volkening, K., Volkening, K. F., Strong, M. J., and Strong, M. J. (2014). RNA-binding proteins as molecular links between cancer and neurodegeneration. *Biogerontology* 15, 587–610. doi: 10.1007/s10522-014-9531-2

Chen, Y., Cairns, R. F., Papandreou, I., Papandreou, I. F., Koong, A., Koong, A. F., et al. (2009). Oxygen consumption can regulate the growth of tumors, a new perspective on the Warburg effect. *PLoS One* 4:e7033. doi: 10.1371/journal.pone.0007033

Clinton, C., and Gazda, H. T. (2009). *Diamond-Blackfan Anemia. BTI – GeneReviews(®)*. Available online at: https://www.ncbi.nlm.nih.gov/books/NBK7047/

Cornelissen, J. J., Versluis, J., Passweg, J. R., van Putten, W. L., Manz, M. G., Maertens, J., et al. (2015). Comparative therapeutic value of post-remission approaches in patients with acute myeloid leukemia aged 40-60 years. *Leukemia* 29, 1041–1050. doi: 10.1038/leu.2014.332

de Rooij, L., Chan, D. C. H., Keyvani Chahi, A., and Hope, K. J. (2019). Post-transcriptional regulation in hematopoiesis: RNA binding proteins take control (1). *Biochem. Cell Biol.* 97, 10–20. doi: 10.1139/bcb-2017-0310

Ding, Q., He, D., He, K., Zhang, Q., Tang, M., Dai, J., et al. (2015). Downregulation of TRIM21 contributes to hepatocellular carcinoma carcinogenesis and indicates poor prognosis of cancers. *Tumor Biol.* 36, 8761–8772. doi: 10.1007/s13277-015-3572-2

Döhner, H. F., Weisdorf, D. J., Weisdorf, D. F., Bloomfield, C. D., and Bloomfield, C. D. (2015). Acute myeloid leukemia. *New Engl. J. Med.* 373, 1136–1152. doi: 10.1056/NEJMra1406184

Döhner, H., Estey, E. F., Amadori, S., Amadori, S. F., Appelbaum, F. R., Appelbaum, F. F., et al. (2010). Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* 115, 453–474. doi: 10.1182/blood-2009-07-235358

Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Büchner, T., et al. (2017). Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 129, 424–447. doi: 10.1182/blood-2016-08-733196

Du, F., Qiao, C., Li, X., Chen, Z., Liu, H., Wu, S., et al. (2019). Forkhead box K2 promotes human colorectal cancer metastasis by upregulating ZEB1 and EGFR. *Theranostics* 9:3879. doi: 10.7150/thno.31716

Ebert, B. L., Pretz, J. F., Bosco, J., Bosco, J. F., Chang, C. Y., Chang, C. F., et al. (2008). Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* 451, 7335–7339. doi: 10.1038/nature06494

Espert, L., Eldin, P. F., Gongora, C., Gongora, C. F., Bayard, B., Bayard, B. F., et al. (2006). The exonuclease ISG20 mainly localizes in the nucleolus and the Cajal (Coiled) bodies and is associated with nuclear SMN protein-containing complexes. *J. Cell. Biochem.* 98, 1320–1333. doi: 10.1002/jcb.20869

Franceschini, A., Szklarczyk, D. F., Frankild, S., Frankild, S. F., Kuhn, M., Kuhn, M. F., et al. (2012). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094

Frans, G., Moens, L., Schaballie, H., Wuyts, G., Liston, A., Poesen, K., et al. (2017). Homozygous N-terminal missense mutation in TRNT1 leads to progressive B-cell immunodeficiency in adulthood. *J. Aller. Clin. Immunol.* 139, 360–363. doi: 10.1016/j.jaci.2016.06.050

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Statist. Softw.* 33:1. doi: 10.18637/jss.v033.i01

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. *Nat. Rev. Genet.* 15, 829–845. doi: 10.1038/nrg3813

Grimwade, D., Hills, R. F., Moorman, A. V., Moorman, A. F., Walker, H., Walker, H. F., et al. (2010). Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* 116, 354–365. doi: 10.1182/blood-2009-11-254441

Hao, J., Chiang, Y. T., Gout, P. W., and Wang, Y. (2016). Elevated XPO6 expression as a potential prognostic biomarker for prostate cancer recurrence. *Front. Biosci. (Schol. Ed.)* 8:44–55. doi: 10.2741/s445

He, H. A.-O., Bing, H., and Liu, G. (2018). TSR2 Induces laryngeal cancer cell apoptosis through inhibiting NF-κB signaling pathway. *Laryngoscope* 128, E130–E134. doi: 10.1002/lary.27035

Iasonos, A., Schrag, D. F., Raj, G. V., Raj, Gv F, Panageas, K. S., and Panageas, K. S. (2008). How to build and interpret a nomogram for cancer prognosis. *J. Clin. Oncol.* 26, 1364–1370. doi: 10.1200/JCO.2007.12.9791

Jin, Y., Zhang, Y., Li, B., Zhang, J., Dong, Z., Hu, X., et al. (2019). TRIM21 mediates ubiquitination of Snail and modulates epithelial to mesenchymal transition in breast cancer cells. *Int. J. Biol. Macromol.* 124, 846–853. doi: 10.1016/j.ijbiomac.2018.11.269

Katano-Toki, A., Satoh, T. F., Tomaru, T., Tomaru, T. F., Yoshino, S., Yoshino, S. F., et al. (2013). THRAP3 interacts with HELZ2 and plays a novel role in adipocyte differentiation. *Mol. Endocrinol.* 27, 769–780. doi: 10.1210/me.2012-1332

Kim, E. J., Kho, J. H., Kang, M. R., and Um, S. J. (2007). Active regulator of SIRT1 cooperates with SIRT1 and facilitates suppression of p53 activity. *Mol. Cell* 28, 277–290. doi: 10.1016/j.molcel.2007.08.030

Knight, J. R., Allison, S. J., and Milner, J. (2013). Active regulator of SIRT1 is required for cancer cell survival but not for SIRT1 activity. *Open Biol.* 3:130130. doi: 10.1098/rsob.130130

Komeno, Y., Huang, Y. J., Qiu, J., Lin, L., Xu, Y., Zhou, Y., et al. (2015). SRSF2 is essential for hematopoiesis, and its myelodysplastic syndrome-related mutations dysregulate alternative pre-mRNA splicing. *Mol. Cell. Biol.* 35, 3071–3082. doi: 10.1128/MCB.00202-15

Koromilas, A. E. (2015). Roles of the translation initiation factor eIF2α serine 51 phosphorylation in cancer formation and treatment. *Biochim. Biophys. Acta (BBA) Gene Regulat. Mech.* 1849, 1871–1880. doi: 10.1016/j.bbagrm.2014.12.007

Lagou, V., Garcia-Perez, J. E., Smets, I., Van Horebeek, L., Vandebergh, M., Chen, L., et al. (2018). Genetic architecture of adaptive immune system identifies key immune regulators. *Cell Rep.* 25, 798–810. doi: 10.1016/j.celrep.2018.09.048

Li, W., Gao, L. N., Song, P. P., and You, C. G. (2020). Development and validation of a RNA binding protein-associated prognostic model for lung adenocarcinoma. *Aging (Albany N.Y.)* 12:3558. doi: 10.18632/aging.102828

Lorson, C. L., Hahnen, E. F., Androphy, E. J., Androphy, E. F., Wirth, B., and Wirth, B. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6307–6311. doi: 10.1073/pnas.96.11.6307

Masuda, K., and Kuwano, Y. (2019). Diverse roles of RNA-binding proteins in cancer traits and their implications in gastrointestinal cancers. *Wiley Interdisc. Rev. RNA* 10:e1520. doi: 10.1002/wrna.1520

Mortera-Blanco, T., Dimitriou, M., Woll, P. S., Karimi, M., Elvarsdottir, E., Conte, S., et al. (2017). SF3B1-initiating mutations in MDS-RSs target lymphomyeloid hematopoietic stem cells. *Blood* 130, 881–890. doi: 10.1182/blood-2017-03-776070

Nguyen, J. Q., and Irby, R. B. (2017). TRIM21 is a novel regulator of Par-4 in colon and pancreatic cancer cells. *Cancer Biol. Ther.* 18, 16–25. doi: 10.1080/15384047.2016.1252880

Nishida, K., Kuwano, Y., Nishikawa, T., Masuda, K., and Rokutan, K. (2017). RNA binding proteins and genome integrity. *Int. J. Mol. Sci.* 18:1341. doi: 10.3390/ijms18071341

Pan, Y. A.-O., Tong, J. H. M., Kang, W. A.-O. X., Lung, R. W. M., Chak, W. P., Chung, L. Y., et al. (2018). EXOSC4 functions as a potential oncogene in development and progression of colorectal cancer. *Mol. Carcinog.* 57, 1780–1791. doi: 10.1002/mc.22896

Rajkumar, T., Sabitha, K. F., Vijayalakshmi, N., Vijayalakshmi, N. F., Shirley, S., Shirley, S. F., et al. (2011). Identification and validation of genes involved in cervical tumourigenesis. *BMC Cancer* 11:80. doi: 10.1186/1471-2407-11-80

Ravandi, F. (2013). Relapsed acute myeloid leukemia: why is there no standard of care? *Best Pract. Res. Clin. Haematol.* 26, 253–259. doi: 10.1016/j.beha.2013.10.005

Robinson, M. D., McCarthy, Dj F, Smyth, G. K., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616

Röllig, C., Bornhäuser, M. F., Thiede, C., Thiede, C. F., Taube, F., Taube, F. F., et al. (2011). Long-term prognosis of acute myeloid leukemia according to the new genetic risk classification of the European LeukemiaNet recommendations: evaluation of the proposed reporting system. *J. Clin. Oncol.* 29, 2758–2765. doi: 10.1200/JCO.2010.32.8500

Shirai, C. L., Ley, J. N., White, B. S., Kim, S., Tibbitts, J., Shao, J., et al. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer Cell* 27, 631–643. doi: 10.1016/j.ccell.2015.04.008

Simon, N., Friedman, J., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for cox's proportional hazards model via coordinate descent. *J. Statist. Softw.* 39:31. doi: 10.18637/jss.v039.i05

Singleton, D. C., and Harris, A. L. (2012). Targeting the ATF4 pathway in cancer therapy. *Exp. Opin. Therap. Targ.* 16, 1189–1202. doi: 10.1517/14728222.2012.728207

Slade, A. F., Kattini, R., Kattini, R. F., Campbell, C., Campbell, C., and Holcik, M. (2020). Diseases associated with defects in tRNA CCA addition. *Int. J. Mol. Sci.* 21:3780. doi: 10.3390/ijms21113780

Slovak, M. L., Kopecky, K. F., Cassileth, P. A., Cassileth, Pa F, Harrington, D. H., Harrington, Dh F, et al. (2000). Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *J. Am. Soc. Hematol.* 96, 4075–4083. doi: 10.1182/blood.V96.13.4075.h8004075_4075_4083

Stavraka, C., and Blagden, S. (2015). The la-related proteins, a family with connections to cancer. *Biomolecules* 5, 2701–2722. doi: 10.3390/biom5042701

Subramanian, A., Tamayo, P. F., Mootha, V. K., Mootha, V. F., Mukherjee, S., Mukherjee, S. F., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Therneau, T. M., and Grambsch, P. M. (2000). "The cox model," in *Modeling Survival Data: Extending the Cox Model*, (Berlin: Springer), 39–77. doi: 10.1007/978-1-4757-3294-8_3 Available online at: https://link.springer.com/chapter/10.1007%2F978-1-4757-3294-8_3

Van Tong, H., Hoan, N. X., Binh, M. T., Quyen, D. T., Meyer, C. G., Song, L. H., et al. (2018). Interferon-stimulated gene 20 kDa protein serum levels and clinical outcome of hepatitis B virus-related liver diseases. *Oncotarget* 9:27858. doi: 10.18632/oncotarget.25559

Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: a novel method for evaluating prediction models. *Med. Dec. Mak.* 26, 565–574. doi: 10.1177/0272989X06295361

Wang, E., Lu, S. X., Pastore, A., Chen, X., Imig, J., Chun-Wei Lee, S., et al. (2019). Targeting an RNA-binding protein network in acute myeloid leukemia. *Cancer Cell* 35, 369–384. doi: 10.1016/j.ccell.2019.01.010

Wang, K., Li, L., Fu, L., Yuan, Y., Dai, H., Zhu, T., et al. (2019). Integrated bioinformatics analysis the function of RNA binding proteins (RBPs) and their prognostic value in breast cancer. *Front. Pharmacol.* 10:140. doi: 10.3389/fphar.2019.00140

Wang, Z. L., Li, B., Luo, Y. X., Lin, Q., Liu, S. R., Zhang, X. Q., et al. (2018). Comprehensive genomic characterization of RNA-binding proteins across human cancers. *Cell Rep.* 22, 286–298. doi: 10.1016/j.celrep.2017.12.035

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Berlin: Springer. doi: 10.1007/978-3-319-24277-4

Xu, T., Ruan, H., Gao, S., Liu, J., Liu, Y., Song, Z., et al. (2020). ISG20 serves as a potential biomarker and drives tumor progression in clear cell renal cell carcinoma. *Aging (Albany N.Y.)* 12:1808. doi: 10.18632/aging.102714

Ye, J., Kumanova, M. F., Hart, L. S., Hart, L. F., Sloane, K., Sloane, K. F., et al. (2010). The GCN2-ATF4 pathway is critical for tumour cell survival and proliferation in response to nutrient deprivation. *EMBO J.* 29, 2082–2096. doi: 10.1038/emboj.2010.81

Yoshida, K., Bohn, J., and Yoshida, M. K. (2020). *Package 'tableone'*. Vienna: R Foundation for Statistical Computing.

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16, 284–287. doi: 10.1089/omi.2011.0118

Zhou, J., and Chng, W. J. (2014). Identification and targeting leukemia stem cells: the path to the cure for acute myeloid leukemia. *World J. Stem Cells* 6:473. doi: 10.4252/wjsc.v6.i4.473

Zhou, W., Zhang, Y., Zhong, C., Hu, J., Hu, H., Zhou, D., et al. (2018). Decreased expression of TRIM21 indicates unfavorable outcome and promotes cell growth in breast cancer. *Cancer management and research* 10, 3687. doi: 10.2147/CMAR.S175470

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

MINI REVIEW
published: 29 September 2021
doi: 10.3389/fgene.2021.757109

RNA Editing and Its Roles in Plant Organelles

Wei Hao[1†], Guoxiang Liu[2†], Weipeng Wang[3†], Wei Shen[4], Yuping Zhao[5], Jialiang Sun[3], Qiuyue Yang[3], Yaxin Zhang[3], Wenjia Fan[3], Shuaishuai Pei[3], Zhuanqing Chen[3], Dongbei Xu[6] and Tengfei Qin[3]**

[1] College of Medical Technology, Beihua University, Jilin City, China, [2] Key Laboratory of Tobacco Improvement and Biotechnology, Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao, China, [3] Henan Collaborative Innovation Center of Modern Biological Breeding, Henan Institute of Sciences and Technology, Xinxiang, China, [4] State Key Laboratory of Agrobiotechnology, School of Life Sciences, The Chinese University of Hong Kong, Hong Kong, SAR China, [5] Beijing City River and Lake Management Office, Beijing, China, [6] College of Agronomy, Sichuan Agricultural University, Chengdu, China

OPEN ACCESS

Edited by:
Yanqiang Li,
Harvard Medical School,
United States

Reviewed by:
Mingxia Zhao,
Kansas State University,
United States
Yanwen Yu,
Henan Agricultural University, China

***Correspondence:**
Tengfei Qin
qintengfeisam@163.com
Dongbei Xu
xudongbei2006@126.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *11 August 2021*
Accepted: *25 August 2021*
Published: *29 September 2021*

Citation:
Hao W, Liu G, Wang W, Shen W, Zhao Y, Sun J, Yang Q, Zhang Y, Fan W, Pei S, Chen Z, Xu D and Qin T (2021) RNA Editing and Its Roles in Plant Organelles. Front. Genet. 12:757109. doi: 10.3389/fgene.2021.757109

RNA editing, a vital supplement to the central dogma, yields genetic information on RNA products that are different from their DNA templates. The conversion of C-to-U in mitochondria and plastids is the main kind of RNA editing in plants. Various factors have been demonstrated to be involved in RNA editing. In this minireview, we summarized the factors and mechanisms involved in RNA editing in plant organelles. Recently, the rapid development of deep sequencing has revealed many RNA editing events in plant organelles, and we further reviewed these events identified through deep sequencing data. Numerous studies have shown that RNA editing plays essential roles in diverse processes, such as the biogenesis of chloroplasts and mitochondria, seed development, and stress and hormone responses. Finally, we discussed the functions of RNA editing in plant organelles.

Keywords: RNA editing, factors, mechanism, plant organelles, deep sequencing

INTRODUCTION

Transcribed RNAs need to undergo a series of processes, such as modification, splicing, and editing, to form mature RNAs. RNA editing, a phenomenon that changes genetic information through nucleotide insertion, deletion, or conversion in messenger RNAs of functional genes, is an important supplement to the central dogma (Keller et al., 1999). In 1986, Benne et al. reported that extra nucleotides were added to the frameshift *coxII* gene through the RNA editing process in trypanosome mitochondria (Benne et al., 1986). To date, different types of RNA editing have been widely discovered in divergent RNAs from diverse organisms, including humans, mice, zebrafish, and plants (Yan et al., 2018; Popitsch et al., 2020).

RNA editing can occur in different forms, including the conversion of cytidine (C) to uridine (U), U-to-C, and adenosine (A) to inosine (I), as well as insertions and deletions of U and insertions of guanosine (G; Small et al., 2020). C-to-U editing events have been largely reported in humans. This editing of apolipoprotein-B (apo-B) mRNA was found to be catalyzed by an enzyme complex. The complex contains the catalytic subunit, deaminase APOBEC1, and the subunit apobec-1 complementation factor for RNA binding (Mehta and Driscoll, 2002). A-to-I editing was found to be mediated by adenosine deaminase acting on RNA enzymes. Adenosine deaminase acting on RNA deaminases usually contain no more than three catalytic

double-stranded RNA binding domains localized to the N-terminus and a catalytic domain localized to the C-terminus. This type of editing occurred in protein-coding regions of a few genes, causing the recoding and subsequent functional alterations of those genes. Common A-to-I RNA editing targets elements within 5′ and 3′UTRs and introns (Nishikura, 2010). U-to-C RNA editing has also been found in several lycophytes and ferns (Knie et al., 2016). Moreover, insertions and deletions of U and insertions of G were reported to exist in protozoa and some viruses, respectively (Small et al., 2020).

Furthermore, RNA editing has been observed in the cell nucleus and cytosol, mitochondria, and plastids. RNA editing mainly occurs in the mitochondria and plastids in plants. The conversion of C-to-U is the main form of RNA editing in plants. This editing phenomenon was described in 1989 in wheat mitochondria (Covello and Gray, 1989). Two years later, Hoch et al. presented evidence that C-to-U editing was responsible for the conversion from a normal ACG codon to an AUG initiation codon in the mRNA of the maize plastome gene *rpl2*. They demonstrated that RNA editing could also occur within plastids. RNA editing phenomena have been widely reported in many plants. RNA editing events have already been observed in *Arabidopsis*, rice, wheat, tobacco, maize, and soybean (Wang et al., 2016; Rodrigues et al., 2020).

Many studies have shown that abnormal RNA editing results in impaired organelle biogenesis, decreased embryo and endosperm development, retarded plant growth, and poor adaptability to abiotic stresses (Yan et al., 2018). These findings indicate that RNA editing plays a vital role in the growth and development of plants. The process of RNA editing between mammals and plant organelles has many similarities. However, plant organelle RNA editing has certain characteristics and machinery. In this review, we briefly summarized the factors involved in plant organelle RNA editing and reviewed the mechanism of RNA editing in plant organelles. In addition, we reviewed RNA editing events identified in plant organelles through deep sequencing data and discussed the functions of RNA editing in plant organelles.

FACTORS INVOLVED IN PLANT ORGANELLE RNA EDITING

Many RNA editing factors have been found in plant organelles. The identification of different types of RNA editing factors broadens our understanding of the RNA editing complex.

Pentatricopeptide repeat (PPR) family proteins are site-specific editing factors that recognize and directly bind to the *cis*-element of the target sequences of RNA (Yan et al., 2018). In 2005, the first RNA editing factor in plant organelles, the PPR family protein CRR4, was reported (Kotera et al., 2005). Pentatricopeptide repeat proteins are nucleus encoded but function in either plastids or mitochondria. They have been discovered to have important effects on RNA editing, splicing, and cleavage (Barkan and Small, 2014). Based on the highly conserved domain modules that are essential for deamination of edited cytidine residues at the C-terminus, these proteins can be separated into extended (E) and DYW subtypes (Cheng et al., 2016). The last C-terminal motif of the PPR family proteins binds to specific nucleotides upstream of the editing site in plant organellar genomes.

Multiple organellar RNA editing factor (MORF) was the second RNA editing factor discovered in the mitochondria of angiosperms (Bentolila et al., 2012). MORF was also identified as an RNA editing factor interacting protein. These factors comprise a central conserved domain that is formed by a core of six β-sheets (Yan et al., 2017). However, the roles of MORF members diverged significantly, and mutation of individual MORFs led to distinct RNA editing deficiencies (Takenaka et al., 2012). Several MORFs were found to interact with PPR family proteins, implying that specific complexes are constructed at different RNA editing sites (Glass et al., 2015).

Organelle RNA recognition motif-containing (ORRM) proteins have been revealed to be critical for RNA editing in the organelles of angiosperms. Disruption of ORRM members can severely affect plant growth (Shi et al., 2017). There is an RNA recognition motif in ORRM proteins that enables them to bind to RNA (Hackett et al., 2017). In addition, additional domains of some ORRM members indicate unique roles for the individual proteins. Organelle RNA recognition motif proteins can associate with other RNA editing factors (e.g., ORRM3/4 with MORF8), suggesting the complexity of the plant RNA editing machinery (Shi et al., 2016b).

Protoporphyrinogen IX oxidase 1 (PPO1) catalyzes the conversion of protoporphyrinogen IX into protoporphyrin IX and has also been identified to function in the RNA editing process in plant plastids. A *ppo1* mutant showed defects in 18 plastid RNA targets, leading to the decreased accumulation of the NDH complex and decreased synthesis of chlorophyll (Zhang et al., 2014). PPO1 can interact with MORF proteins but was not found to interact with PPR factors in chloroplasts. Therefore, PPO1 was speculated to have effects on RNA editing efficiency by coordinating the association of MORF proteins in chloroplasts.

Organelle zinc finger 1 (OZ1) is a RanBP2-type zinc finger protein family member. It was observed to act as an important RNA editing factor by interacting with ORRM1 (Sun et al., 2015). The *oz1* mutant showed decreased editing efficiency at 16 sites and impairment at 14 sites in plastids and consequently exhibited a yellow phenotype, suggesting its vital role in plant development. Organelle zinc finger 1 can interact with other factors, including PPR, ORRM1, and MORF proteins, indicating, that is, has a function in the assembly of the plastid RNA editing complex (Sandoval et al., 2019).

NUWA is identified as a P-class PPR protein. NUWA can interact with SLO2 or CLB19 (Guillaumot et al., 2017). The *nuwa* mutants showed a decreased level of RNA editing at various sites in both the plastids and mitochondria of mature leaves. In addition to PPR proteins, NUWA proteins were found to be coimmunoprecipitated with MORF proteins *in vivo*, which revealed their collaboration with other factors in RNA editing (Bayer-Csaszar et al., 2017).

THE MECHANISM OF PLANT ORGANELLE RNA EDITING

In plastids of tobacco, editing an ACG codon of the *psbL* mRNA to an AUG codon caused translational initiation. A chimaeric RNA containing *psbL* deletion derivatives and kanamycin resistance genes was constructed to investigate RNA editing in transgenic plants. This plastid transformation technique revealed that 22 nucleotides are sufficient to direct the editing of *psbL* in the chloroplasts of tobacco. Using this approach, *cis*-elements were determined to be vital for RNA editing site recognition. Another *in vitro* study using extracts of tobacco chloroplasts also detected *cis*-elements that are essential for RNA editing (Shikanai, 2006). Furthermore, the *cis*-elements essential for RNA editing in plant mitochondria were also demonstrated. Using an *in vivo* technique, 16 nucleotides upstream and 6 nucleotides downstream were found to be required for the efficient editing of *coxII* mRNA (Farre et al., 2001). Another group of scientists also identified the *cis*-elements that were required for RNA editing using pea mitochondrial extract (Takenaka et al., 2004).

PPR factors function as site recognition factors in plant organelles. Moreover, PPR proteins could also bind to *cis*-elements specifically. Yin et al. identified the crystal structures of PPR10 in two different RNA states. They showed that PPR10 interacts with RNA in the 5′ to 3′ direction for the target single-stranded RNA (ssRNA; Yin et al., 2013). The PPR–RNA complex, together with other factors, such as ORRM proteins and MORF proteins, forms a higher ordered editosome. Multiple organellar RNA editing factor proteins can act as connectors to form heterodimers and homodimers and selectively interact with different PPR proteins. In *Arabidopsis* and maize, evidence has shown that ORRM proteins are involved in RNA editing. ORRM1 contains two truncated MORF domains and one RRM domain. *Arabidopsis* mutants (*orrm1*) and maize mutants (Zm-*orrm1*) lost 12 and 9 RNA editing sites, respectively (Sun et al., 2013). Three ORRM proteins, ORRM2, ORRM3, and ORRM4, were also found to be essential for RNA editing sites in mitochondria (Shi et al., 2016a).

RNA EDITING EVENTS IDENTIFIED IN PLANT ORGANELLES FROM TRANSCRIPTOME DATA

In recent years, with the development of deep sequencing technologies, various plant organellar genomes have been released, resulting in the surprising discovery of a number of novel RNA editing events.

Grimes et al. screened RNA C-to-U editing sites of the tobacco mitochondrial transcriptome by deep sequencing. In total, they identified 635 editing sites, of which 557 were found within protein-coding genes, 73 in noncoding regions, and five in tRNA genes (Grimes et al., 2014). Zheng et al. identified nearly 570 C-to-U editing site mitochondria-encoded ORFs in rice by Sanger sequencing and publicly available RNA-seq data. Among these identified editing sites, 85.41% were identified on one of the first two bases of a codon, thus altering the corresponding amino acid (Zheng et al., 2020). Edera et al. identified 10,217 editing sites within protein-coding genes in the mitochondria by analyzing publicly available RNA-seq data of 17 diverse angiosperms. They revealed that the majority of the RNA editing sites was conserved across these angiosperms except for some specific sites in different species (Edera et al., 2018).

Oldenkott et al. detected more than 3000 C-to-U RNA editing events by analyzing the transcriptome of the chloroplast of the lycophyte *Selaginella uncinata* (Oldenkott et al., 2014). Lin et al. detected 41 C-to-U editing sites within the transcripts of chloroplast genes of *Vigna radiata* using RNA-seq reads, 5 and 34 of which altered one of the first two nucleotides of a codon (Lin et al., 2015a). Chen et al. identified 137 editing sites by screening the whole plastid transcriptomes of moth orchids, among which 93 were novel edits and 79 were on protein-coding genes (Chen et al., 2017).

Using deep sequencing data, Zheng et al. explored the impact of temperature on the RNA editing process in grape organelles. They identified 627 and 122 RNA editing sites in mitochondria and chloroplasts, respectively. They found that the overall editing level and the expression level of most *PPR* genes were negatively correlated with temperature, suggesting decreased RNA editing efficiency at high temperatures. The authors demonstrated that RNA editing events were susceptible to environmental heat stress (Zhang et al., 2020).

FUNCTIONS OF PLANT ORGANELLE RNA EDITING

RNA editing events mainly occur at one of the first two positions of the codon, thus affecting the amino acids encoded in plant organelles. By using publicly available RNA-seq data, Edera et al. found that many RNA editing events caused altered amino acids across angiosperm evolution, and this mainly occurred by substituting editing sites with thymidines (Edera et al., 2018).

Some studies have revealed that mitochondrial RNA editing plays an essential role in various plants. Sung et al. reported that a PPR protein, SLOW GROWTH1 (SLO1), plays a vital role in RNA editing of *NADH dehydrogenase 4* (*nad4*) and *nad9* in the mitochondria of *Arabidopsis*. The *slo1* mutants displayed abnormal phenotypes, such as darker and shrunken seeds, late germination, and delayed development (Sung et al., 2010). Furthermore, another study showed that the mitochondrial editing factor SLO2 was involved in the electron transport chain of mitochondria and hormone and stress responses in *Arabidopsis* (Zhu et al., 2014). *Opaque and growth retardation 1* (*OGR1*) encodes a PPR protein in rice. Kim et al. generated an *ogr1* RNA editing mutant in rice and found that *OGR1* is required for mitochondrial RNA editing in rice and is essential for seed germination and the growth and development of plants (Kim et al., 2009). *Pentatricopeptide repeat 2263* (*PPR2263*)

encodes a DYW-subgroup PPR protein responsible for mitochondrial RNA editing in *nad5* and *cytochrome b* (*cob*). Sosso et al. identified the *ppr2263* mutation in maize and found that the mutants exhibited editing defects at the *nad5*-1,550 and *cob*-908 sites. In addition, the authors demonstrated that mitochondrial RNA editing is vital for mitochondrial biogenesis and the growth in maize (Sosso et al., 2012). Another group of scientists showed that the PPR-DYW protein empty pericarp5 is required for mitochondrial RNA editing and seed development in maize (Liu et al., 2013).

Furthermore, abnormal RNA editing could also have adverse effects on the biogenesis of chloroplasts. In *Arabidopsis*, *yellow seedlings 1* (*YS1*) encodes a PPR protein that is localized to chloroplasts and contains a DYW motif. Zhou et al. found that YS1 is essential for the editing of the housekeeping gene *rpoB*. During the early stage, YS1 is required for the differentiation of chloroplasts (Zhou et al., 2009). In rice, *albino seedling lethality 3* (*ASL3*) encodes a chloroplast-localized PPR protein and contains 10 tandem PPR motifs. Lin et al. revealed that interruption of *ASL3* disturbed the transcriptional levels of genes involved in chloroplast development and photosynthesis and caused impaired development of chloroplasts and the growth of seedlings (Lin et al., 2015b).

CONCLUSION AND PERSPECTIVES

RNA editing has aroused great interest because it is a critical supplement to the central dogma. It has been more than 30 years since the RNA editing phenomenon was discovered. Various reactions are required for the RNA editing process in plant organelles. To date, some *cis*-elements and different kinds of editing factors have been recognized in plant organelles. These factors are responsible for editing events, including the most common forms of C-to-U and A-to-I conversions. These editing factors are thought to act together with one another to form an RNA editosome. However, there are still many undiscovered *cis*-elements and RNA editing factors. Furthermore, the significance of individual factors and their organization within the editosome is still largely unknown. The mechanism of plant organelle RNA editing needs more exploration in the future.

In recent years, the rapid development of deep sequencing has led to the discovery of numerous RNA editing events in plant organelles. Nevertheless, RNA editing events are not well annotated in some databases, such as DDBJ, GenBank, and ENA. However, some specialized bioinformatics resources have been developed for an appropriate description of plant RNA editing events. Plant RNA editing events are annotated in some databases, including CloroplastDB for plant chloroplast genomes, GOBASE for organellar genomes, PREPACT 3.0, which integrates information on characterized editing factors (Lenz et al., 2018), and REDIdb 3.0, which also shows the changes in amino acids caused by RNA editing events in functional domains and protein secondary structures (Lo Giudice et al., 2018). These databases also facilitate discoveries and functional studies of RNA editing sites in both plant mitochondria and plastids.

AUTHOR CONTRIBUTIONS

TQ and DX contributed equally to the design and coordination of the study. WH, GL, and WW collected the data with the help from the WS, YZ, JS, QY, YZ, WF, SP, and ZC wrote the manuscript. All of the authors reviewed and edited the manuscript.

FUNDING

This research was supported by Science and Technology Program of Sichuan Province (2020YJ0406), National Natural Science Foundation of China (31900256), Natural Science Foundation of Shandong Province (ZR2020QC026), Modern Agricultural Industry Technical Economic Evaluation System Green Development Position of Henan Province, Soft Science Project of Henan Province (202400410185) and State Key Laboratory of Cotton Biology Open Fund.

REFERENCES

Barkan, A., and Small, I. (2014). Pentatricopeptide repeat proteins in plants. *Annu. Rev. Plant Biol.* 65, 415–442. doi: 10.1146/annurev-arplant-050213-040159

Bayer-Csaszar, E., Haag, S., Jorg, A., Glass, F., Hartel, B., Obata, T., et al. (2017). The conserved domain in MORF proteins has distinct affinities to the PPR and E elements in PPR RNA editing factors. *Biochim. Biophys. Acta Gene Regul. Mech.* 1860, 813–828. doi: 10.1016/j.bbagrm.2017.05.004

Benne, R., van de Burg, J., Brakenhoff, J. P., Sloof, P., van Boom, J. H., and Tromp, M. C. (1986). Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46, 819–826. doi: 10.1016/0092-8674(86)90063-2

Bentolila, S., Heller, W. P., Sun, T., Babina, A. M., Friso, G., van Wijk, K. J., et al. (2012). RIP1, a member of an Arabidopsis protein family, interacts with the protein RARE1 and broadly affects RNA editing. *Proc. Natl. Acad. Sci. U. S. A.* 109, E1453–E1461. doi: 10.1073/pnas.1121465109

Chen, T. C., Liu, Y. C., Wang, X., Wu, C. H., Huang, C. H., and Chang, C. C. (2017). Whole plastid transcriptomes reveal abundant RNA editing sites and differential editing status in *Phalaenopsis aphrodite* subsp. *formosana*. *Bot. Stud.* 58:38. doi: 10.1186/s40529-017-0193-7

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., et al. (2016). Redefining the structural motifs that determine RNA binding and RNA editing by pentatricopeptide repeat proteins in land plants. *Plant J.* 85, 532–547. doi: 10.1111/tpj.13121

Covello, P. S., and Gray, M. W. (1989). RNA editing in plant mitochondria. *Nature* 341, 662–666. doi: 10.1038/341662a0

Edera, A. A., Gandini, C. L., and Sanchez-Puerta, M. V. (2018). Towards a comprehensive picture of C-to-U RNA editing sites in angiosperm mitochondria. *Plant Mol. Biol.* 97, 215–231. doi: 10.1007/s11103-018-0734-9

Farre, J. C., Leon, G., Jordana, X., and Araya, A. (2001). Cis recognition elements in plant mitochondrion RNA editing. *Mol. Cell. Biol.* 21, 6731–6737. doi: 10.1128/MCB.21.20.6731-6737.2001

Glass, F., Hartel, B., Zehrmann, A., Verbitskiy, D., and Takenaka, M. (2015). MEF13 requires MORF3 and MORF8 for RNA editing at eight targets in mitochondrial mRNAs in *Arabidopsis thaliana*. *Mol. Plant* 8, 1466–1477. doi: 10.1016/j.molp.2015.05.008

Grimes, B. T., Sisay, A. K., Carroll, H. D., and Cahoon, A. B. (2014). Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed

ORFs and numerous editing sites outside coding regions. *BMC Genomics* 15:31. doi: 10.1186/1471-2164-15-31

Guillaumot, D., Lopez-Obando, M., Baudry, K., Avon, A., Rigaill, G., Falcon de Longevialle, A., et al. (2017). Two interacting PPR proteins are major *Arabidopsis* editing factors in plastid and mitochondria. *Proc. Natl. Acad. Sci. U. S. A.* 114, 8877–8882. doi: 10.1073/pnas.1705780114

Hackett, J. B., Shi, X., Kobylarz, A. T., Lucas, M. K., Wessendorf, R. L., Hines, K. M., et al. (2017). An organelle RNA recognition motif protein is required for photosystem II subunit psbF transcript editing. *Plant Physiol.* 173, 2278–2293. doi: 10.1104/pp.16.01623

Keller, W., Wolf, J., and Gerber, A. (1999). Editing of messenger RNA precursors and of tRNAs by adenosine to inosine conversion. *FEBS Lett.* 452, 71–76. doi: 10.1016/S0014-5793(99)00590-6

Kim, S. R., Yang, J. I., Moon, S., Ryu, C. H., An, K., Kim, K. M., et al. (2009). Rice OGR1 encodes a pentatricopeptide repeat-DYW protein and is essential for RNA editing in mitochondria. *Plant J.* 59, 738–749. doi: 10.1111/j.1365-313X.2009.03909.x

Knie, N., Grewe, F., Fischer, S., and Knoop, V. (2016). Reverse U-to-C editing exceeds C-to-U RNA editing in some ferns - a monilophyte-wide comparison of chloroplast and mitochondrial RNA editing suggests independent evolution of the two processes in both organelles. *BMC Evol. Biol.* 16:134. doi: 10.1186/s12862-016-0707-z

Kotera, E., Tasaka, M., and Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature* 433, 326–330. doi: 10.1038/nature03229

Lenz, H., Hein, A., and Knoop, V. (2018). Plant organelle RNA editing and its specificity factors: enhancements of analyses and new database features in PREPACT 3.0. *BMC Bioinf.* 19:255. doi: 10.1186/s12859-018-2244-9

Lin, D., Gong, X., Jiang, Q., Zheng, K., Zhou, H., Xu, J., et al. (2015b). The rice ALS3 encoding a novel pentatricopeptide repeat protein is required for chloroplast development and seedling growth. *Rice* 8:17. doi: 10.1186/s12284-015-0050-9

Lin, C. P., Ko, C. Y., Kuo, C. I., Liu, M. S., Schafleitner, R., and Chen, L. F. (2015a). Transcriptional slippage and RNA editing increase the diversity of transcripts in chloroplasts: insight from deep sequencing of *Vigna radiata* genome and transcriptome. *PLoS One* 10:e0129396. doi: 10.1371/journal.pone.0129396

Liu, Y. J., Xiu, Z. H., Meeley, R., and Tan, B. C. (2013). Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. *Plant Cell* 25, 868–883. doi: 10.1105/tpc.112.106781

Lo Giudice, C., Pesole, G., and Picardi, E. (2018). REDIdb 3.0: a comprehensive collection of RNA editing events in plant organellar genomes. *Front. Plant Sci.* 9:482. doi: 10.3389/fpls.2018.00482

Mehta, A., and Driscoll, D. M. (2002). Identification of domains in apobec-1 complementation factor required for RNA binding and apolipoprotein-B mRNA editing. *RNA* 8, 69–82. doi: 10.1017/S1355838202015649

Nishikura, K. (2010). Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* 79, 321–349. doi: 10.1146/annurev-biochem-060208-105251

Oldenkott, B., Yamaguchi, K., Tsuji-Tsukinoki, S., Knie, N., and Knoop, V. (2014). Chloroplast RNA editing going extreme: more than 3400 events of C-to-U editing in the chloroplast transcriptome of the lycophyte *Selaginella uncinata*. *RNA* 20, 1499–1506. doi: 10.1261/rna.045575.114

Popitsch, N., Huber, C. D., Buchumenski, I., Eisenberg, E., Jantsch, M., von Haeseler, A., et al. (2020). A-to-I RNA editing uncovers hidden signals of adaptive genome evolution in animals. *Genome Biol. Evol.* 12, 345–357. doi: 10.1093/gbe/evaa046

Rodrigues, N. F., Nogueira, F. C. S., Domont, G. B., and Margis, R. (2020). Identification of soybean trans-factors associated with plastid RNA editing sites. *Genet. Mol. Biol.* 43:e20190067. doi: 10.1590/1678-4685-gmb-2019-0067

Sandoval, R., Boyd, R. D., Kiszter, A. N., Mirzakhanyan, Y., Santibanez, P., Gershon, P. D., et al. (2019). Stable native RIP9 complexes associate with C-to-U RNA editing activity, PPRs, RIPs, OZ1, ORRM1 and ISE2. *Plant J.* 99, 1116–1126. doi: 10.1111/tpj.14384

Shi, X., Bentolila, S., and Hanson, M. R. (2016a). Organelle RNA recognition motif-containing (ORRM) proteins are plastid and mitochondrial editing factors in *Arabidopsis*. *Plant Signal. Behav.* 11:e1167299. doi: 10.1080/15592324.2016.1167299

Shi, X., Castandet, B., Germain, A., Hanson, M. R., and Bentolila, S. (2017). ORRM5, an RNA recognition motif-containing protein, has a unique effect on mitochondrial RNA editing. *J. Exp. Bot.* 68, 2833–2847. doi: 10.1093/jxb/erx139

Shi, X., Germain, A., Hanson, M. R., and Bentolila, S. (2016b). RNA recognition motif-containing protein ORRM4 broadly affects mitochondrial RNA editing and impacts plant development and flowering. *Plant Physiol.* 170, 294–309. doi: 10.1104/pp.15.01280

Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cell. Mol. Life Sci.* 63, 698–708. doi: 10.1007/s00018-005-5449-9

Small, I. D., Schallenberg-Rudinger, M., Takenaka, M., Mireau, H., and Ostersetzer-Biran, O. (2020). Plant organellar RNA editing: what 30 years of research has revealed. *Plant J.* 101, 1040–1056. doi: 10.1111/tpj.14578

Sosso, D., Mbelo, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., et al. (2012). PPR2263, a DYW-subgroup pentatricopeptide repeat protein, is required for mitochondrial nad5 and cob transcript editing, mitochondrion biogenesis, and maize growth. *Plant Cell* 24, 676–691. doi: 10.1105/tpc.111.091074

Sun, T., Germain, A., Giloteaux, L., Hammani, K., Barkan, A., Hanson, M. R., et al. (2013). An RNA recognition motif-containing protein is required for plastid RNA editing in *Arabidopsis* and maize. *Proc. Natl. Acad. Sci. U. S. A.* 110, E1169–E1178. doi: 10.1073/pnas.1220162110

Sun, T., Shi, X., Friso, G., Van Wijk, K., Bentolila, S., and Hanson, M. R. (2015). A zinc finger motif-containing protein is essential for chloroplast RNA editing. *PLoS Genet.* 11:e1005028. doi: 10.1371/journal.pgen.1005028

Sung, T. Y., Tseng, C. C., and Hsieh, M. H. (2010). The SLO1 PPR protein is required for RNA editing at multiple sites with similar upstream sequences in *Arabidopsis* mitochondria. *Plant J.* 63, 499–511. doi: 10.1111/j.1365-313X.2010.04258.x

Takenaka, M., Neuwirt, J., and Brennicke, A. (2004). Complex cis-elements determine an RNA editing site in pea mitochondria. *Nucleic Acids Res.* 32, 4137–4144. doi: 10.1093/nar/gkh763

Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Hartel, B., and Brennicke, A. (2012). Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. U. S. A.* 109, 5104–5109. doi: 10.1073/pnas.1202452109

Wang, M., Liu, H., Ge, L., Xing, G., Wang, M., Weining, S., et al. (2016). Identification and analysis of RNA editing sites in the chloroplast transcripts of *Aegilops tauschii* L. *Gen. Dent.* 8:13. doi: 10.3390/genes8010013

Yan, J., Zhang, Q., Guan, Z., Wang, Q., Li, L., Ruan, F., et al. (2017). MORF9 increases the RNA-binding activity of PLS-type pentatricopeptide repeat protein in plastid RNA editing. *Nat. Plants* 3:17037. doi: 10.1038/nplants.2017.37

Yan, J., Zhang, Q., and Yin, P. (2018). RNA editing machinery in plant organelles. *Sci. China Life Sci.* 61, 162–169. doi: 10.1007/s11427-017-9170-3

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. *Nature* 504, 168–171. doi: 10.1038/nature12651

Zhang, A., Jiang, X., Zhang, F., Wang, T., and Zhang, X. (2020). Dynamic response of RNA editing to temperature in grape by RNA deep sequencing. *Funct. Integr. Genomics* 20, 421–432. doi: 10.1007/s10142-019-00727-7

Zhang, F., Tang, W., Hedtke, B., Zhong, L., Liu, L., Peng, L., et al. (2014). Tetrapyrrole biosynthetic enzyme protoporphyrinogen IX oxidase 1 is required for plastid RNA editing. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2023–2028. doi: 10.1073/pnas.1316183111

Zheng, P., Wang, D., Huang, Y., Chen, H., Du, H., and Tu, J. (2020). Detection and analysis of C-to-U RNA editing in rice mitochondria-encoded ORFs. *Plan. Theory* 9:1277. doi: 10.3390/plants9101277

Zhou, W., Cheng, Y., Yap, A., Chateigner-Boutin, A. L., Delannoy, E., Hammani, K., et al. (2009). The *Arabidopsis* gene YS1 encoding a DYW protein is required for editing of rpoB transcripts and the rapid development of chloroplasts during early growth. *Plant J.* 58, 82–96. doi: 10.1111/j.1365-313X.2008.03766.x

Zhu, Q., Dugardeyn, J., Zhang, C., Muhlenbock, P., Eastmond, P. J., Valcke, R., et al. (2014). The *Arabidopsis thaliana* RNA editing factor SLO2, which affects the mitochondrial electron transport chain, participates in multiple stress and hormone responses. *Mol. Plant* 7, 290–310. doi: 10.1093/mp/sst102

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 15 October 2021
doi: 10.3389/fgene.2021.726369

N6-Methyladenosine-Related Long Non-coding RNA Signature Associated With Prognosis and Immunotherapeutic Efficacy of Clear-Cell Renal Cell Carcinoma

Tianming Ma[1,2], *Xiaonan Wang*[2,3], *Jiawen Wang*[1,2], *Xiaodong Liu*[1,2], *Shicong Lai*[1,2], *Wei Zhang*[1], *Lingfeng Meng*[1,2], *Zijian Tian*[1,2] *and Yaoguang Zhang*[1,2]*

[1]*Department of Urology, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China,* [2]*Graduate School of Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China,* [3]*Department of Radiology, Beijing Hospital, National Center of Gerontology, Institute of Geriatric Medicine, Chinese Academy of Medical Sciences, Beijing, China*

OPEN ACCESS

Edited by:
Jia Meng,
Xi'an Jiaotong-Liverpool University, China

Reviewed by:
Da Huang,
Shanghai Jiao Tong University School of Medicine, China
Yao Lin,
Fujian University of Traditional Chinese Medicine, China

***Correspondence:**
Yaoguang Zhang
zhang003887@sina.com

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *16 June 2021*
Accepted: *30 September 2021*
Published: *15 October 2021*

Citation:
Ma T, Wang X, Wang J, Liu X, Lai S, Zhang W, Meng L, Tian Z and Zhang Y (2021) N6-Methyladenosine-Related Long Non-coding RNA Signature Associated With Prognosis and Immunotherapeutic Efficacy of Clear-Cell Renal Cell Carcinoma. Front. Genet. 12:726369. doi: 10.3389/fgene.2021.726369

Increasing evidence suggests that N6-methyladenosine (m6A) and long non-coding RNAs (lncRNAs) play important roles in cancer progression and immunotherapeutic efficacy in clear-cell renal cell carcinoma (ccRCC). In this study, we conducted a comprehensive ccRCC RNA-seq analysis using The Cancer Genome Atlas data to establish an m6A-related lncRNA prognostic signature (m6A-RLPS) for ccRCC. Forty-four prognostic m6A-related lncRNAs (m6A-RLs) were screened using Pearson correlation analysis ($|R| > 0.7$, $p < 0.001$) and univariable Cox regression analysis ($p < 0.01$). Using consensus clustering, the patients were divided into two clusters with different overall survival (OS) rates and immune status according to the differential expression of the lncRNAs. Gene set enrichment analysis corroborated that the clusters were enriched in immune-related activities. Twelve prognostic m6A-RLs were selected and used to construct the m6A-RLPS through least absolute shrinkage and selection operator Cox regression. We validated the differential expression of the 12 lncRNAs between tumor and non-cancerous samples, and the expression levels of four m6A-RLs were further validated using Gene Expression Omnibus data and Lnc2Cancer 3.0 database. The m6A-RLPS was verified to be an independent and robust predictor of ccRCC prognosis using univariable and multivariable Cox regression analyses. A nomogram based on age, tumor grade, clinical stage, and m6A-RLPS was generated and showed high accuracy and reliability at predicting the OS of patients with ccRCC. The prognostic signature was found to be strongly correlated to tumor-infiltrating immune cells and immune checkpoint expression. In conclusion, we established a novel m6A-RLPS with a favorable prognostic value for patients with ccRCC. The 12 m6A-RLs included in the signature may provide new

Abbreviations: AUC, area under curve; ccRCC, clear-cell renal cell carcinoma; C-index, concordance index; DSS, disease-specific survival; GSEA, gene set enrichment analysis; ICI, immune-checkpoint inhibitor; lncRNA, long non-coding RNA; m6A-RL, m6A-related lncRNA; m6A-RLPS, m6A-related lncRNA prognostic signature; OS, overall survival; RNA-seq, RNA-sequencing; ROC, receiver operating-characteristic; TIC, tumor-infiltrating immune cell; TME, tumor microenvironment; TKI, tyrosine kinase inhibitor.

insights into the tumorigenesis and allow the prediction of the treatment response of ccRCC.

Keywords: clear-cell renal cell carcinoma, N6-methyladenosine, lncRNA, prognostic signature, immune infiltration

INTRODUCTION

Kidney cancer, of which renal cell carcinoma (RCC) is the major histological type, is a common type of cancer with a high mortality rate and steadily rising morbidity rate worldwide (Bray et al., 2018). In 2020, there were approximately 73,750 new cases of kidney cancer and approximately 14,830 deaths due to kidney cancer in United States (Siegel et al., 2020). Clear-cell (cc)RCC is the most common subtype of RCC, accounting for more than 70% of cases (Rini et al., 2009). Although surgery is an effective treatment for early-stage ccRCC, relapse or metastasis after curative treatment occurs in 30% of cases. Unfortunately, the prognosis for patients with metastasis is generally dismal because ccRCC is insensitive to conventional radiotherapy and chemotherapy, even to tyrosine kinase inhibitors (TKIs) and immune-checkpoint inhibitors (ICIs) (Ljungberg et al., 2019; Mori et al., 2021). Therefore, it is imperative to elucidate the mechanisms underlying ccRCC development to identify effective treatments.

N6-methyladenosine (m6A), a reversible and abundant modification on mRNAs and non-coding RNAs, has been demonstrated to affect various aspects of RNA metabolism, including splicing, stability, nuclear export, and translation (Huang et al., 2020). Several studies support the hypothesis that aberrant expression of m6A regulators, i.e., "writers" (methyltransferases), "readers" (binding proteins), and "erasers" (demethylases), potentially participates in carcinogenesis, cancer development, or tumor-suppressive activities in various types of cancer, including ccRCC (Wang Q. et al., 2020; Huang et al., 2020). For instance, high expression of methyltransferase 3 (METTL3), an m6A methyltransferase, reportedly promotes the progression of gastric cancer, lung cancer, and prostate cancer (Cai et al., 2019; Wang T. et al., 2020; Wanna-Udom et al., 2020). In contrast, downregulation of METTL3 in endometrial cancer and decreased the expression of alkB homolog 5 (ALKBH5) and fat mass and obesity-associated protein (FTO), which are both "erasers," in ccRCC, has been suggested to be related to tumorigenicity and dismal prognosis (Liu et al., 2018; Strick et al., 2020).

Long non-coding RNAs (lncRNAs) also actively regulate various biological processes, including tumorigenesis, proliferation, and immunity (Peng et al., 2017; Taniue and Akimitsu, 2021). In ccRCC, lncRNA SNHG1 modulates immune escape by targeting miR-129-3p to activate STAT3 and PD-L1 (Tian et al., 2021), and LINC00641 facilitates ccRCC progression by sponging miR-340-5p (Zhang et al., 2021). Abundant evidence supports that m6A modifications and lncRNAs interactively affect the growth and development of various cancers (Wang X. et al., 2020; Huang et al., 2020; Yang et al., 2020). However, few studies have investigated the exact mechanism by which m6A modifications are involved in lncRNA-dependent ccRCC progression and prognosis. Recently, several prognostic signatures for ccRCC based on m6A regulators or lncRNAs alone have been identified (Zeng et al., 2019; Chen et al., 2020). However, to the best of our knowledge, an accurate and applicable prognostic signature based on m6A-related lncRNAs for patients with ccRCC has not been formulated. Therefore, we aimed to investigate the prognostic significance of m6A-related lncRNAs (m6A-RLs), and to develop an m6A-related lncRNA prognostic signature (m6A-RLPS) to predict the survival outcomes in patients with ccRCC.

MATERIALS AND METHODS

Data Sources

We collected transcriptome RNA-sequencing (RNA-seq) data of 539 ccRCC samples and 72 adjacent non-tumor samples with corresponding clinical data from the Cancer Genome Atlas (TCGA) database. Detailed clinicopathological information of the patients is provided in **Supplementary Table S1**. Patients with incomplete information were excluded. Expression matrices of 23 m6A-related genes (*METTL3*, *METTL14*, *METTL16*, *WTAPI*, *VIRMA*, *ZC3H13*, *RBM15*, *RBM15B*, *YTHDC1*, *YTHDC2*, *YTHDF1*, *YTHDF2*, *YTHDF3*, *HNRNPC*, *FMR1*, *LRPPRC*, *HNRNPA2B1*, *IGFBP1*, *IGFBP2*, *IGFBP3*, *RBMX*, *FTO*, and *ALKBH5*) were also obtained according to recent publications (Deng et al., 2018; Chen et al., 2019). To validate the expression of target lncRNAs, we extracted gene expression profile data from four datasets with more than 60 ccRCC samples in the Gene Expression Omnibus (GEO) database, including GSE17895 (tumor: n = 151, normal: n = 9), GSE40435 (tumor: n = 101, normal: n = 101), GSE46699 (tumor: n = 67, normal: n = 63), and GSE53757 (tumor: n = 72, normal: n = 72) (Dalgliesh et al., 2010; Wozniak et al., 2013; Eckel-Passow et al., 2014; von Roemeling et al., 2014). In addition, an updated Lnc2Cancer 3.0 (http://bio-bigdata.hrbmu.edu.cn/lnc2cancer) database, which hosts comprehensive data on experimentally supported lncRNAs and circular RNAs associated with human cancers (Gao et al., 2021), was used to analyze lncRNA expression further.

m6A-Related lncRNA Identification

The lncRNA profile obtained from TCGA was first screened against the human reference genome (GRCh38. p12; https://www.ncbi.nlm.nih.gov/genome). Pearson correlation analysis was used to identify m6A-RLs ($|\text{Pearson's R}| > 0.7$, $p < 0.001$). Then, univariable Cox regression analysis was performed to determine prognostic m6A-RLs related to the overall survival (OS) ($p < 0.01$). The Wilcoxon rank-sum test was utilized to compare the expression levels (visualized using heatmaps) of the prognostic m6A-RLs between tumor and normal tissues.

Consensus Clustering Analysis

To explore the expression features of prognostic m6A-RLs further, we clustered the ccRCC samples into different groups according to the differential expression of the lncRNAs using the ConsensusClusterPlus R package (Chen et al., 2020). Survival analysis and the chi-square test or Fisher's exact test were used to compare the survival rates between the clusters and to determine the relationships between the clinicopathological parameters and the clusters. Heatmaps were created using the "pheatmap" R package to visualize differential expression of the m6A-RLs and clinicopathological parameters in the different groups (Li, 2012).

Gene Set Enrichment Analysis and Tumor-Infiltrating Immune Cell Profiling

GSEA was conducted using the Hallmark, C2 KEGG v.7.1, C5 GO, and C7 v.6.2 gene sets in GSEA v.4.0.3 (http://www.broadinstitute.org/gsea) to investigate potential m6A-RL pathways and functions. The ESTIMATE algorithm was used to determine immune, stromal, and ESTIMATE scores for each sample. These scores represent the ratios of immune and stromal components and the total proportion of these components in the tumor microenvironment (TME) (Yoshihara et al., 2013). The CIBERSORT algorithm was used to estimate the abundance of 22 tumor-infiltrating immune cells (TICs) (Chen et al., 2018). Only tumor samples with $p < 0.05$ were retained for subsequent analysis.

Analysis of the Associations Between Clusters and Tumor-Infiltrating Immune Cells

The Wilcoxon rank-sum test was used to determine the relationships between TICs in the different clusters. Some common immune checkpoints, including CTLA-4, LAG-3, HAVCR2 (TIM-3), PDCD1 (PD-1), and TIGHT, were screened out to compare differences in expression between tumor and normal tissues and in the differential clusters, as well as to investigate the correlations between the expression levels of the genes and hub m6A-RLs.

Establishment and Validation of the m6A-Related lncRNA Prognostic Signature

The entire cohort was randomly divided into training and first validation groups at a 1:1 ratio. Thereafter, all the samples were randomly divided into the second and third validation cohorts at a ratio of 3:7. The training cohort was used to construct the m6A-RLPS to predict the prognosis of ccRCC patients. The selected m6A-related candidate lncRNAs mentioned above were subjected to least absolute shrinkage and selection operator (LASSO) Cox regression analysis for establishing an m6A-RLPS for ccRCC (Friedman et al., 2010). A risk score was then calculated for each patient using the following formula: risk score = $\sum_{i=1}^{n} Coef_i * Expr_i$, where $Coef_i$ is the coefficient and $Expr_i$ is the expression value of the *i*th lncRNA in the signature. Univariable Cox regression and Kaplan–Meier survival analyses were used to analyze the prognostic significance of the selected lncRNAs and their co-expressed m6A genes. The Wilcoxon rank-sum test was used to investigate differential expression of the selected lncRNAs and co-expressed m6A genes between tumor and non-cancerous samples in TCGA, which was further validated using GEO datasets and Lnc2Cancer 3.0.

Next, patients were classified into low- and high-risk groups according to the median risk score. The OS and disease-specific survival (DSS) were compared between the groups based on Kaplan–Meier curves. Time-dependent receiver operating-characteristic (ROC) curves with area under curve (AUC) values for the 1-, 3-, and 5-years OS and DSS rates were used to estimate the prognostic prediction efficacy of the m6A-RLPS. Heatmaps were generated to visualize the differential expression of the prognostic m6A-RLs in the low- and high-risk groups. Univariable and multivariable Cox regression analyses were used to investigate the independent predictive value of the risk score and clinicopathological parameters for the survival of patients with ccRCC. Survival analysis was conducted to further elucidate the prognostic ability of the risk score in subgroups stratified by age, sex, grade, clinical stage, and T stage. A nomogram was built based on the multivariate Cox regression analysis results and was comprehensively evaluated using the concordance index (C-index) and calibration curves.

Correlation Between the m6A-Related lncRNA Prognostic Signature and Clinicopathological Parameters

Associations between the clinicopathological parameters, immune score, clusters, and m6A-RLPS risk levels were assessed using the chi-square test or Fisher's exact test (shown in heatmaps). Student's *t*-test was used to determine the relationships between the risk scores and clinicopathological parameters, including age, sex, grade, clinical stage, T stage, cluster, immune score, stromal score, and ESTIMATE score. Additionally, survival analysis was employed to explore associations between all these factors and patient OS.

Correlation Between the m6A-Related lncRNA Prognostic Signature and Immune-Related Features

Correlations between the m6A-RLPS and immune cells were evaluated using the Wilcoxon rank-sum test and Spearman correlation analysis. In addition, we downloaded data of different cancer cell lines from the NCI-60 database (https://discover.nci.nih.gov/cellminer/home.do). Thereafter, the m6A-RLPS was also comprehensively analyzed to determine its relationship with the expression of some immune checkpoint proteins and drug sensitivity of some TKIs involved in ccRCC.

Statistical Analysis

All analyses were performed using R v.4.0.3 (http://www.R-project.org). The Wilcoxon rank-sum test was used to

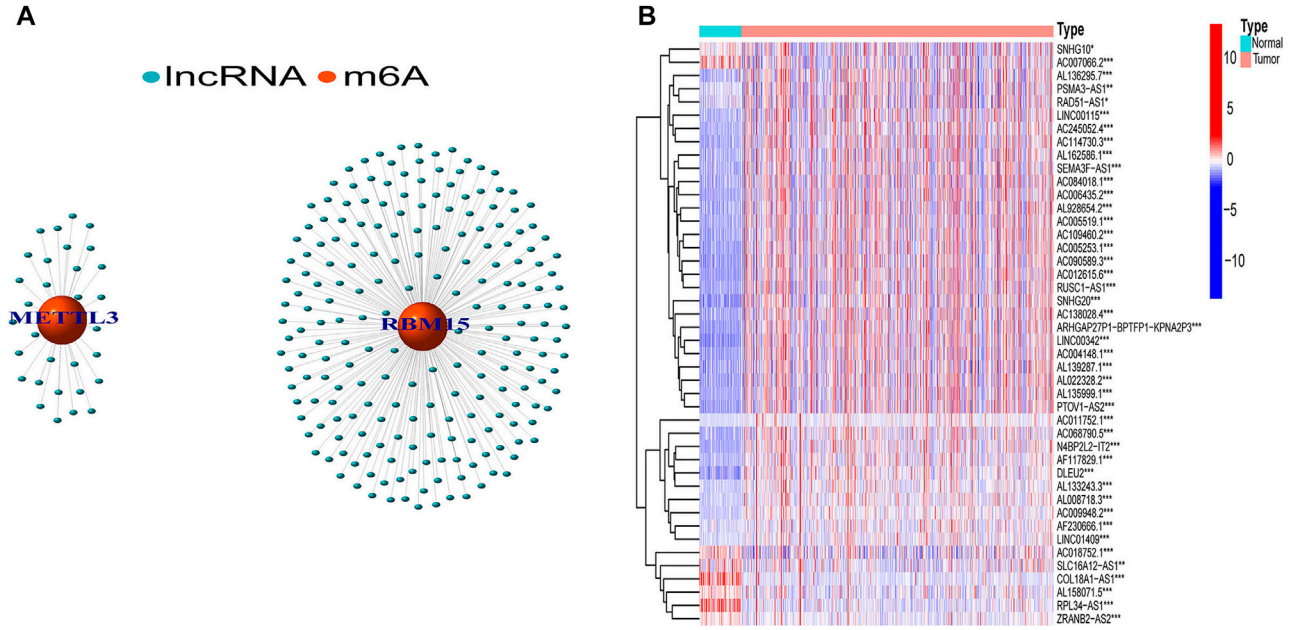

FIGURE 1 | Screening of hub lncRNAs. **(A)** Network of the 23 selected m6A-related genes and their associated lncRNAs. **(B)** Differential expression of the 44 prognostic m6A-RLs in tumor and normal samples. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

investigate differential expression of lncRNAs and TICs. The Pearson correlation analysis was performed to identify m6A-RLs. Kaplan-Meier method and log rank test were used for comparing the OS and DSS between various groups, incorporating the high and low-risk groups and other subgroups based on the expression of each of the 12 m6A-RLs. Student's *t*-test was used to determine the relationships between the risk scores and clinicopathological parameter. Chi-square test or Fisher's exact test was used to analyze associations between the clinicopathological parameters, immune score, clusters, and m6A-RLPS. Univariable and multivariable Cox regression analyses were applied to investigate the independent predictive value of the m6A-RLPS for the OS and DSS of patients with ccRCC. Statistical significance was set at $p < 0.05$ (two-tailed).

RESULTS

Identification of Prognostic m6A-RLs in Clear-Cell Renal Cell Carcinoma

In total, 239 m6A-related lncRNAs were found to be significantly correlated with 23 m6A-related genes using Pearson correlation analysis ($|R| > 0.7$, $p < 0.001$, **Figure 1A**, **Supplementary Table S2**). After excluding patients without cancer or survival data, we merged the survival data with lncRNA expression data of each patient (final patient number = 530). Then, we identified 44 lncRNAs related to prognosis ($p < 0.01$, **Supplementary Table S3**) using univariable Cox regression analysis. As shown in **Figure 1B**, the expression of these prognostic m6A-RLs differed significantly between normal and ccRCC tissues.

Consensus Clustering of m6A-Related lncRNAs Identified in Molecular Subtypes Clear-Cell Renal Cell Carcinoma

We categorized all samples into groups according to the expression profiles of the prognostic m6A-RLs, using a consensus clustering algorithm. We found that k = 2 was the most suitable choice to sort the entire cohort into two groups: cluster 1 ($n = 332$) and cluster 2 ($n = 198$) (**Figures 2A–C**). Kaplan–Meier curves revealed that patients in cluster 2 had a worse OS than those in cluster 1 ($p < 0.001$, **Figure 2D**). Nevertheless, as shown in **Figure 2E**, there were no distinct differences in clinical variables (age, sex, grade, clinical stage, and T stage) between the clusters. Thus, the consensus clustering results were clearly associated with the survival of patients with ccRCC.

GSEA and Immune-Related Analysis of the Two Clusters

To determine the potential pathways and functions associated with the prognostic m6A-RLs in ccRCC, GSEA was applied to the two clusters. Notably, several tumor hallmarks, such as coagulation, glycolysis, and mTORC1 signaling, were predominantly enriched in cluster 1 (nominal $p < 0.05$ and false discovery rate-corrected $q < 0.05$, **Figure 3A**). C2 Kyoto Encyclopedia of Genes and Genomes (KEGG) and C5 Gene Ontology (GO) analyses similarly revealed multiple tumor-related signaling pathways in the samples (**Figures 3B,C**). Notably, several enriched immune-related signaling pathways and genes were identified through both C2 KEGG and C5 GO

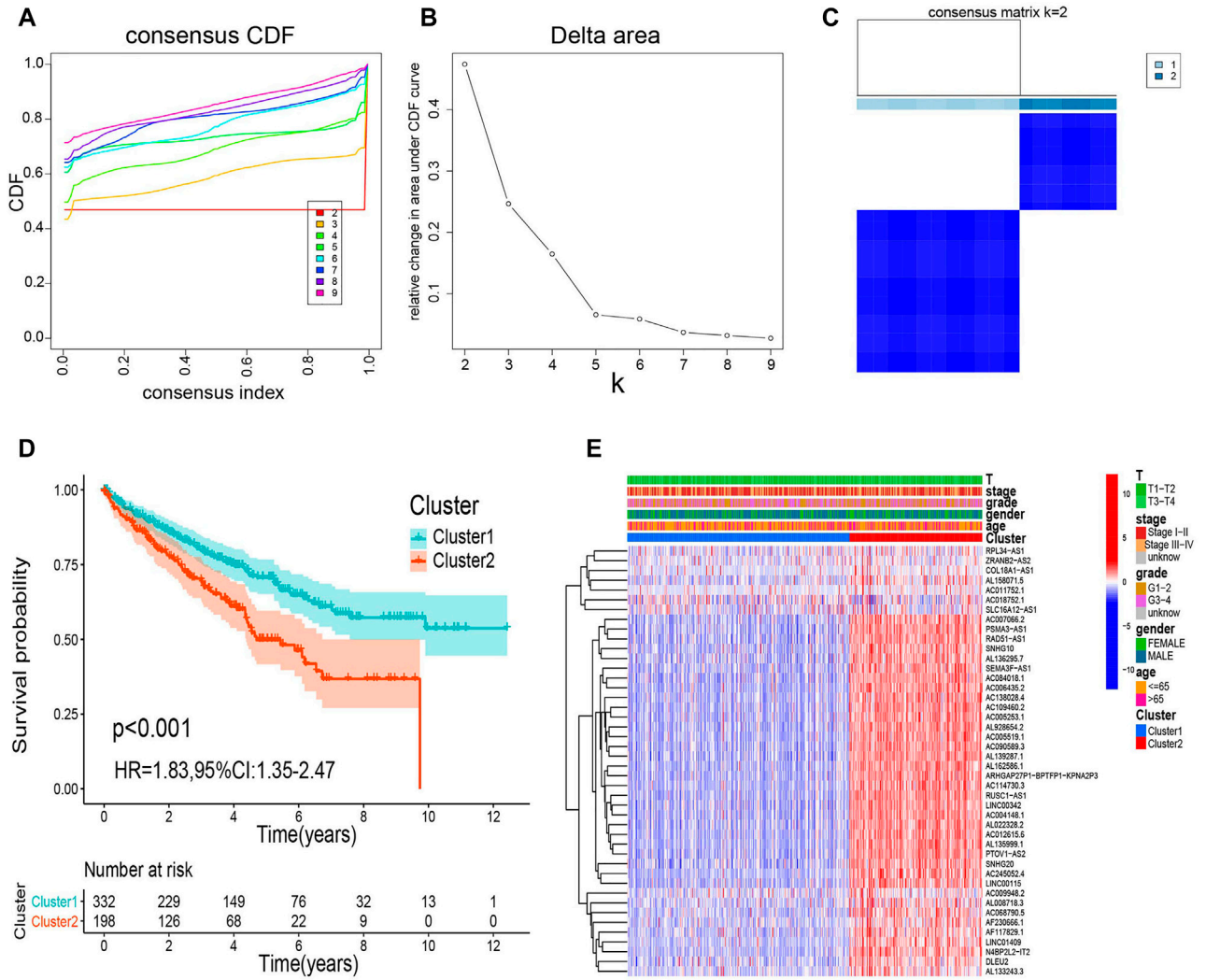

FIGURE 2 | Consensus clustering analysis of m6A-RL expression. **(A)** Consensus clustering cumulative distribution function for k = 2 to 9. **(B)** Relative change in the area under the cumulative distribution function curve for k = 2 to 9. **(C)** Consensus matrix for k = 2. **(D)** Kaplan–Meier survival analysis of the overall survival (OS) for the two clusters. **(E)** Relationships between m6A-RL expression and clinicopathological parameters.

analyses. C7 collection analysis indicated that multiple immune-functional gene sets were enriched in clusters 1 and 2 (**Figure 3D**).

Additionally, we thoroughly evaluated the correlations between the two clusters and TICs in the ccRCC samples. Activated dendritic cells, activated memory CD4 T cells, CD8 T cells, follicular helper T cells, gamma delta T cells, M0 macrophages, naive B cells, neutrophils, and resting natural killer cells were significantly associated with the clusters (**Figure 4A**). Given the potential immunomodulatory effects of the expression of the m6A-RLs, we determined the correlations between the clusters and some immune checkpoints. Notably, all five immune checkpoints evaluated, except TIM-3, were highly expressed in tumor samples and cluster 2 (**Figure 4B**). The correlations between the 44 lncRNAs and the five immune checkpoints are shown in **Supplementary Figure S1**, and most of these correlations were significant. Based on these results, we speculated that the poor prognosis of patients in cluster 2 was probably due to the upregulation of the immune checkpoints.

Construction and Validation of the m6A-Related lncRNA Prognostic Signature

The entire cohort with complete survival information was randomly divided into training (n = 266) and the first validation (n = 264) cohort at a ratio of 1:1 and was then divided into the second (n = 156) and third (n = 374) validation cohorts at a ratio of 3:7. The risk model was developed using the training cohort and validated using the validation cohorts. We conducted LASSO Cox regression analysis to screen the 44 prognostic m6A-RLs and the 12

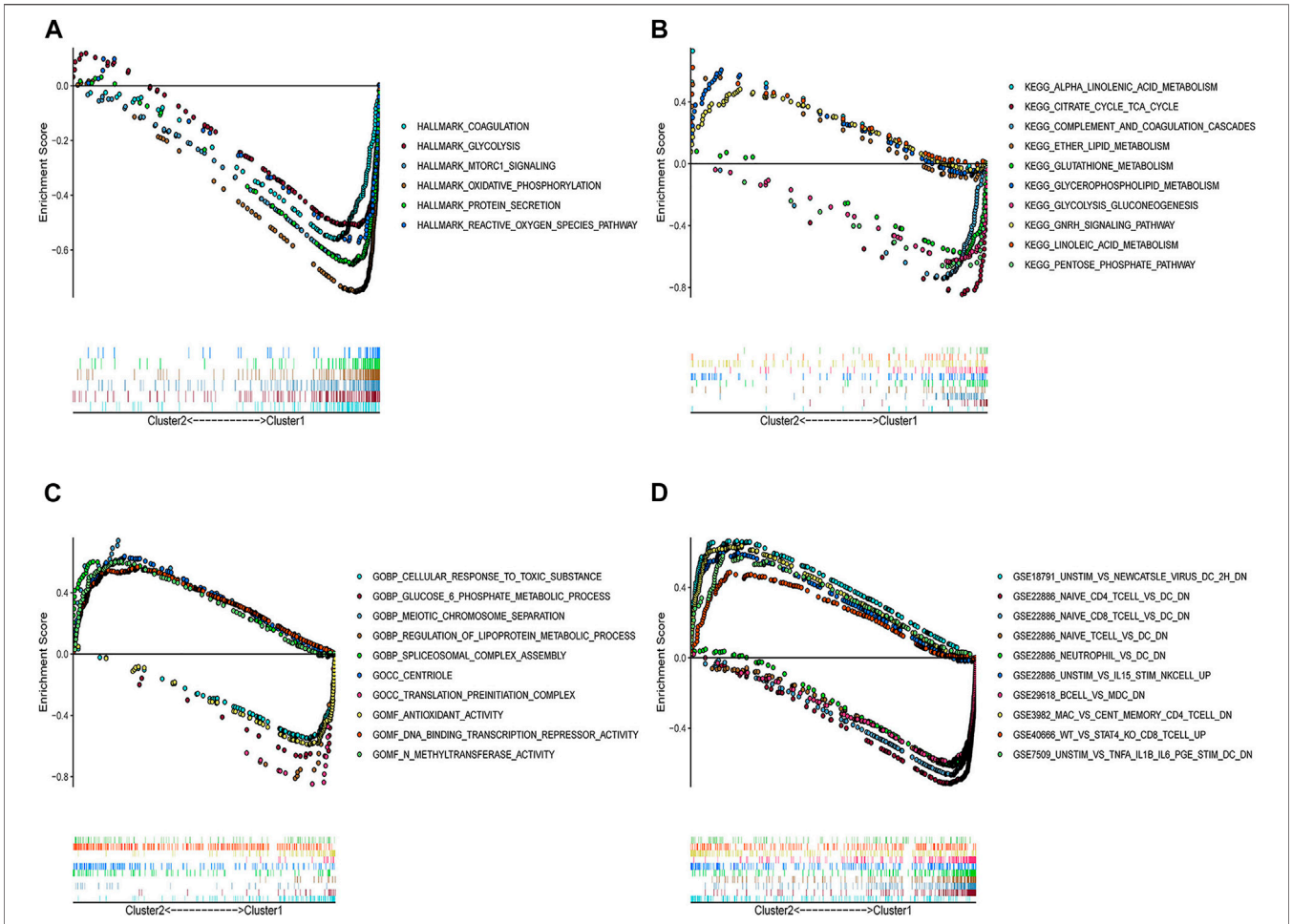

FIGURE 3 | GSEA for the two clusters. **(A)** Tumor hallmarks were enriched in cluster 1. Multi-GSEA enrichment curves for the **(B)** C2 KEGG collection, **(C)** C5 GO collection, and **(D)** C7 collection for two clusters. All the | NES | > 1, Nominal $p < 0.05$, and false discovery rate-corrected $q < 0.25$. NES: Normalized enrichment score.

lncRNAs incorporated into the m6A-RLPS for lncRNAs that would allow prediction of the OS of patients with ccRCC (**Figures 5A–C**). The risk score for each patient was calculated as follows: risk score = (1.083 × AC009948.2 expression) + (−0.303 × AC011752.1 expression) + (−0.148 × AC018752.1 expression) + (−0.002 × AF117829.1 expression) + (0.616 × AL008718.3 expression) + (0.334 × AL133243.3 expression) + (−0.943 × AL158071.5 expression) + (−0.363 × COL18A1-AS1 expression) + (0.141 × DLEU2 expression) + (0.217 × LINC00115 expression) + (−2.645 × RPL34-AS1 expression) + (0.053 × SNHG10 expression) (expression refers to the normalized log levels of each lncRNA). Both univariable Cox regression analysis and Kaplan–Meier log-rank test-based survival analysis supported the remarkable prognostic significance of all 12 lncRNAs, among which AC011752.1, AC018752.1, AL158071.5, COL18A1-AS1, and RPL34-AS1 were protective factors with a hazard ratio < 1, whereas all others were risk factors (**Figure 5D–P**).

The relative expression levels of the 12 signature lncRNAs differed significantly between normal and ccRCC tissues based on TCGA data (**Figures 6A,B**). In addition, we collected the expression data from the GEO DataSets platform for validation analysis of the 12 lncRNAs. Ultimately, only GSE17895 (contained DLEU2), GSE40435 (contained DLEU2, LINC00115, and SNHG10), GSE46699 and GSE53757 (both contained COL18A1-AS1, DLEU2, LINC00115, and SNHG10) were included in the subsequent analyses. The results showed that DLEU2 and LINC00115 were markedly overexpressed in ccRCC samples compared with normal tissues, whereas the expression levels of COL18A1-AS1 and SNHG10 were remarkably lower in tumor tissues than those in the adjacent non-tumor samples (**Figure 6C**–N). Furthermore, Wilcoxon rank-sum test in Lnc2Cancer 3.0 showed the similar differential expression of COL18A1-AS1, DLEU2, LINC00115, RPL34-AS1, and SNHG10 (**Supplementary Figure S2**). These findings were in accordance with the results of TCGA analysis. Moreover, the co-expression of m6A genes (*METTL3* and RNA binding motif protein 15 (*RBM15*)) of the 12 lncRNAs also differed between tumor and normal tissues and was associated with OS in ccRCC (**Supplementary Figure S3**).

Next, the patients were classified into high- and low-risk groups based on the median m6A-RLPS risk score.

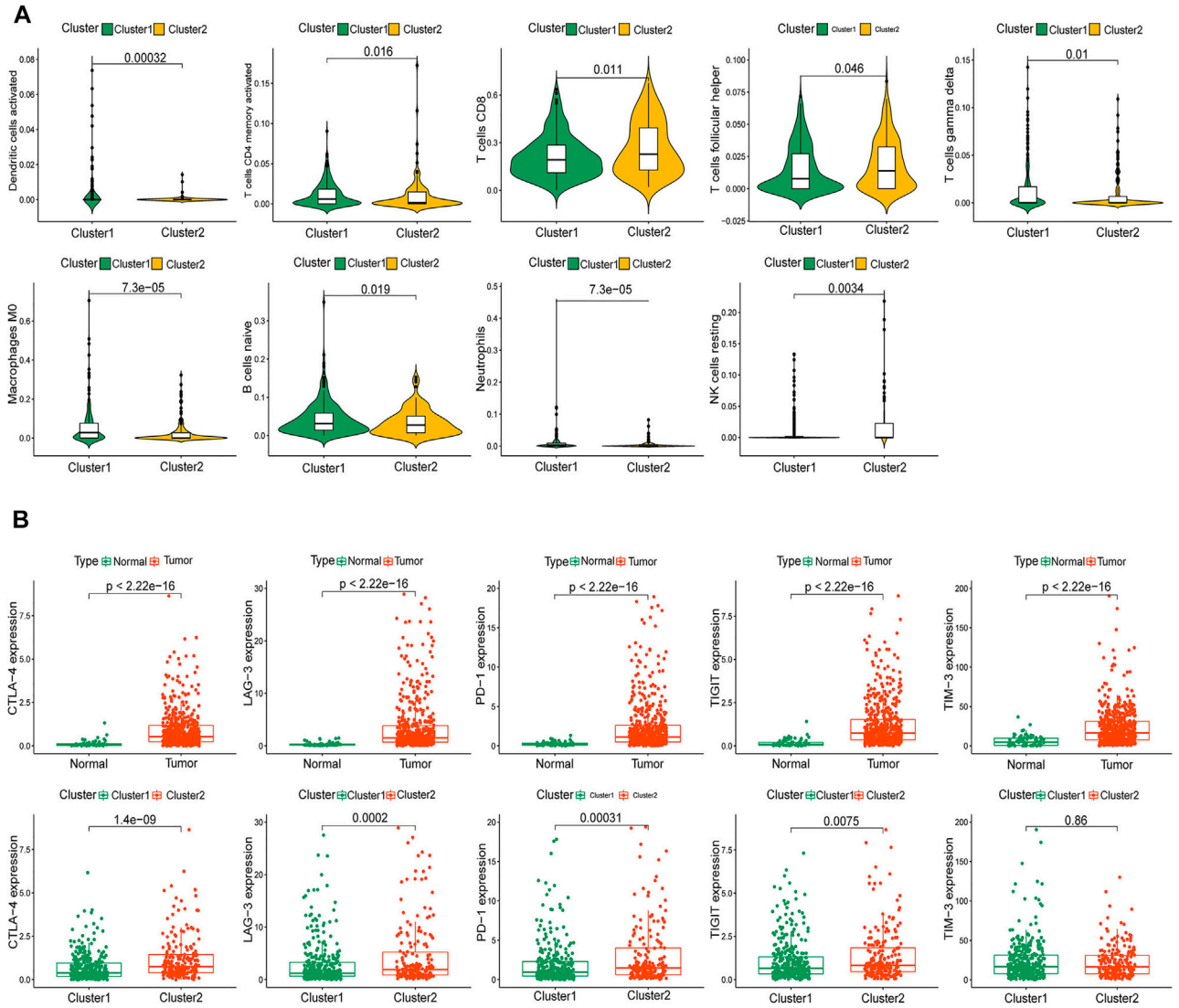

FIGURE 4 | Immune status of patients in the two clusters. **(A)** Identification of nine types of tumor-infiltrating immune cells (TICs) in the two clusters with $p < 0.05$. **(B)** Distribution of five immune checkpoints.

Kaplan–Meier curves showed that in the training cohort, patients in the low-risk group had an improved OS compared to those in the high-risk group ($p < 0.001$, **Figure 7A**). As shown in **Figure 7B**, in the training cohort, the low-risk groups had an obviously higher survival rate and lower values for the risk score. Moreover, as the risk score increased, the expression of the protective lncRNAs (AL158071.5, AC011752.1, AC018752.1, RPL34-AS1, and COL18A1-AS1) decreased, whereas those of the risk lncRNAs (DLEU2, SNHG10, LINC00115, AL008718.3, AC009948.2, AL133243.3, and AF117829.1) increased. The ROC curves showed that the m6A-RLPS was accurate in predicting OS in the training cohort, and the AUCs for the 1-, 3-, and 5-years OS rates were 0.735, 0.760, and 0.779, respectively (**Figure 7C**). Similar findings were obtained in the first (**Figures 7D–F**), second (**Supplementary Figure S4A–C**), and third (**Supplementary Figure S4D–F**) validation cohorts. In particular, the efficacy of m6A-RLPS for predicting DSS in all the cohorts was also satisfactory (**Supplementary Figure S5**). These findings indicated that the prognostic signature has a robust and stable predictive efficacy.

The m6A-Related lncRNA Prognostic Signature Is an Independent Prognostic Indicator in Clear-Cell Renal Cell Carcinoma

We used univariable analysis to evaluate the independent prognostic significance of the m6A-RLPS risk score and several clinical features, including age, sex, tumor grade, and tumor stage (TNM stage was excluded due to incomplete information), in ccRCC. Age, tumor grade, tumor stage, and risk score were closely associated with OS and DSS, and

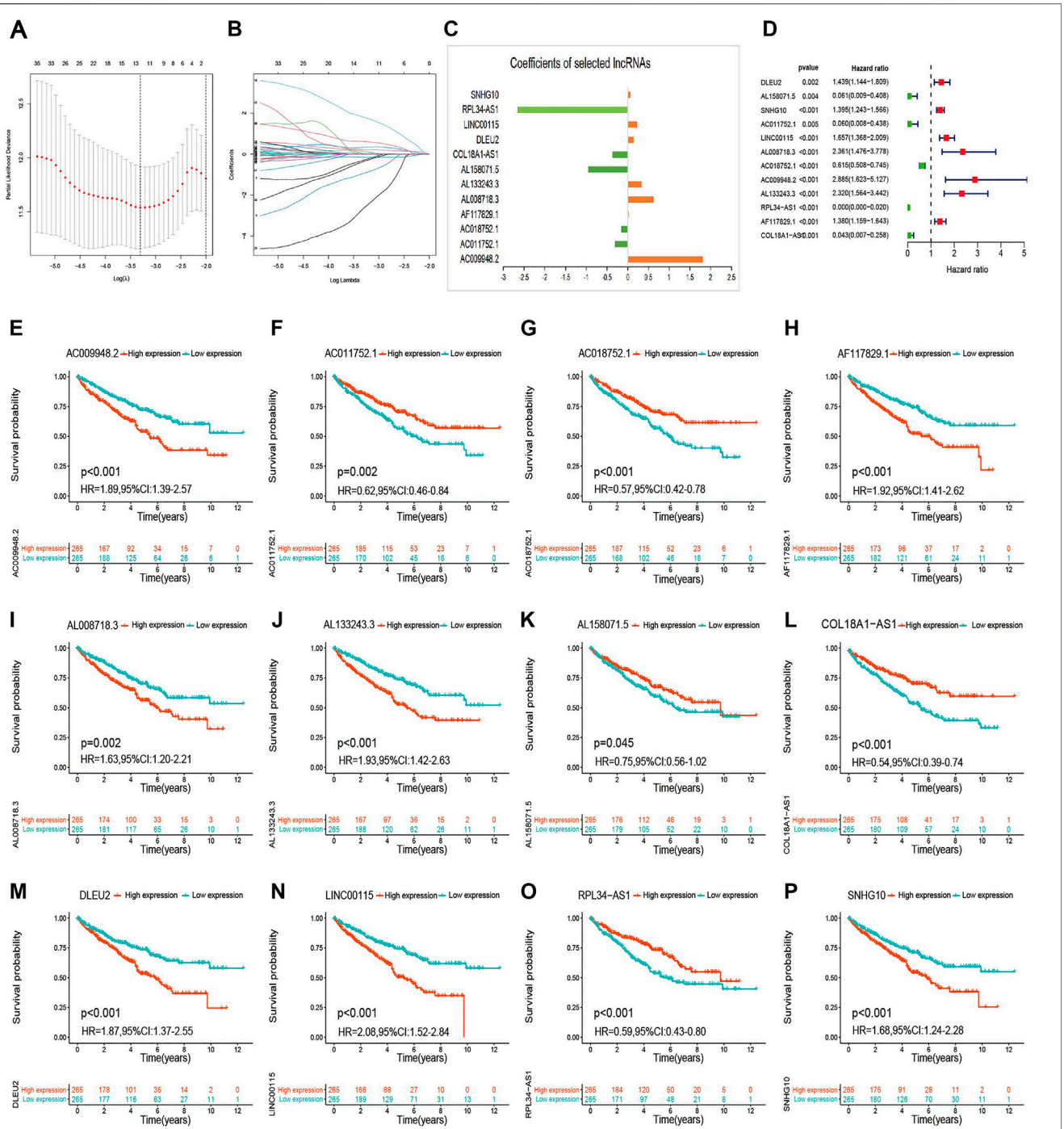

FIGURE 5 | Establishment of the m6A-RLPS. **(A–C)** LASSO Cox regression analysis to determine the 12 m6A-RLs and their corresponding coefficients. **(D)** Forest plot of the 12 m6A-RLs. **(E–P)** Kaplan–Meier curves indicating different overall survival (OS) of patients with different expression levels of the 12 m6A-RLs.

multivariate analysis validated that these four variables were independent prognostic factors in patients with ccRCC (**Figures 8A–D**, **Supplementary Figure S6A–D**). Tests for interaction revealed no significant association between the effect of risk score and clinicopathological factors on the OS and DSS (all p values for interactions > 0.05) (**Supplementary Table S4**). Stratification survival analysis to estimate the predictive ability of m6A-RLPS in ccRCC patients with different clinicopathology revealed that high-risk patients had a worse OS and DSS than low-risk patients in every subgroup (**Figure 8E–N**, **Supplementary Figure S6E-N**), indicating the good predictive performance of the m6A-RLPS.

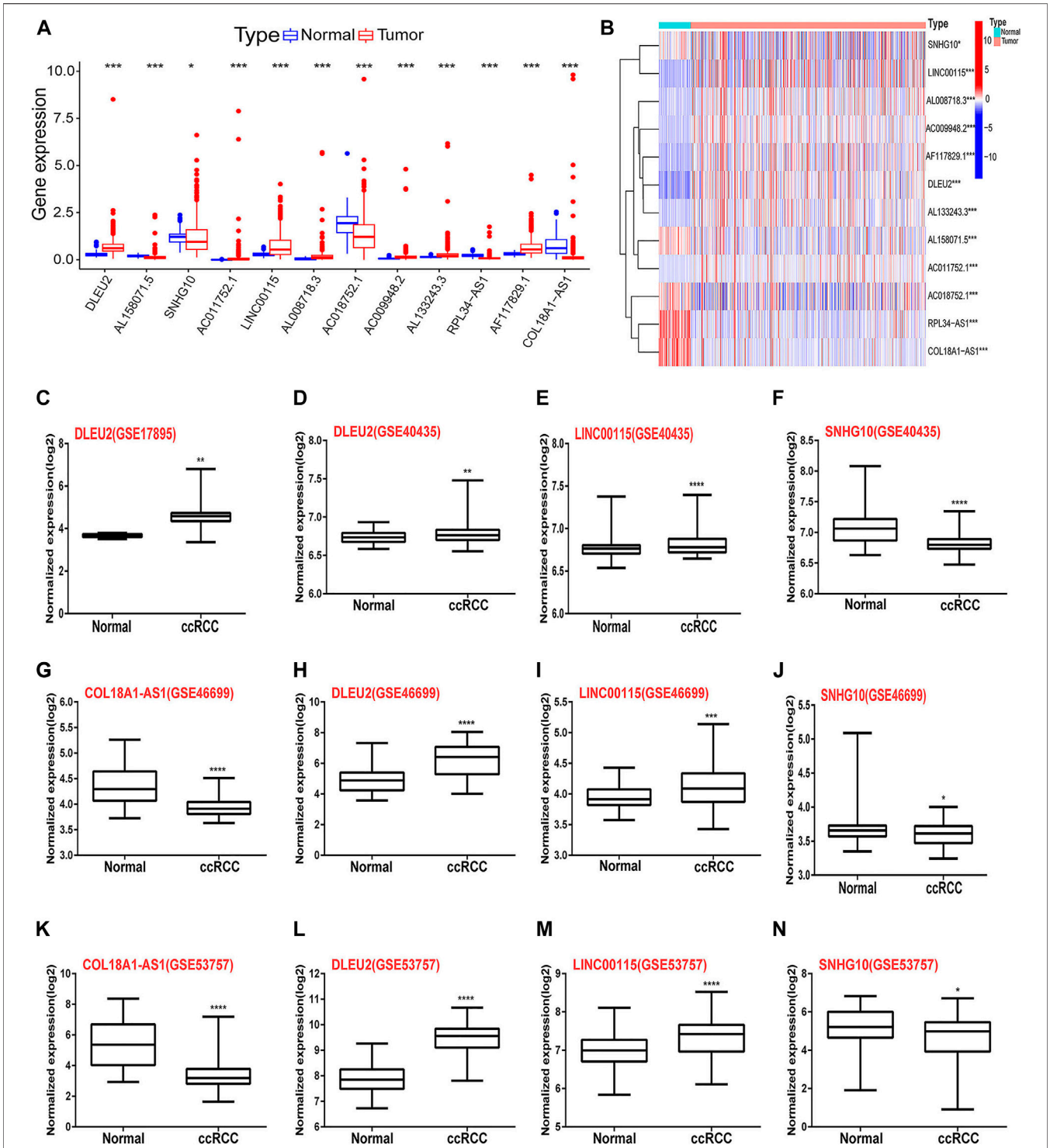

FIGURE 6 | Differential expression of the 12 m6A-RLs between tumor and normal renal tissues. **(A)** Boxplots and **(B)** heatmap of the expression of the 12 m6A-RLs based on TCGA data. **(C–N)** Verification of the differential expression of several lncRNAs among the 12 m6A-RLs based on the GEO data. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$, ****$p < 0.0001$.

A nomogram for the 3- and 5-years OS rates based on the independent predictors determined using the multivariate analysis are shown in **Supplementary Figure S7A**. A point was plotted for each covariate, and a total nomogram score correlated with the 3- and 5-years OS rates was calculated for each patient. The nomogram showed a favorable accuracy in

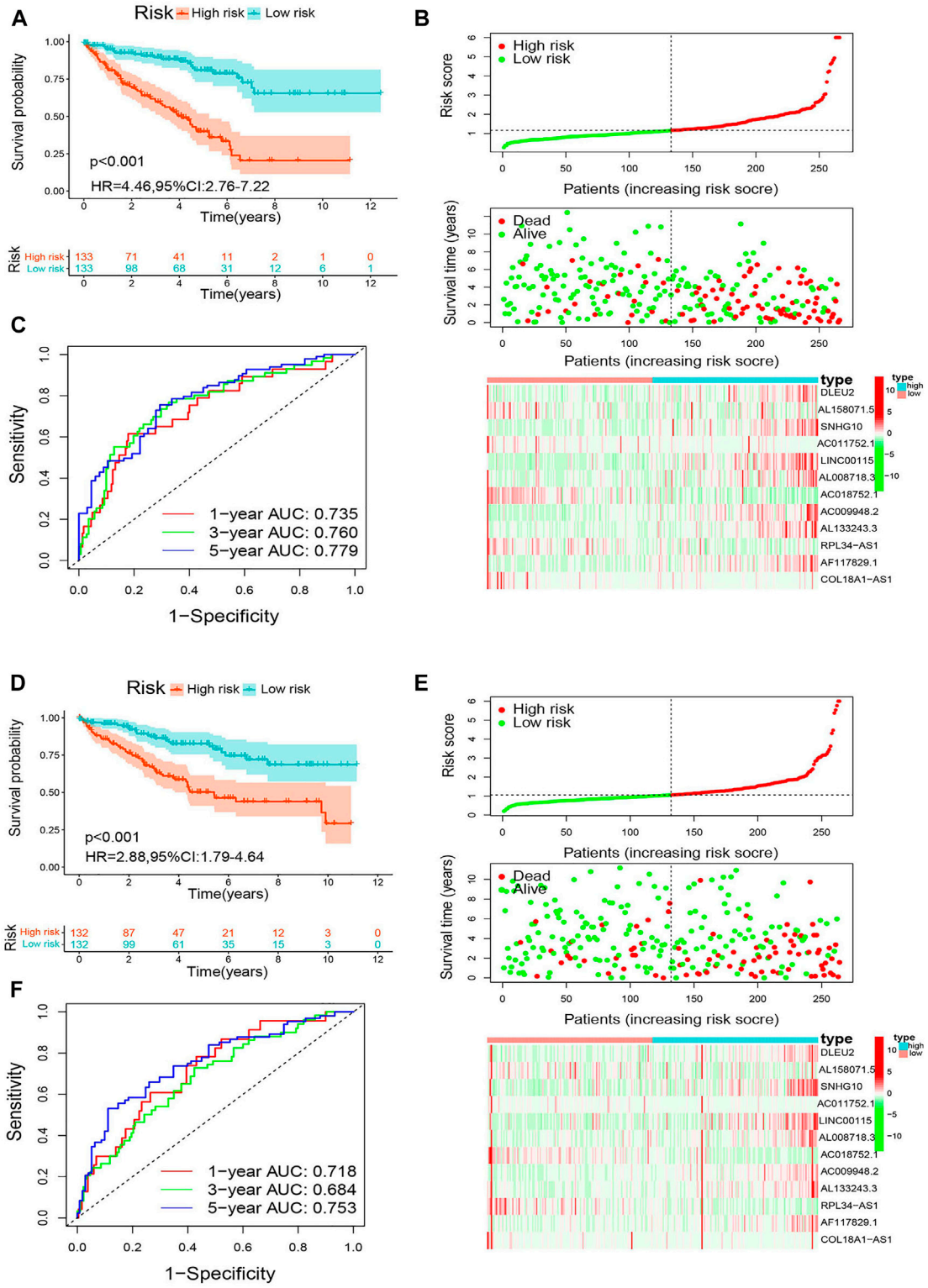

FIGURE 7 | Analysis of the m6A-RLPS efficacy in patients stratified by risk level. **(A)** Kaplan–Meier curves of overall survival (OS) for the m6A-RLPS **(B)** Distributions of risk scores, survival status, and relative lncRNA expression; and **(C)** ROC curves for predicting 1-, 3-, and 5-years OS rates in the training cohort. (D) Kaplan–Meier curves of OS for the m6A-RLPS (E) Distributions of risk scores, survival status, and relative lncRNA expression; and (F) ROC curves for predicting 1-, 3-, and 5-years OS rates in the first validation cohort.

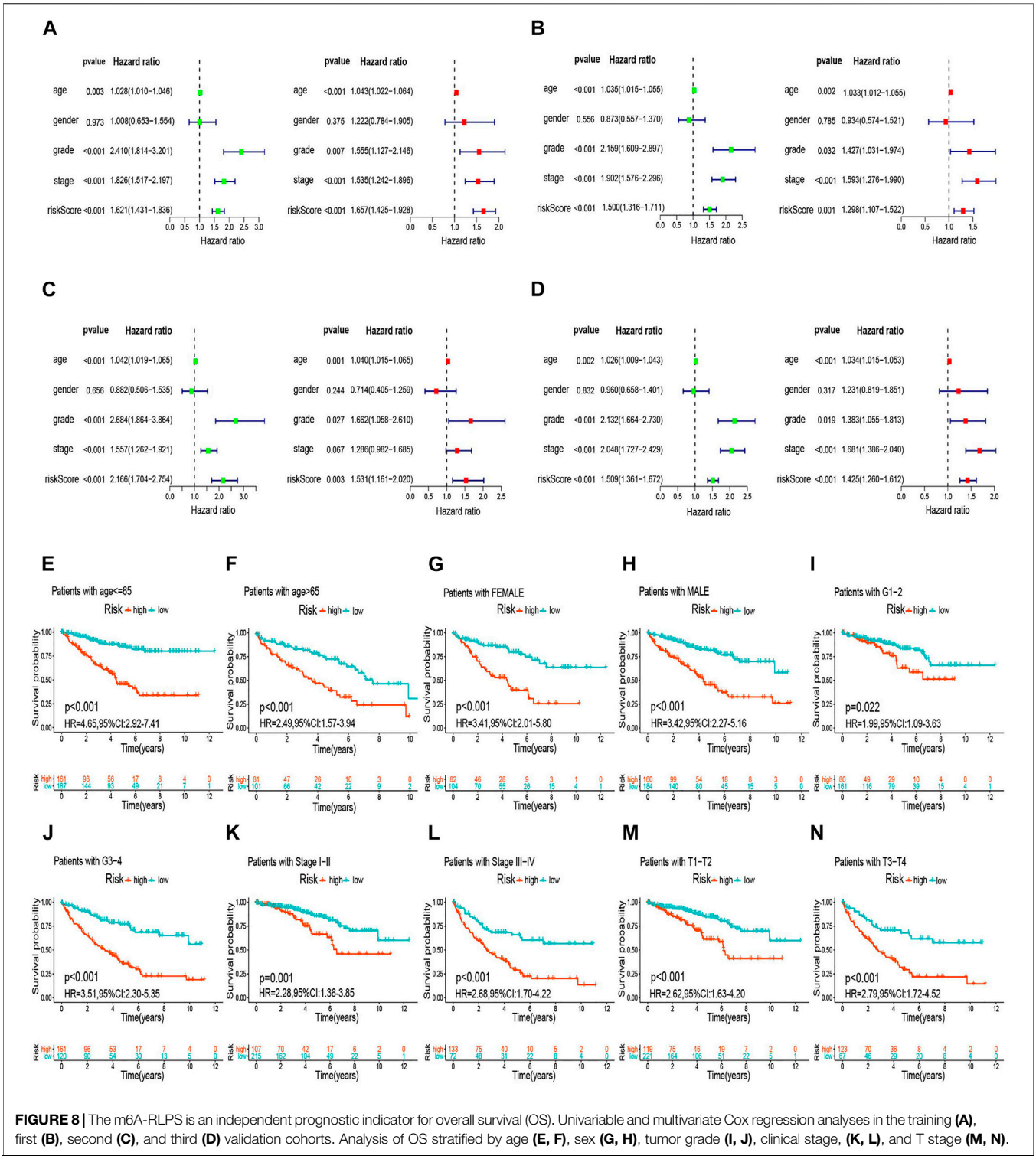

FIGURE 8 | The m6A-RLPS is an independent prognostic indicator for overall survival (OS). Univariable and multivariate Cox regression analyses in the training **(A)**, first **(B)**, second **(C)**, and third **(D)** validation cohorts. Analysis of OS stratified by age **(E, F)**, sex **(G, H)**, tumor grade **(I, J)**, clinical stage, **(K, L)**, and T stage **(M, N)**.

predicting OS, with a C-index of 0.80 (95% CI: 0.76–0.84), 0.76 (95% CI: 0.70–0.82), 0.78 (95% CI: 0.72–0.84), and 0.78 (95% CI: 0.74–0.82) for the training, first, second, and third validation cohorts, respectively. Moreover, calibration curves revealed that there was an appreciable agreement between the predictive outcome and actual survival, and similar findings were obtained in the validation cohorts (**Supplementary Figure S7B–I**).

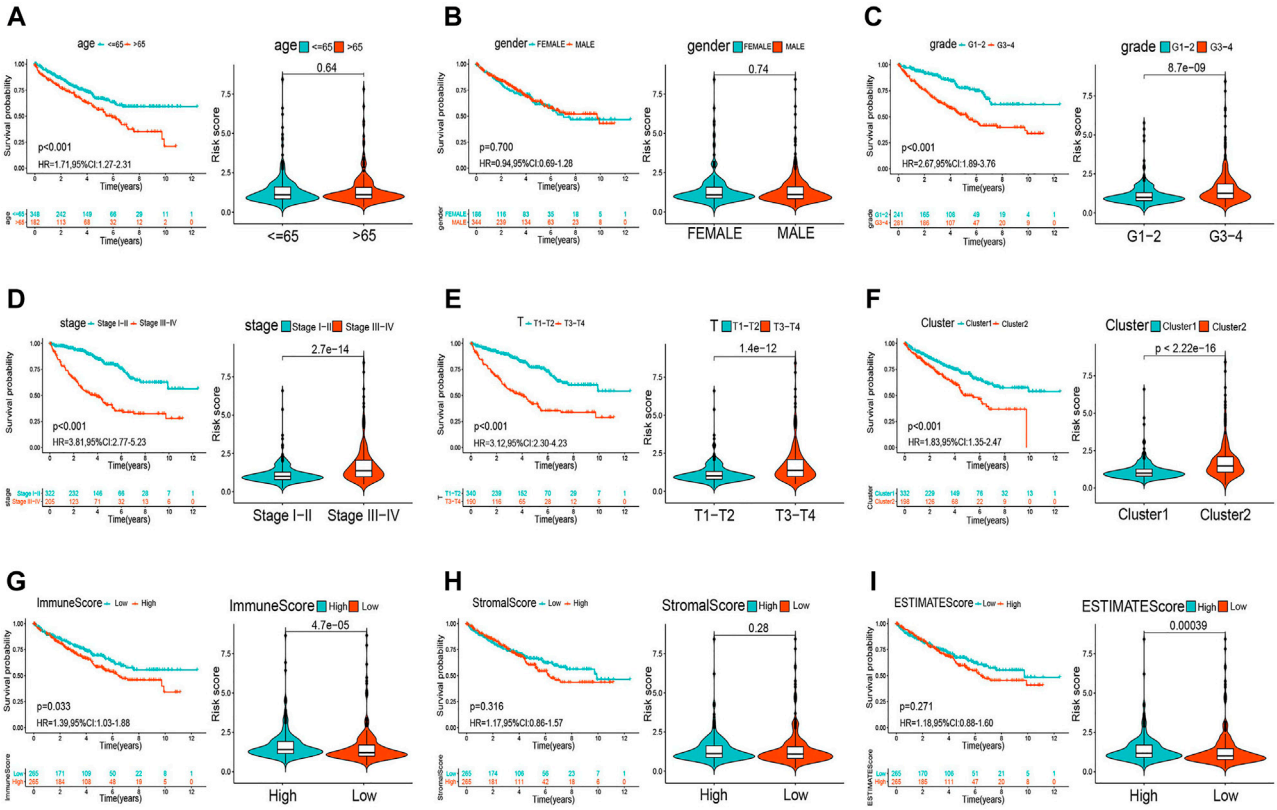

FIGURE 9 | Relationships between the risk score and clinicopathological parameters. Distribution of risk scores according to age **(A)**, sex **(B)**, tumor grade **(C)**, clinical stage **(D)**, T stage **(E)**, cluster **(F)**, immune score **(G)**, stromal score **(H)**, and ESTIMATE score **(I)**.

The m6A-Related lncRNA Prognostic Signature Correlates With Clinicopathology and TME Immune Activity

The heatmap in **Supplementary Figure S8** shows that the expression of the lncRNAs included in the m6A-RLPS was significantly correlated with cluster type, immune score, tumor grade, tumor stage, and T stage (all $p < 0.01$). In addition, a Student *t*-test revealed that the risk score increased with increasing tumor grade, clinical stage, T stage, immune score, and ESTIMATE score (**Figures 9A–E**, G–I), and that high-risk patients tended to be classified into cluster 2 (**Figure 9F**). These findings suggested that the lncRNAs included in the m6A-RLPS can influence the progression, malignancy, and survival outcomes of ccRCC.

Given the strong association between clinicopathology and TME immune activity, we comprehensively investigated the correlation between the m6A-RLPS risk score and immune cell infiltration. By combining difference and correlation analyses, we found that 10 types of TICs, including activated dendritic cells, resting dendritic cells, resting mast cells, M0, M1, and M2 macrophages, monocytes, CD8 T cells, follicular helper T cells, and regulatory T cells (Tregs), were strikingly associated with the m6A-RLPS (**Figure 10**). Of these cell types, four were positively correlated with the m6A-RLPS (M0 macrophages, CD8 T cells, follicular helper T cells, and Tregs), whereas the others were negative correlated.

Finally, we compared drug sensitivity to some TKIs, including axitinib, pazopanib, sorafenib, and sunitinib, between the risk groups. The risk stratification was significantly associated with pazopanib and sunitinib sensitivity (**Figure 11A**). In addition, the m6A-RLPS had a strong positive correlation with the TME scores obtained using the ESTIMATE algorithm (**Figure 11B**). Given the significant correlations of the routine immune checkpoints with the clusters, we next explored the associations between the m6A-RLPS and these immune checkpoints. The expression levels of all immune checkpoints, except TIM-3, were increased in high-risk patients and positively correlated with the risk score, reflecting the effect of the immune checkpoints on the TME and poor oncological outcomes (**Figures 11C,D**).

DISCUSSION

Due to the complexity and heterogeneity of ccRCC, patients with ccRCC generally have a high risk of recurrence and metastasis, as well as a dismal prognosis (Shingarev and Jaimes, 2017). Because of the limited predictive capability of traditional prognostic models, an m6A-related lncRNA-based prognostic model may improve the understanding and management of ccRCC (Veeratterapillay et al., 2012). Therefore, in this study, by systematically screening and identifying target lncRNAs, we

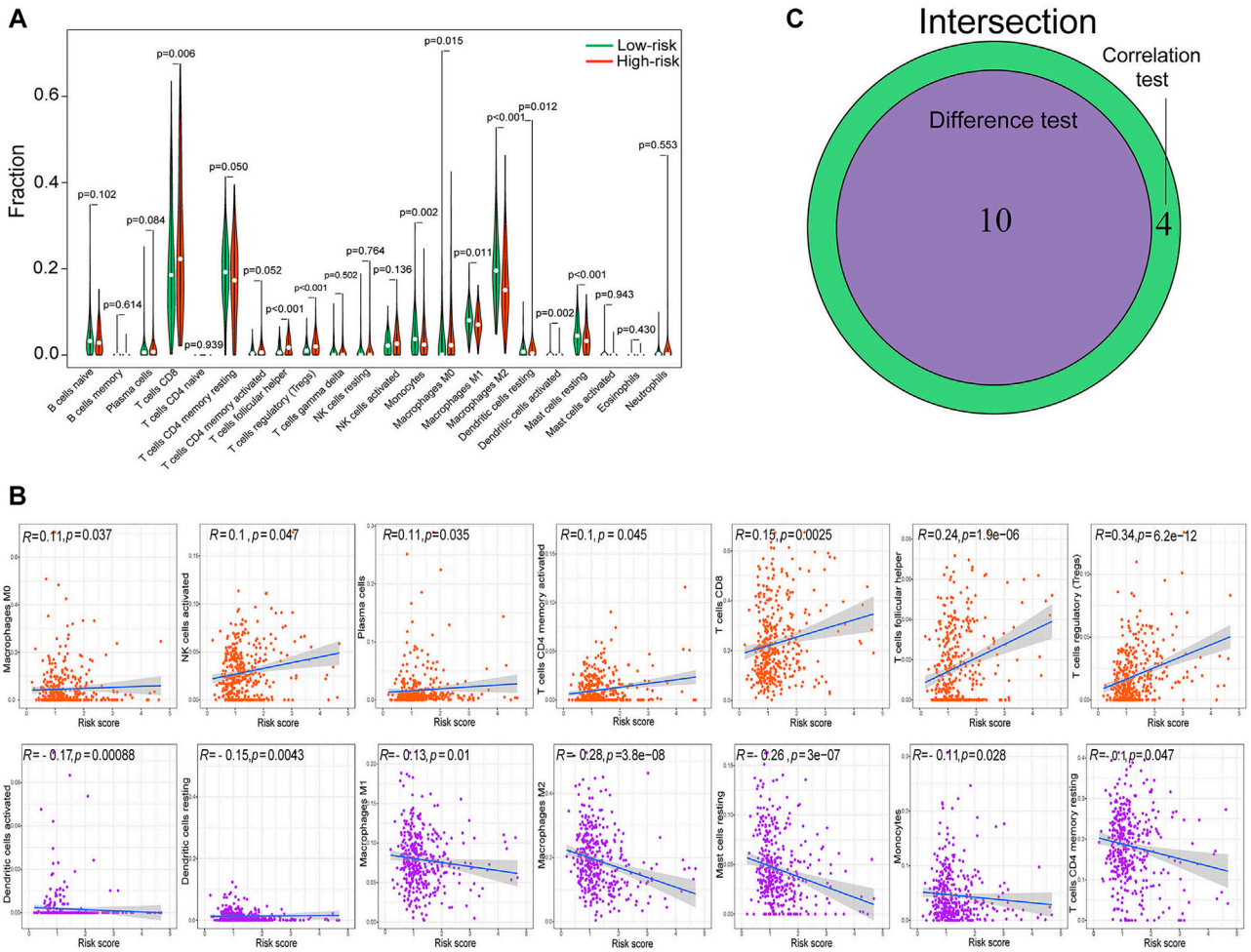

FIGURE 10 | The m6A-RLPS is significantly associated with tumor-infiltrating immune cells (TICs). **(A)** Violin plots of the difference analysis confirming 10 types of TICs, including activated dendritic cells, resting dendritic cells, resting mast cells, M0, M1, and M2 macrophages, monocytes, CD8 T cells, follicular helper T cells, and Tregs, all $p < 0.05$. **(B)** Correlation analysis determined 14 types of TICs. **(C)** Venn diagram of common TICs.

created a reliable prognostic model with a satisfactory predictive capability for ccRCC outcomes.

We conducted a comprehensive ccRCC RNA-seq analysis using data of 539 patients with ccRCC downloaded from TCGA database. In total, 239 m6A-RLs were filtered out, 44 of which were determined to have prognostic value. Next, we divided all ccRCC samples into two clusters by consensus clustering to shed light on the potential effects of the m6A-RLs. The cluster subtypes were strongly linked to the OS of the patients with ccRCC, and cluster 2, comprising patients with worse prognosis, was associated with malignancy-related signaling pathways and numerous immune-related activities as indicated by GSEA. Further analyses emphasized the close relationships among the m6A-RLs, TICs, and immune checkpoints, which indicated that the expression pattern of m6A-RLs is remarkably associated with immunity and oncogenesis. Twelve lncRNAs of the forty-four m6A-RLs were used to construct an m6A-RLPS, which stratified the patients with ccRCC into low- and high-risk groups and exhibited a substantial prognostic prediction performance. The expression of four lncRNAs (COL18A1-AS1, DLEU2, LINC00115, and SNHG10) of the 12 m6A-RLs in ccRCC and normal tissues was evaluated using the GEO data. Multivariate Cox regression analysis showed that the m6A-RLPS was an independent prognostic factor for OS and DSS, and the prognostic predictive capacity of the signature was validated in patients stratified according to clinicopathological parameters. By integrating the m6A-RLPS, age, and tumor grade and stage, we constructed a quantitative nomogram, which was highly accurate and reliable in estimating patient survival. In fact, previous studies have also shown some novel prognostic signatures of ccRCC. Zeng et al. (2019) constructed a prognostic model based on six-lncRNAs, and Chen et al. (2020) identified an m6A RNA methylation regulator-based signature, which were used for predicting the OS of ccRCC and showed a good prediction efficiency, consistent with our model. However, each of the previous studies mainly focused on differentially expressed lncRNAs or m6A regulators; they did not combine and explore the association between m6A modification and lncRNAs, and did not validate the efficacy of signature in predicting DSS of ccRCC.

Numerous studies have highlighted the fact that m6A modification in specific lncRNAs affects tumorigenesis and

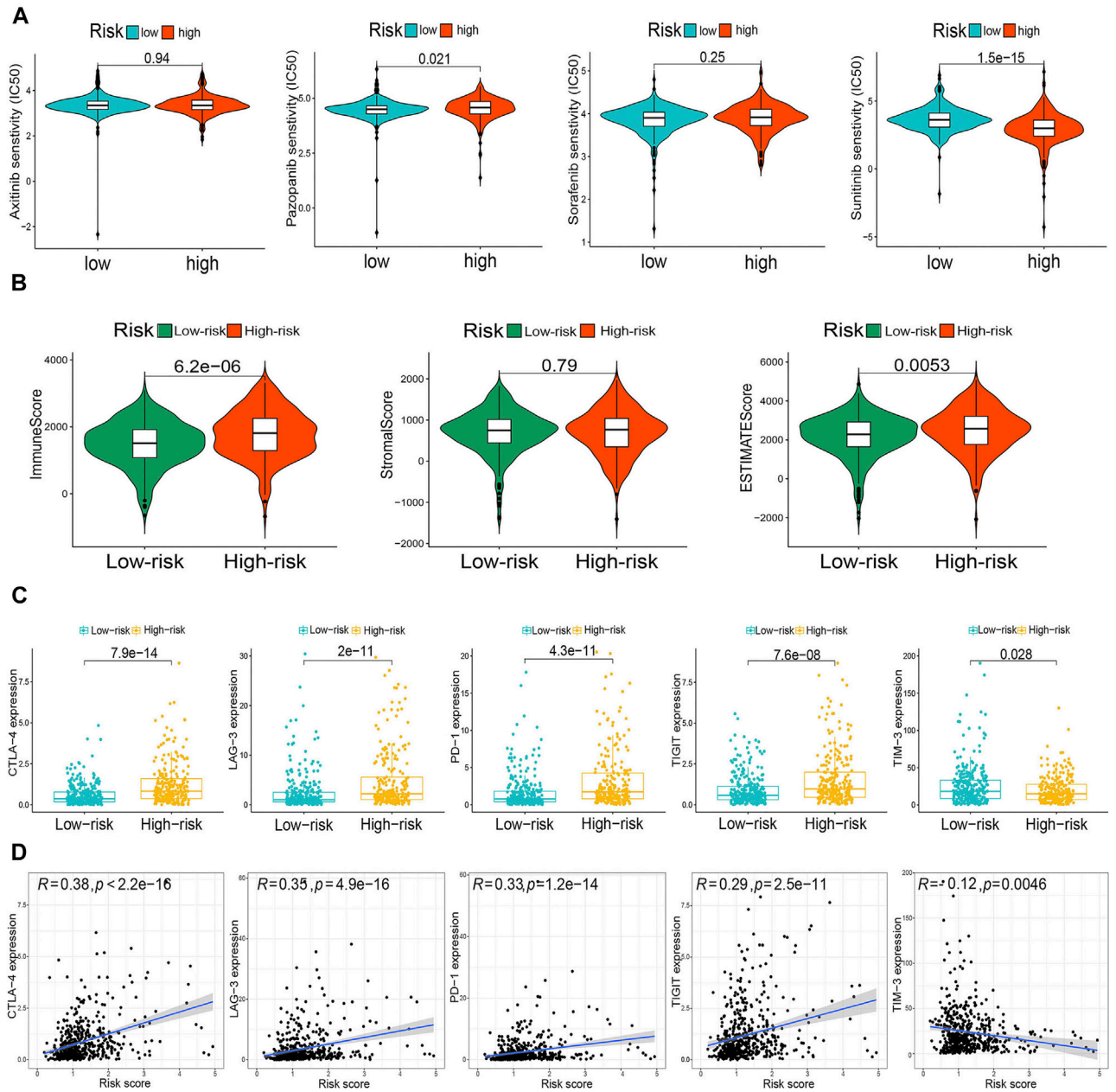

FIGURE 11 | The m6A-RLPS is related with the tumor microenvironment (TME) immune reaction. **(A)** Differential drug sensitivity. **(B)** Distribution of TME scores in the risk groups. **(C)** Differential expression of five immune checkpoints in the risk groups. **(D)** Nearly all selected immune checkpoints correlated positively with the risk score.

metastasis in cancer through various mechanisms; however, how it affects particular lncRNAs in ccRCC remains unclear. In hepatocellular carcinoma, METTL3-mediated m6A modification stabilizes the lncRNA LINC00958 transcript to increase the levels of hepatoma-derived growth factor, thus facilitating tumor growth (Zuo et al., 2020). YTHDF3 inhibits colorectal cancer progression by negatively modulating the lncRNA GAS5 to generate a GAS5-YAP-YTHDF3 negative feedback loop (Ni et al., 2019). Furthermore, lncRNAs can interact with m6A regulators to regulate their functions. For example, the lncRNA GAS5-AS1 enhances the stability of GAS5 by interacting with ALKBH5 to suppress the proliferation of cervical cancer (Wang et al., 2019). We found METTL3 to be a risk factor associated with poor prognosis of ccRCC, which was consistent with findings in one previous study, but in contrast to those in another study (Li et al., 2017; Chen et al., 2020); thus, the role of MELLT3 in ccRCC remains uncertain. Moreover, studies have revealed that METTL3 and RBM15/15B can modulate m6A modification of lncRNA-XIST, which has been shown to facilitate ccRCC tumorigenicity via the miR-302c/SDC1 axis (Patil et al., 2016; Zhang et al., 2017). Together, the findings

suggest that interactions between m6A modification and lncRNAs play vital roles in cancer development and improve our understanding of ccRCC and may aid in identifying potential prognostic markers for this disease.

Among the 12 prognostic m6A modification-related lncRNAs included in the m6A-RLPS, DLEU2 has been associated with the prognosis of multiple types of cancers (Ghafouri-Fard et al., 2021). Chen et al. (2017) reported that DLEU2 downregulates miR-30a-5p, resulting in poor prognosis in patients with ccRCC. LINC00115 can target miR-489-3p through the PI3K/AKT/mTOR pathway to promote the progression of colorectal cancer (Feng et al., 2020), and it can be activated by TGF-β to regulate the tumorigenicity of glioblastoma (Tang et al., 2019). Zhu et al. (2020) reported that lncRNA SNHG10 contributes to the proliferation and invasion of osteosarcoma by activating Wnt/β-catenin signaling via sponging miR-182-5p. Yuan et al. (2020) suggested that SNHG10 accelerates gastric cancer cell proliferation and migration via targeting the miR-495-3p/CTNNB1 axis. Unfortunately, to date, only few studies have been conducted in ccRCC, and the remaining nine lncRNAs have been rarely studied. However, our results corroborate the prognostic value of these m6A-targeted lncRNAs and offer insights into their potential roles in ccRCC oncogenesis and progression.

There is increasing evidence for the significance of m6A modification and lncRNAs in the regulation of cancer immunity, including immune cell infiltration in the TME, and immune resistance and activation (Denaro et al., 2019; Gu et al., 2021). The TME plays an important role in the tumorigenesis and progression of ccRCC (Hinshaw and Shevde, 2019). Recently, ICIs combined with TKIs have been recommended as a first-line treatment for metastatic RCC and the TME is actively involved in the therapeutic efficacy of these treatments (Simonaggio et al., 2021). Our findings indicated that the m6A-RLs play vital roles in ccRCC immune status. Therefore, we thoroughly analyzed the association between the m6A-RLPS and TME immune activity. The results showed that the risk score was strongly related with the TME score. Among TICs, the levels of M0 macrophages, CD8 T cells, follicular helper T cells, and Tregs notably differed between the high- and low-risk groups and were positively correlated with the risk score. Notable, these TICs reportedly are associated with tumorigenesis, progression, and immunotherapy efficacy in ccRCC (Santagata et al., 2017; Kadomoto et al., 2019; Xiong et al., 2020). Recent studies focusing on immune checkpoints, such as CTLA-4, LAG-3, PD-1, TIGHT, and TIM-3, have uncovered that these can significantly regulate the cancer immune function of TICs, leading to the inhibition of immune surveillance (Carosella et al., 2015; Manieri et al., 2017; Andrews et al., 2019). Therefore, immunotherapy has emerged as a promising cancer therapeutic option, and the clinical utility of combinations of ICIs and TKIs in metastatic ccRCC is being extensively researched (Motzer et al., 2019; Powles et al., 2020). In our study, the risk subgroups were notably related to the sensitivity to pazopanib and sunitinib. More importantly, the expression levels of nearly all the above immune checkpoints were positively correlated with the risk score calculated with the prognostic signature, suggesting that an immunosuppressive microenvironment is causally related with poor prognosis, and indicating the potential utility of our m6A-RLPS in estimating the response to ICIs.

We employed and analyzed clinical and survival data from a large cohort of patients with ccRCC from the TCGA database and validated the differential expression of several lncRNAs between normal and tumor tissues based on the GEO data. Nonetheless, this study had some limitations. First, because of a lack of ccRCC samples and large independent clinical data, we were not able to clinically validate the findings. In addition, the detailed functional roles of the m6A-related lncRNAs in ccRCC remain to be identified.

In conclusion, we identified 12 m6A-related lncRNAs with a potential prognostic value and developed a prognostic and predictive m6A-RLPS, which may be useful in the investigation of the functional and molecular mechanisms involved in ccRCC oncogenesis and in the determination of treatment strategies and efficacy in patients with ccRCC.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: The Cancer Genome Atlas (https://portal.gdc.cancer.gov/), Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/), and Lnc2Cancer 3.0 (http://bio-bigdata.hrbmu.edu.cn/lnc2cancer).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Study concept and design: TM and YZ; data acquisition and data analysis/interpretation: TM and XW; manuscript drafting or manuscript revision for important intellectual content: TM, XL,and JW; literature study: SL, WZ, LM, and ZT. All authors contributed to the article and approved the submitted version

FUNDING

This study was supported by the National Key Research and Development Program of China (grant number: 2018YFC2002202).

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.cn) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.726369/full#supplementary-material

REFERENCES

Andrews, L. P., Yano, H., and Vignali, D. A. A. (2019). Inhibitory Receptors and Ligands beyond PD-1, PD-L1 and CTLA-4: Breakthroughs or Backups. *Nat. Immunol.* 20, 1425–1434. doi:10.1038/s41590-019-0512-0

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68, 394–424. doi:10.3322/caac.21492

Cai, J., Yang, F., Zhan, H., Situ, J., Li, W., Mao, Y., et al. (2019). RNA m6A Methyltransferase METTL3 Promotes the Growth of Prostate Cancer by Regulating Hedgehog Pathway. *Ott* 12, 9143–9152. doi:10.2147/OTT.S226796

Carosella, E. D., Ploussard, G., LeMaoult, J., and Desgrandchamps, F. (2015). A Systematic Review of Immunotherapy in Urologic Cancer: Evolving Roles for Targeting of CTLA-4, PD-1/pd-L1, and HLA-G. *Eur. Urol.* 68, 267–279. doi:10.1016/j.eururo.2015.02.032

Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M., and Alizadeh, A. A. (2018). Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol. Biol.* 1711, 243–259. doi:10.1007/978-1-4939-7493-1_12

Chen, J., Yu, K., Zhong, G., and Shen, W. (2020). Identification of a m6A RNA Methylation Regulators-Based Signature for Predicting the Prognosis of clear Cell Renal Carcinoma. *Cancer Cel Int* 20, 157. doi:10.1186/s12935-020-01238-3

Chen, X.-Y., Zhang, J., and Zhu, J.-S. (2019). The Role of m6A RNA Methylation in Human Cancer. *Mol. Cancer* 18, 103. doi:10.1186/s12943-019-1033-z

Chen, Z., Zhang, J., Zhang, Z., Feng, Z., Wei, J., Lu, J., et al. (2017). The Putative Tumor Suppressor microRNA-30a-5p Modulates clear Cell Renal Cell Carcinoma Aggressiveness through Repression of ZEB2. *Cell Death Dis* 8, e2859. doi:10.1038/cddis.2017.252

Dalgliesh, G. L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., et al. (2010). Systematic Sequencing of Renal Carcinoma Reveals Inactivation of Histone Modifying Genes. *Nature* 463, 360–363. doi:10.1038/nature08672

Denaro, N., Merlano, M. C., and Lo Nigro, C. (2019). Long Noncoding RNA S as Regulators of Cancer Immunity. *Mol. Oncol.* 13, 61–73. doi:10.1002/1878-0261.12413

Deng, X., Su, R., Weng, H., Huang, H., Li, Z., and Chen, J. (2018). RNA N6-Methyladenosine Modification in Cancers: Current Status and Perspectives. *Cell Res* 28, 507–517. doi:10.1038/s41422-018-0034-6

Eckel-Passow, J. E., Serie, D. J., Bot, B. M., Joseph, R. W., Cheville, J. C., and Parker, A. S. (2014). ANKS1B Is a Smoking-Related Molecular Alteration in clear Cell Renal Cell Carcinoma. *BMC Urol.* 14, 14. doi:10.1186/1471-2490-14-14

Feng, W., Li, B., Wang, J., Zhang, H., Liu, Y., Xu, D., et al. (2020). Long Non-coding RNA LINC00115 Contributes to the Progression of Colorectal Cancer by Targeting miR-489-3p via the PI3K/AKT/mTOR Pathway. *Front. Genet.* 11, 567630. doi:10.3389/fgene.2020.567630

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Soft.* 33, 1–22. doi:10.18637/jss.v033.i01

Gao, Y., Shang, S., Guo, S., Li, X., Zhou, H., Liu, H., et al. (2021). Lnc2Cancer 3.0: an Updated Resource for Experimentally Supported lncRNA/circRNA Cancer Associations and Web Tools Based on RNA-Seq and scRNA-Seq Data. *Nucleic Acids Res.* 49 (D1), D1251–D1258. doi:10.1093/nar/gkaa1006

Ghafouri-Fard, S., Dashti, S., Farsi, M., and Taheri, M. (2021). Deleted in Lymphocytic Leukemia 2 (DLEU2): an lncRNA with Dissimilar Roles in Different Cancers. *Biomed. Pharmacother.* 133, 111093. doi:10.1016/j.biopha.2020.111093

Gu, Y., Wu, X., Zhang, J., Fang, Y., Pan, Y., Shu, Y., et al. (2021). The Evolving Landscape of N6-Methyladenosine Modification in the Tumor Microenvironment. *Mol. Ther.* 29, 1703–1715. doi:10.1016/j.ymthe.2021.04.009

Hinshaw, D. C., and Shevde, L. A. (2019). The Tumor Microenvironment Innately Modulates Cancer Progression. *Cancer Res.* 79, 4557–4566. doi:10.1158/0008-5472.CAN-18-3962

Huang, H., Weng, H., and Chen, J. (2020). m6A Modification in Coding and Non-coding RNAs: Roles and Therapeutic Implications in Cancer. *Cancer Cell* 37, 270–288. doi:10.1016/j.ccell.2020.02.004

Kadomoto, S., Izumi, K., Hiratsuka, K., Nakano, T., Naito, R., Makino, T., et al. (2019). Tumor-associated Macrophages Induce Migration of Renal Cell Carcinoma Cells via Activation of the CCL20-CCR6 axis. *Cancers* 12, 89. doi:10.3390/cancers12010089

Li, W. (2012). Volcano Plots in Analyzing Differential Expressions with mRNA Microarrays. *J. Bioinform. Comput. Biol.* 10 (6), 1231003. doi:10.1142/s0219720012310038

Li, X., Tang, J., Huang, W., Wang, F., Li, P., Qin, C., et al. (2017). The M6A Methyltransferase METTL3: Acting as a Tumor Suppressor in Renal Cell Carcinoma. *Oncotarget* 8, 96103–96116. doi:10.18632/oncotarget.21726

Liu, J., Eckert, M. A., Harada, B. T., Liu, S.-M., Lu, Z., Yu, K., et al. (2018). m6A mRNA Methylation Regulates AKT Activity to Promote the Proliferation and Tumorigenicity of Endometrial Cancer. *Nat. Cel Biol.* 20, 1074–1083. doi:10.1038/s41556-018-0174-4

Ljungberg, B., Albiges, L., Abu-Ghanem, Y., Bensalah, K., Dabestani, S., Fernández-Pello, S., et al. (2019). European Association of Urology Guidelines on Renal Cell Carcinoma: the 2019 Update. *Eur. Urol.* 75, 799–810. doi:10.1016/j.eururo.2019.02.011

Manieri, N. A., Chiang, E. Y., and Grogan, J. L. (2017). TIGIT: a Key Inhibitor of the Cancer Immunity Cycle. *Trends Immunol.* 38, 20–28. doi:10.1016/j.it.2016.10.002

Mori, K., Mostafaei, H., Miura, N., Karakiewicz, P. I., Luzzago, S., Schmidinger, M., et al. (2021). Systemic Therapy for Metastatic Renal Cell Carcinoma in the First-Line Setting: a Systematic Review and Network Meta-Analysis. *Cancer Immunol. Immunother.* 70, 265–273. doi:10.1007/s00262-020-02684-8

Motzer, R. J., Penkov, K., Haanen, J., Rini, B., Albiges, L., Campbell, M. T., et al. (2019). Avelumab Plus Axitinib versus Sunitinib for Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* 380, 1103–1115. doi:10.1056/NEJMoa1816047

Ni, W., Yao, S., Zhou, Y., Liu, Y., Huang, P., Zhou, A., et al. (2019). Long Noncoding RNA GAS5 Inhibits Progression of Colorectal Cancer by Interacting with and Triggering YAP Phosphorylation and Degradation and Is Negatively Regulated by the m6A Reader YTHDF3. *Mol. Cancer* 18, 143. doi:10.1186/s12943-019-1079-y

Patil, D. P., Chen, C.-K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m6A RNA Methylation Promotes XIST-Mediated Transcriptional Repression. *Nature* 537, 369–373. doi:10.1038/nature19342

Peng, W.-X., Koirala, P., and Mo, Y.-Y. (2017). LncRNA-mediated Regulation of Cell Signaling in Cancer. *Oncogene* 36, 5661–5667. doi:10.1038/onc.2017.184

Powles, T., Plimack, E. R., Soulières, D., Waddell, T., Stus, V., Gafanov, R., et al. (2020). Pembrolizumab Plus Axitinib versus Sunitinib Monotherapy as First-Line Treatment of Advanced Renal Cell Carcinoma (KEYNOTE-426): Extended Follow-Up from a Randomised, Open-Label, Phase 3 Trial. *Lancet Oncol.* 21 (e553), 1563–1573. doi:10.1016/S1470-2045(20)30436-8

Rini, B. I., Campbell, S. C., and Escudier, B. (2009). Renal Cell Carcinoma. *The Lancet* 373, 1119–1132. doi:10.1016/S0140-6736(09)60229-4

Santagata, S., Napolitano, M., D'Alterio, C., Desicato, S., Maro, S. D., Marinelli, L., et al. (2017). Targeting CXCR4 Reverts the Suppressive Activity of T-Regulatory Cells in Renal Cancer. *Oncotarget* 8, 77110–77120. doi:10.18632/oncotarget.20363

Shingarev, R., and Jaimes, E. A. (2017). Renal Cell Carcinoma: New Insights and Challenges for a Clinician Scientist. *Am. J. Physiology-Renal Physiol.* 313, F145–F154. doi:10.1152/ajprenal.00480.2016

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70, 7–30. doi:10.3322/caac.21590

Simonaggio, A., Epaillard, N., Pobel, C., Moreira, M., Oudard, S., and Vano, Y.-A. (2021). Tumor Microenvironment Features as Predictive Biomarkers of Response to Immune Checkpoint Inhibitors (ICI) in Metastatic clear Cell Renal Cell Carcinoma (mccRCC). *Cancers* 13, 231. doi:10.3390/cancers13020231

Strick, A., Hagen, F., Gundert, L., Klümper, N., Tolkach, Y., Schmidt, D., et al. (2020). The N 6 -methyladenosine (M 6 A) Erasers Alkylation Repair Homologue 5 (ALKBH5) and Fat Mass and Obesity-associated Protein (FTO) Are Prognostic Biomarkers in Patients with clear Cell Renal Carcinoma. *BJU Int.* 125, 617–624. doi:10.1111/bju.15019

Tang, J., Yu, B., Li, Y., Zhang, W., Alvarez, A. A., Hu, B., et al. (2019). TGF-β-activated lncRNA LINC00115 Is a Critical Regulator of Glioma Stem-like Cell Tumorigenicity. *EMBO Rep.* 20, e48170. doi:10.15252/embr.201948170

Taniue, K., and Akimitsu, N. (2021). The Functions and Unique Features of lncRNAs in Cancer Development and Tumorigenesis. *Ijms* 22, 632. doi:10.3390/ijms22020632

Tian, P., Wei, J. X., Li, J., Ren, J. K., and Yang, J. J. (2021). LncRNA SNHG1 Regulates Immune Escape of Renal Cell Carcinoma by Targeting miR-129-3p to Activate STAT3 and PD-L1. *Cell Biol. Int.* 45, 1546–1560. doi:10.1002/cbin.11595

Veeratterapillay, R., Simren, R., El-Sherif, A., Johnson, M. I., Soomro, N., and Heer, R. (2012). Accuracy of the Revised 2010 TNM Classification in Predicting the Prognosis of Patients Treated for Renal Cell Cancer in the north east of England. *J. Clin. Pathol.* 65, 367–371. doi:10.1136/jclinpath-2011-200468

von Roemeling, C. A., Radisky, D. C., Marlow, L. A., Cooper, S. J., Grebe, S. K., Anastasiadis, P. Z., et al. (2014). Neuronal Pentraxin 2 Supports clear Cell Renal Cell Carcinoma by Activating the AMPA-Selective Glutamate Receptor-4. *Cancer Res.* 74, 4796–4810. doi:10.1158/0008-5472.CAN-14-0210

Wang, Q., Chen, C., Ding, Q., Zhao, Y., Wang, Z., Chen, J., et al. (2020a). METTL3-mediated m6A Modification of HDGF mRNA Promotes Gastric Cancer Progression and Has Prognostic Significance. *Gut* 69, 1193–1205. doi:10.1136/gutjnl-2019-319639

Wang, T., Kong, S., Tao, M., and Ju, S. (2020b). The Potential Role of RNA N6-Methyladenosine in Cancer Progression. *Mol. Cancer* 19, 88. doi:10.1186/s12943-020-01204-7

Wang, X., Zhang, J., and Wang, Y. (2019). Long Noncoding RNA GAS5-AS1 Suppresses Growth and Metastasis of Cervical Cancer by Increasing GAS5 Stability. *Am. J. Transl. Res.* 11, 4909–4921.

Wang, X., Xie, H., Ying, Y., Chen, D., and Li, J. (2020c). Roles of N 6 -methyladenosine (M 6 A) RNA Modifications in Urological Cancers. *J. Cel. Mol. Med.* 24, 10302–10310. doi:10.1111/jcmm.15750

Wanna-Udom, S., Terashima, M., Lyu, H., Ishimura, A., Takino, T., Sakari, M., et al. (2020). The m6A Methyltransferase METTL3 Contributes to Transforming Growth Factor-Beta-Induced Epithelial-Mesenchymal Transition of Lung Cancer Cells through the Regulation of JUNB. *Biochem. Biophysical Res. Commun.* 524, 150–155. doi:10.1016/j.bbrc.2020.01.042

Wozniak, M. B., Le Calvez-Kelm, F., Abedi-Ardekani, B., Byrnes, G., Durand, G., Carreira, C., et al. (2013). Integrative Genome-wide Gene Expression Profiling of clear Cell Renal Cell Carcinoma in Czech Republic and in the United States. *PLOS ONE* 8, e57886. doi:10.1371/journal.pone.0057886

Xiong, Y., Wang, Z., Zhou, Q., Zeng, H., Zhang, H., Liu, Z., et al. (2020). Identification and Validation of Dichotomous Immune Subtypes Based on Intratumoral Immune Cells Infiltration in clear Cell Renal Cell Carcinoma Patients. *J. Immunother. Cancer* 8, e000447. doi:10.1136/jitc-2019-000447

Yang, X., Hu, X., Liu, J., Wang, R., Zhang, C., Han, F., et al. (2020). N6-methyladenine Modification in Noncoding RNAs and its Function in Cancer. *Biomark. Res.* 8, 61. doi:10.1186/s40364-020-00244-x

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Yuan, X., Yang, T., Xu, Y., Ou, S., Shi, P., Cao, M., et al. (2020). SNHG10 Promotes Cell Proliferation and Migration in Gastric Cancer by Targeting miR-495-3p/CTNNB1 Axis. *Dig. Dis. Sci.* 66 (8), 2627–2636. doi:10.1007/s10620-020-06576-w

Zeng, J.-h., Lu, W., Liang, L., Chen, G., Lan, H.-h., Liang, X.-Y., et al. (2019). Prognosis of clear Cell Renal Cell Carcinoma (ccRCC) Based on a Six-lncRNA-Based Risk Score: an Investigation Based on RNA-Sequencing Data. *J. Transl. Med.* 17, 281. doi:10.1186/s12967-019-2032-y

Zhang, J., Cao, Z., Ding, X., Wei, X., Zhang, X., Hou, J., et al. (2017). The lncRNA XIST Regulates the Tumorigenicity of Renal Cell Carcinoma Cells via the miR-302c/SDC1 axis. *Int. J. Clin. Exp. Pathol.* 10, 7481–7491.

Zhang, J., Jin, S., Xiao, W., Zhu, X., Jia, C., and Lin, Z. (2021). Long Noncoding RNA LINC00641 Promotes Renal Cell Carcinoma Progression via Sponging microRNA-340-5p. *Cancer Cel Int* 21, 210. doi:10.1186/s12935-021-01895-y

Zhu, S., Liu, Y., Wang, X., Wang, J., and Xi, G. (2020). lncRNA SNHG10 Promotes the Proliferation and Invasion of Osteosarcoma via Wnt/β-Catenin Signaling. *Mol. Ther. - Nucleic Acids* 22, 957–970. doi:10.1016/j.omtn.2020.10.010

Zuo, X., Chen, Z., Gao, W., Zhang, Y., Wang, J., Wang, J., et al. (2020). M6A-mediated Upregulation of LINC00958 Increases Lipogenesis and Acts as a Nanotherapeutic Target in Hepatocellular Carcinoma. *J. Hematol. Oncol.* 13, 5. doi:10.1186/s13045-019-0839-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

MINI REVIEW
published: 18 October 2021
doi: 10.3389/fgene.2021.765580

Research Progress of PPR Proteins in RNA Editing, Stress Response, Plant Growth and Development

Tengfei Qin[1†], Pei Zhao[2†], Jialiang Sun[1†], Yuping Zhao[3], Yaxin Zhang[1], Qiuyue Yang[1], Weipeng Wang[1], Zhuanqing Chen[1], Tengfei Mai[1], Yingying Zou[1], Guoxiang Liu[4] and Wei Hao[5*]*

[1]Henan Collaborative Innovation Center of Modern Biological Breeding, Henan Institute of Sciences and Technology, Xinxiang, China, [2]State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang, China, [3]Beijing River and Lake Management Office, Beijing, China, [4]Key Laboratory of Tobacco Improvement and Biotechnology, Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao, China, [5]College of Medical Technology, Beihua University, Jilin City, China

OPEN ACCESS

Edited by:
Yanqiang Li,
Boston Children's Hospital and Harvard Medical School, United States

Reviewed by:
Mingkun Huang,
The Chinese University of Hong Kong, China
Zhengrui Qin,
Zhejiang University, China

***Correspondence:**
Guoxiang Liu
liuguoxiang@caas.cn
Wei Hao
haowei9111@gmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 27 August 2021
Accepted: 04 October 2021
Published: 18 October 2021

Citation:
Qin T, Zhao P, Sun J, Zhao Y, Zhang Y, Yang Q, Wang W, Chen Z, Mai T, Zou Y, Liu G and Hao W (2021) Research Progress of PPR Proteins in RNA Editing, Stress Response, Plant Growth and Development. Front. Genet. 12:765580. doi: 10.3389/fgene.2021.765580

RNA editing is a posttranscriptional phenomenon that includes gene processing and modification at specific nucleotide sites. RNA editing mainly occurs in the genomes of mitochondria and chloroplasts in higher plants. In recent years, pentatricopeptide repeat (PPR) proteins, which may act as trans-acting factors of RNA editing have been identified, and the study of PPR proteins has become a research focus in molecular biology. The molecular functions of these proteins and their physiological roles throughout plant growth and development are widely studied. In this minireview, we summarize the current knowledge of the PPR family, hoping to provide some theoretical reference for future research and applications.

Keywords: pentatricopeptide repeat, RNA editing, biogenesis, development, mechanism

INTRODUCTION

PPR family is one of the largest gene families in higher plants. PPR proteins contain an array of 2–30 tandem repetitions of a degraded unit containing 30–40 amino acid (aa) motifs (Lurin et al., 2004). PPR proteins are classified into two subfamilies based on their domain architecture: P and PPR-like (PLS), which are distinguished by motifs with no space and motifs with interspaced PPR-like motifs respectively. The PLS subfamily can be subdivided into five subgroups based on domain assembly at the C-terminus of a PPR protein: PLS, E1, E2, E+, and DYW (**Figure 1**) (Cheng et al., 2016; Xing et al., 2018). PPR proteins have been found in a variety of terrestrial plants since their discovery in yeast (*Saccharomyces cerevisiae* L.) (Manthey and McEwen, 1995). To date, PPR proteins have been found in many different plants, including *Arabidopsis* (Lurin et al., 2004), foxtail millet (Xing et al., 2018), poplar (Liu et al., 2016), maize (Chen L. et al., 2018), and rice (Chen G. et al., 2018), containing 441, 486, 626, 491, and 477 members of the PPR family, respectively. Additionally, PPR proteins have been discovered to have RNA-binding characteristics, allowing them to mediate gene expression via posttranscriptional mechanisms involving transcripts in the mitochondria, chloroplast, and nucleus. PPR proteins play vital roles in the plant organelle RNA editing machinery. PPRs could not only act as site recognition factors but also bind to *cis*-elements specifically. The resultant PPR–RNA complex and other editing factors, such as ORRM proteins and MORF proteins, can form a higher ordered editosome. As a result of their participation in different posttranscriptional processes, such as RNA editing (Hayes et al., 2015), RNA splicing (Ichinose et al., 2012), and RNA processing (Hao et al., 2019), PPR proteins are considered to have a substantial influence on organelle stability, including

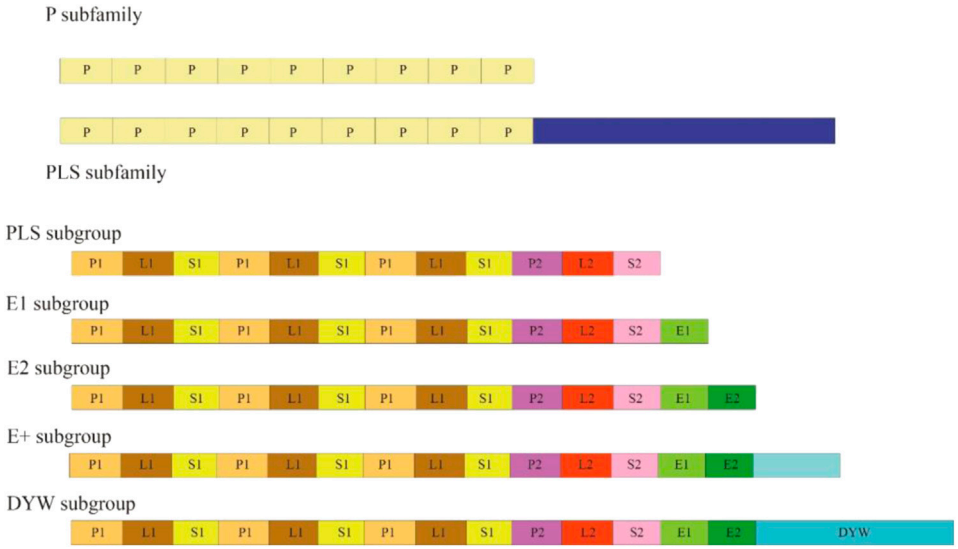

FIGURE 1 | The architecture of different PPR subfamilies and subgroups.

biogenesis and function. Furthermore, plant growth and development have been linked to the activities of PPR proteins. Previously, we briefly reviewed RNA editing in plant organelles, including the factors and mechanism of RNA editing, the editing events identified through deep sequencing data, and the roles of RNA editing (Ichinose et al., 2012). In this review, we emphasize on the recent discoveries of PPR proteins, including their roles in manipulating CMS-related genes, chloroplast biogenesis, embryogenesis, and stress responses. Furthermore, we also discuss the roles of PPR proteins in fruit growth, ripening, plant flesh colour, and fibre development.

The number of motifs in each protein can vary from 2 to 35, and the first motif can be any of P, P1, L1, S1 or SS. The E+ subgroup consists of proteins with a degenerate or truncated DYW domain.

PPR Proteins Manipulate Cytoplasmic Male Sterility-Related Genes

Plant cytoplasmic male sterility (CMS) is a maternally inherited trait that maintains female fertility but results in abortive pollen. At present, CMS is considered to be jointly controlled by mitochondrial genes and their corresponding nuclear restorer (RF) genes, which can change the expression of CMS-related genes in mitochondria. Most *Rf* genes belong to the PPR gene family with several exceptions. Radish Rfo is a PPR protein containing 16 motifs. Inhibition of the *Rfo* translation process by binding to an *orf125*-transcribed mRNA inhibits the accumulation of the CMS-inducible protein ORF125 and restores CMS fertility (Koizuka et al., 2003). In addition, it is worth noting that sorghum *Rf1* encodes PPR13, which belongs to the PLS-E class and can restore male fertility by recruiting RNA editing enzymes (Klein et al., 2005). *OsRF1A* encodes pentatricopeptide repeat proteins, plays an additional role in promoting the editing of *atp6* mRNAs, and restores cytoplasmic male sterility in rice (Wang et al., 2006). Moreover, the PPR protein OsRF5 forms a complex with GRP162 to process *atp6-orf79* and restore fertility in the Honglian CMS line (Hu et al., 2012). Similarly, the interaction between the PPR proteins OsRF6 and Oshxk6 for coprocessing *atp6-orf79* rescues the male fertility of Honglian CMS lines (Huang et al., 2015). PPR756, a member of the PLS-E subclass, participates in RNA editing events of *atp6*, *ccmC*, and *nad7*. The loss of PPR756 could cause abortive pollen development in rice (Zhang Q. et al., 2020). The P-subfamily PPR protein OsPPR939, which can be phosphorylated by OsS6K1, regulates plant growth and pollen development by splicing mitochondrial *nad5* introns 1, 2, and 3 (Zheng et al., 2021). PPS1 is a mitochondria-localized PPR protein, while OsPGLl is a PPR protein that is localized to both mitochondria and chloroplasts. PPS1 is involved in 5 consecutive editing sites of the *nad3* transcript in mitochondria, while OsPGL1 is involved in RNA editing at a single site in both chloroplast and mitochondrial transcripts. The function of these two PPR proteins results in a decrease in the RNA editing efficiency at the specific site and ultimately in a defective phenotype concerning reproductive growth in rice.

PPRs Are Required for Chloroplast Biogenesis

Previous studies have reported that 21 PPR and PPR-related proteins are required for plastid RNA editing in thale cress (Lu, 2018). Two PPR proteins (OsPPR4 and OsPPR6) have been shown to participate in the editing of a single plastid RNA editing site in rice; they are also involved in the editing of chloroplast RNA and are required for chloroplast biogenesis (Asano et al., 2013; Tang et al., 2017). The PPR protein AtPDM2 is located in plastids and regulates the expression of plastid genes related to chloroplast development by interacting

with the organelle RNA editing factors morf2 and morf9. T-DNA insertion of the *AtPDM2* leads to the loss of pigment in *Arabidopsis* cotyledons and cotyledon albinism (Du et al., 2017). Another PPR protein in *Arabidopsis*, Hcf152 is located in chloroplasts and regulates the metabolism of chloroplast RNA by combining with the *petb* operon. This gene mutation will affect the accumulation of the cytochrome b6f complex (Meierhoff et al., 2003); A PPR gene in Poaceae, *Clb19* is required for editing the chloroplast transcripts *rpoA* and *rpoB*. Mutation of this gene will lead to impaired chloroplast development, a yellowing phenotype of seedlings and even death (Hein and Knoop, 2018). OsPPR16, a PLS-DYW subfamily PPR protein, is responsible for RNA editing of the RNA polymerase subunit RpoB and affects Chl synthesis and efficient chloroplast development in rice (Huang et al., 2020). The PPR protein DUA1 interacts with sigma factor 1 to form the PPR-SIG complex, and the module regulates chloroplast gene expression and chloroplast development in response to light and temperature (Du et al., 2021). The PPR protein OsPGL1 disrupts chloroplast RNA editing of *ndhD-878*, which is involved in the conversion of serine to leucine. Loss of OsPGL1 leads to the dysfunction of chloroplasts and the photosynthetic complex (Xiao et al., 2018). The PPR-SMR protein ATP4 participates in C-to-U editing of *rps8* RNA in rice and maize, which is required for the formation of photosynthetic complexes (Zhang J. et al., 2020).

PPRs Regulate Embryogenesis

Previous studies have demonstrated that PPR proteins are essential for *Arabidopsis* and maize kernel formation, with the loss of function of specific PPR proteins resulting in empty pericarps and tiny, malformed kernels in various genetic backgrounds. AtEmb175 is located in chloroplasts and is the first PPR gene found in *Arabidopsis* that is related to early embryo death. Its mutation will lead to continuous cell division, embryo expansion, and abnormal tissue formation (Cushing et al., 2005). *ZmSmk1* encodes a protein containing a PPR structural domain that regulates seed embryo and endosperm development by regulating RNA editing of the mitochondrial gene *nad7* in maize (Li et al., 2014). The maize PPR-like protein EMP9 (EMPTY PERICARP9) regulates seed development by regulating RNA editing of the mitochondrial genes *ccmB* and *rps4* (Yang et al., 2017). ZmEMP21, a PPR-DYW protein that is needed for the editing of 81 mitochondrial target sites, is required for mitochondrial complex assembly as well as embryo and endosperm development (Wang et al., 2019). ZmPPR-SMR1 interacts with ZmCSF1 and is essential for the splicing of numerous group II introns, mitochondrial functions, embryogenesis, and endosperm development (Chen et al., 2019). The E-subgroup PPR protein DEK55 affects mitochondrial RNA processing, which is important for maize kernel formation (Ren et al., 2020). ZmPPR27 interacts with ZmMORF1 (MULTIPLE ORGANELLAR RNA EDITING FACTOR 1), regulates the RNA editing rate of mitochondrial genes such as *ccmFN* and affects the formation of the key mitochondrial protein complex, thus hindering seed embryo and endosperm development (Liu et al., 2020). The maize *DEK46* (DEFECTIVE KERNEL 46) gene encodes a protein containing a PPR domain that edits a specific site in the intron of the mitochondrial *nad7* gene. When DEK46 is functionally absent, the percentage of selective splicing of the intron of the mitochondrial *nad7* gene is reduced, which in turn affects normal seed development in maize (Xu et al., 2020). A DYW domain–containing PPR protein, PPR2263, was found to be involved in RNA editing in the mitochondrial NADH dehydrogenase 5 (*nad5*) and cytochrome b (*cob*) in maize. The *ppr2263* mutant showed reduced embryo and endosperm growth, resulting in growth defects in kernels and seedlings (Sosso et al., 2012). A PPR protein, defective kernel 2 (Dek2), is required for *nad1* mRNA splicing in maize. The *dek2* mutant displayed small kernel and tardy development. (Qi et al., 2017). Another E+ subgroup PPR protein, DEK40, involved in the processing of cox3, nad2, and nad5 was identified to be essential for mitochondrial function and kernel development in maize (Ren et al., 2019).

PPRs Participate in Stress Responses

RNA editing may have contributed to the adaptation of land plants to extreme temperature, UV, and oxidative stress during the early stages of land plant formation (Fujii and Small, 2011). Increasing molecular evidence has revealed that many PPRs are involved in the response to a variety of biotic and abiotic stresses. In *Arabidopsis*, salt, oxidative, and ABA stressors all increased the expression of the *PPR96* gene (Oren et al., 2001). The PPR protein GUN1 is associated with plastid-to-nucleus retrograde communication, control of *ABI4* expression, and photooxidative stress responses in *Arabidopsis* (Koussevitzky et al., 2007). In *Arabidopsis*, PPR40 is known to offer a signaling connection between mitochondrial electron transport elements. PPR40 knockout resulted in increased reactive oxygen species (ROS) accumulation, lipid peroxidation, and superoxide dismutase activity (Zsigmond et al., 2008). The PPR protein ABA overly sensitive 5 (ABO5/At1g51965) is required for NADH dehydrogenase subunit 2 (NAD2) intron 3 splicing in mitochondria. Compared to the wild type, the *abo5* mutant accumulated more H_2O_2 in roots (Liu et al., 2010). MITOCHONDRIAL RNA EDITING FACTOR 11 (MEF11)/ LOVASTATIN INSENSITIVE 1 (LOI1) controls isoprenoid production, which is known to influence defense gene expression in response to wounding and pathogen infection (Kobayashi et al., 2007; Tang et al., 2010).

The *Arabidopsis* PPR-like protein AHG11 (ABA HYPERSENSITIVE GERMINATION 11) can edit the mRNA of the mitochondrial gene *nad4* to maintain normal intracellular levels of reactive oxygen species (Murayama et al., 2012). PGN (PENTATRICOPEPTIDE REPEAT PROTEIN FOR GERMINATION ON NACl) has been shown to be involved in biotic and abiotic stress responses (Laluk et al., 2011). Functional disruption of the PPR protein SLG1 impacts mitochondrial RNA editing, plant growth, and abiotic stress responses in *Arabidopsis* (Yuan and Liu, 2012).

Plant development is also regulated by another PPR protein, SLO2. Stress-sensitive genes have higher transcript levels in the *slo2* mutant. Furthermore, the *slo2* mutant is hypersensitive to

osmotic and ABA stressors at various phases of seed germination, although their mature plants have a high resistance to salt and drought stresses (Zhu et al., 2012; Zhu et al., 2014). SVR7 (SUPPRESSOR OF VARIATION 7) is needed for chloroplast ATP synthase subunit translation in *Arabidopsis*. SVR7 knockout led to increased ROS production, increased sensitivity to H_2O_2, and decreased photosynthetic activity (Lv et al., 2014). SOAR1 (suppressor of ABAR-overexpressor 1), which encodes a nucleocytoplasmic localized PPR protein, has recently been discovered to be a positive regulator of the responses to different stresses, including those to drought, salt, and cold (Jiang et al., 2015).

PPRs Regulate Fruit Growth, Ripening, Plant Flesh Colour, and Fibre Development

With the discovery of growing molecular evidence, researchers have found that PPR genes are involved in a variety of fruit functions, including as growth, ripening, colouration, and fiber formation. *GUN1*, which encodes a plastid-localized PPR protein, has been involved in the plastid-to-nucleus retrograde signalling route during tomato fruit development and ripening (Pesaresi et al., 2014). A number of tomato mutants, such as *Cnr*, were found to have dramatically reduced the expression of ripening-related PPR genes, resulting in mature fruits with colorless pericarp tissue, indicating that PPR proteins play a role in the development of fruits (Eriksson et al., 2004). *CmPPR1*, which encodes a plastid-targeted P-type PPR protein in melon (*Cucumis melo* L.), has been identified as a potential main quantitative trait locus (QTL) that determines flesh colour intensity (Galpaz et al., 2018). It was shown that genotyping 70 lines utilizing four SNPs from four *ClaPPRs* resulted in match rates above 0.87 for each validated SNP in association with the distinct phenotypes of skin colour. These findings contribute to a better knowledge of PPR genes and their functions in watermelon fruit growth and ripening, which may be useful in watermelon cultivar improvement (Subburaj et al., 2020). *GhImA*, which encodes a PPR protein, is involved in mitochondrial *nad7* splicing, respiratory metabolism, and cotton fibre formation via ATP supply and ROS balancing (Zhang et al., 2021).

CONCLUSION AND PERSPECTIVES

In general, PPR proteins participate in plant growth and development by RNA editing or RNA stabilization and splicing. Recently, many attempts have been made to engineer special proteins for efficient RNA editing in plant organelles. Shen et al. observed the interactions between different PPR codes and RNA bases at the atomic level, revealing the molecular basis for the modular and specific recognition patterns of the RNA bases U, C, A, and G (Zhang et al., 2021). PPR family proteins have the ability to enter organelles and bind to single-stranded RNA, and more recognition codes are being studied and verified. On this basis, according to the single-stranded sequence characteristics of target RNA, the corresponding PPR proteins have been artificially designed. Developing RNA-targeting tools will likely accelerate the functional study of RNA editing and the biological study of chloroplasts and mitochondria. By fusing the corresponding functional domains, the artificial design of fused PPR proteins is expected to become the next generation of biotechnology that can regulate the expression of organelle genes.

AUTHOR CONTRIBUTIONS

WH and GL contributed equally to the design and coordination of the study; TQ, PZ, and JS collected the data; With the help from the YPZ, QY, YXZ, WW, ZC, TM, and YYZ wrote the manuscript. All of the authors reviewed and edited the manuscript.

FUNDING

This research was supported by Science and Technology Program of Sichuan Province (2020YJ0406), National Natural Science Foundation of China (31900256), Natural Science Foundation of Shandong Province (ZR2020QC026), Modern Agricultural Industry Technical Economic Evaluation System Green Development Position of Henan Province, Soft Science Project of Henan Province (202400410185) and State Key Laboratory of Cotton Biology Open Fund (CB2021A02).

REFERENCES

Asano, T., Miyao, A., Hirochika, H., Kikuchi, S., and Kadowaki, K.-i. (2013). A Pentatricopeptide Repeat Gene of rice Is Required for Splicing of Chloroplast Transcripts and RNA Editing of ndhA. *Plant Biotechnol.* 12, 1217. doi:10.5511/plantbiotechnology.12.1217a

Chen, G., Zou, Y., Hu, J., and Ding, Y. (2018a). Genome-wide Analysis of the rice PPR Gene Family and Their Expression Profiles under Different Stress Treatments. *BMC genomics* 19 (1), 1–14. doi:10.1186/s12864-018-5088-9

Chen, L., Li, Y. X., Li, C., Shi, Y., Song, Y., Zhang, D., et al. (2018b). Genome-wide Analysis of the Pentatricopeptide Repeat Gene Family in Different maize Genomes and its Important Role in Kernel Development. *BMC Plant Biol.* 18 (1), 366–414. doi:10.1186/s12870-018-1572-2

Chen, Z., Wang, H.-C., Shen, J., Sun, F., Wang, M., Xu, C., et al. (2019). PPR-SMR1 Is Required for the Splicing of Multiple Mitochondrial Introns, Interacts with Zm-mCSF1, and Is Essential for Seed Development in maize. *J. Exp. Bot.* 70 (19), 5245–5258. doi:10.1093/jxb/erz305

Cheng, S., Gutmann, B., Zhong, X., Ye, Y., Fisher, M. F., Bai, F., et al. (2016). Redefining the Structural Motifs that Determine RNA Binding and RNA Editing by Pentatricopeptide Repeat Proteins in Land Plants. *Plant J.* 85 (4), 532–547. doi:10.1111/tpj.13121

Cushing, D. A., Forsthoefel, N. R., Gestaut, D. R., and Vernon, D. M. (2005). Arabidopsis Emb175 and Other Ppr Knockout Mutants Reveal Essential Roles for Pentatricopeptide Repeat (PPR) Proteins in Plant Embryogenesis. *Planta* 221 (3), 424–436. doi:10.1007/s00425-004-1452-x

Du, L., Zhang, J., Qu, S., Zhao, Y., Su, B., Lv, X., et al. (2017). The Pentratricopeptide Repeat Protein Pigment-Defective Mutant2 Is Involved in the Regulation of Chloroplast Development and Chloroplast Gene Expression in Arabidopsis. *Plant Cel Physiol.* 58 (4), 747–759. doi:10.1093/pcp/pcx004

Du, Y., Mo, W., Ma, T., Tang, W., Tian, L., and Lin, R. (2021). A Pentatricopeptide Repeat Protein DUA1 Interacts with Sigma Factor 1 to Regulate Chloroplast

Gene Expression in Rice. *Photosynth Res.* 147 (2), 131–143. doi:10.1007/s11120-020-00793-0

Eriksson, E. M., Bovy, A., Manning, K., Harrison, L., Andrews, J., De Silva, J., et al. (2004). Effect of the Colorless Non-ripening Mutation on Cell wall Biochemistry and Gene Expression during Tomato Fruit Development and Ripening. *Plant Physiol.* 136 (4), 4184–4197. doi:10.1104/pp.104.045765

Fujii, S., and Small, I. (2011). The Evolution of RNA Editing and Pentatricopeptide Repeat Genes. *New Phytol.* 191 (1), 37–47. doi:10.1111/j.1469-8137.2011.03746.x

Galpaz, N., Gonda, I., Shem-Tov, D., Barad, O., Tzuri, G., Lev, S., et al. (2018). Deciphering Genetic Factors that Determine Melon Fruit-quality Traits Using RNA -Seq-based High-resolution QTL and eQTL Mapping. *Plant J.* 94 (1), 169–191. doi:10.1111/tpj.13838

Hao, Y., Wang, Y., Wu, M., Zhu, X., Teng, X., Sun, Y., et al. (2019). The Nuclear-Localized PPR Protein OsNPPR1 Is Important for Mitochondrial Function and Endosperm Development in rice. *J. Exp. Bot.* 70 (18), 4705–4720. doi:10.1093/jxb/erz226

Hayes, M. L., Dang, K. N., Diaz, M. F., and Mulligan, R. M. (2015). A Conserved Glutamate Residue in the C-Terminal Deaminase Domain of Pentatricopeptide Repeat Proteins Is Required for RNA Editing Activity. *J. Biol. Chem.* 290 (16), 10136–10142. doi:10.1074/jbc.m114.631630

Hein, A., and Knoop, V. (2018). Expected and Unexpected Evolution of Plant RNA Editing Factors CLB19, CRR28 and RARE1: Retention of CLB19 Despite a Phylogenetically Deep Loss of its Two Known Editing Targets in Poaceae. *BMC Evol. Biol.* 18 (1), 85–15. doi:10.1186/s12862-018-1203-4

Hu, J., Wang, K., Huang, W., Liu, G., Gao, Y., Wang, J., et al. (2012). The Rice Pentatricopeptide Repeat Protein RF5 Restores Fertility in Hong-Lian Cytoplasmic Male-Sterile Lines via a Complex with the Glycine-Rich Protein GRP162. *The Plant Cell* 24 (1), 109–122. doi:10.1105/tpc.111.093211

Huang, W., Yu, C., Hu, J., Wang, L., Dan, Z., Zhou, W., et al. (2015). Pentatricopeptide-repeat Family Protein RF6 Functions with Hexokinase 6 to rescue rice Cytoplasmic Male Sterility. *Proc. Natl. Acad. Sci. USA* 112 (48), 14984–14989. doi:10.1073/pnas.1511748112

Huang, W., Zhang, Y., Shen, L., Fang, Q., Liu, Q., Gong, C., et al. (2020). Accumulation of the RNA Polymerase Subunit RpoB Depends on RNA Editing by OsPPR16 and Affects Chloroplast Development during Early Leaf Development in rice. *New Phytol.* 228 (4), 1401–1416. doi:10.1111/nph.16769

Ichinose, M., Tasaki, E., Sugita, C., and Sugita, M. (2012). A PPR-DYW Protein Is Required for Splicing of a Group II Intron of Cox1 Pre-mRNA in Physcomitrella Patens. *Plant J.* 70 (2), 271–278. doi:10.1111/j.1365-313x.2011.04869.x

Jiang, S. C., Mei, C., Liang, S., Yu, Y. T., Lu, K., Wu, Z., et al. (2015). Crucial Roles of the Pentatricopeptide Repeat Protein SOAR1 in Arabidopsis Response to Drought, Salt and Cold Stresses. *Plant Mol. Biol.* 88 (4), 369–385. doi:10.1007/s11103-015-0327-9

Klein, R. R., Klein, P. E., Mullet, J. E., Minx, P., Rooney, W. L., and Schertz, K. F. (2005). Fertility Restorer Locus Rf1 of Sorghum (Sorghum Bicolor L.) Encodes a Pentatricopeptide Repeat Protein Not Present in the Colinear Region of rice Chromosome 12. *Theor. Appl. Genet.* 111 (6), 994–1012. doi:10.1007/s00122-005-2011-y

Kobayashi, K., Suzuki, M., Tang, J., Nagata, N., Ohyama, K., Seki, H., et al. (2007). Lovastatin Insensitive 1, a Novel Pentatricopeptide Repeat Protein, Is a Potential Regulatory Factor of Isoprenoid Biosynthesis in Arabidopsis. *Plant Cel. Physiol.* 48 (2), 322–331. doi:10.1093/pcp/pcm005

Koizuka, N., Imai, R., Fujimoto, H., Hayakawa, T., Kimura, Y., Kohno-Murase, J., et al. (2003). Genetic Characterization of a Pentatricopeptide Repeat Protein Gene, Orf687, that Restores Fertility in the Cytoplasmic Male-Sterile Kosena Radish. *Plant J.* 34 (4), 407–415. doi:10.1046/j.1365-313x.2003.01735.x

Koussevitzky, S., Nott, A., Mockler, T. C., Hong, F., Sachetto-Martins, G., Surpin, M., et al. (2007). Signals from Chloroplasts Converge to Regulate Nuclear Gene Expression. *Science* 316 (5825), 715–719. doi:10.1126/science.1140516

Laluk, K., AbuQamar, S., and Mengiste, T. (2011). The Arabidopsis Mitochondria-Localized Pentatricopeptide Repeat Protein PGN Functions in Defense against Necrotrophic Fungi and Abiotic Stress Tolerance. *Plant Physiol.* 156 (4), 2053–2068. doi:10.1104/pp.111.177501

Li, X.-J., Zhang, Y.-F., Hou, M., Sun, F., Shen, Y., Xiu, Z.-H., et al. (2014). Small Kernel 1encodes a Pentatricopeptide Repeat Protein Required for Mitochondrialnad7transcript Editing and Seed Development in maize(Zea mays)and rice(Oryza Sativa). *Plant J.* 79 (5), 797–809. doi:10.1111/tpj.12584

Liu, J. M., Xu, Z. S., Lu, P. P., Li, W. W., Chen, M., Guo, C. H., et al. (2016). Genome-wide Investigation and Expression Analyses of the Pentatricopeptide Repeat Protein Gene Family in Foxtail Millet. *BMC genomics* 17 (1), 1–16. doi:10.1186/s12864-016-3184-2

Liu, R., Cao, S.-K., Sayyed, A., Yang, H.-H., Zhao, J., Wang, X., et al. (2020). The DYW-Subgroup Pentatricopeptide Repeat Protein PPR27 Interacts with ZmMORF1 to Facilitate Mitochondrial RNA Editing and Seed Development in maize. *J. Exp. Bot.* 71 (18), 5495–5505. doi:10.1093/jxb/eraa273

Liu, Y., He, J., Chen, Z., Ren, X., Hong, X., and Gong, Z. (2010). ABA Overly-Sensitive 5 (ABO5), Encoding a Pentatricopeptide Repeat Protein Required for Cis-Splicing of Mitochondrial Nad2 Intron 3, Is Involved in the Abscisic Acid Response in Arabidopsis. *Plant J.* 63 (5), 749–765. doi:10.1111/j.1365-313x.2010.04280.x

Lu, Y. (2018). RNA Editing of Plastid-Encoded Genes. *Photosynt.* 56 (1), 48–61. doi:10.1007/s11099-017-0761-9

Lurin, C., Andreés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., et al. (2004). Genome-Wide Analysis of Arabidopsis Pentatricopeptide Repeat Proteins Reveals Their Essential Role in Organelle Biogenesis[W]. *The Plant Cell* 16 (8), 2089–2103. doi:10.1105/tpc.104.022236

Lv, H.-X., Huang, C., Guo, G.-Q., and Yang, Z.-N. (2014). Roles of the Nuclear-Encoded Chloroplast SMR Domain-Containing PPR Protein SVR7 in Photosynthesis and Oxidative Stress Tolerance in Arabidopsis. *J. Plant Biol.* 57 (5), 291–301. doi:10.1007/s12374-014-0041-1

Manthey, G. M., and McEwen, J. E. (1995). The Product of the Nuclear Gene PET309 Is Required for Translation of Mature mRNA and Stability or Production of Intron-Containing RNAs Derived from the Mitochondrial COX1 Locus of *Saccharomyces cerevisiae*. *EMBO J.* 14 (16), 4031–4043. doi:10.1002/j.1460-2075.1995.tb00074.x

Meierhoff, K., Felder, S., Nakamura, T., Bechtold, N., and Schuster, G. (2003). HCF152, an Arabidopsis RNA Binding Pentatricopeptide Repeat Protein Involved in the Processing of Chloroplast psbB-psbT-psbH-petB-petD RNAs. *Plant Cell* 15 (6), 1480–1495. doi:10.1105/tpc.010397

Murayama, M., Hayashi, S., Nishimura, N., Ishide, M., Kobayashi, K., Yagi, Y., et al. (2012). Isolation of Arabidopsis Ahg11, a Weak ABA Hypersensitive Mutant Defective in Nad4 RNA Editing. *J. Exp. Bot.* 63 (14), 5301–5310. doi:10.1093/jxb/ers188

Oren, R., Ellsworth, D. S., Johnsen, K. H., Phillips, N., Ewers, B. E., Maier, C., et al. (2001). Soil Fertility Limits Carbon Sequestration by forest Ecosystems in a CO2-enriched Atmosphere. *Nature* 411 (6836), 469–472. doi:10.1038/35078064

Pesaresi, P., Mizzotti, C., Colombo, M., and Masiero, S. (2014). Genetic Regulation and Structural Changes during Tomato Fruit Development and Ripening. *Front. Plant Sci.* 5, 124. doi:10.3389/fpls.2014.00124

Qi, W., Yang, Y., Feng, X., Zhang, M., and Song, R. (2017). Mitochondrial Function and maize Kernel Development Requires Dek2, a Pentatricopeptide Repeat Protein Involved in Nad1 mRNA Splicing. *Genetics* 205 (1), 239–249. doi:10.1534/genetics.116.196105

Ren, R. C., Yan, X. W., Zhao, Y. J., Wei, Y. M., Lu, X., Zang, J., et al. (2020). The Novel E-Subgroup Pentatricopeptide Repeat Protein DEK55 Is Responsible for RNA Editing at Multiple Sites and for the Splicing of Nad1 and Nad4 in maize. *BMC Plant Biol.* 20 (1), 1–15. doi:10.1186/s12870-020-02765-x

Ren, R. C., Lu, X., Zhao, Y. J., Wei, Y. M., Wang, L. L., Zhang, L., et al. (2019). Pentatricopeptide Repeat Protein DEK40 Is Required for Mitochondrial Function and Kernel Development in maize. *J. Exp. Bot.* 70 (21), 6163–6179. doi:10.1093/jxb/erz391

Sosso, D., Mbelo, S., Vernoud, V., Gendrot, G., Dedieu, A., Chambrier, P., et al. (2012). PPR2263, a DYW-Subgroup Pentatricopeptide Repeat Protein, Is Required for Mitochondrial Nad5 and Cob Transcript Editing, Mitochondrion Biogenesis, and maize Growth. *Plant Cell* 24 (2), 676–691. doi:10.1105/tpc.111.091074

Subburaj, S., Tu, L., Lee, K., Park, G.-S., Lee, H., Chun, J.-P., et al. (2020). A Genome-Wide Analysis of the Pentatricopeptide Repeat (PPR) Gene Family and PPR-Derived Markers for Flesh Color in Watermelon (*Citrullus lanatus*). *Genes* 11 (10), 1125. doi:10.3390/genes11101125

Tang, J., Zhang, W., Wen, K., Chen, G., Sun, J., Tian, Y., et al. (2017). OsPPR6, a Pentatricopeptide Repeat Protein Involved in Editing and Splicing Chloroplast

RNA, Is Required for Chloroplast Biogenesis in rice. *Plant Mol. Biol.* 95 (4), 345–357. doi:10.1007/s11103-017-0654-0

Tang, J., Kobayashi, K., Suzuki, M., Matsumoto, S., and Muranaka, T. (2010). The Mitochondrial PPR Protein LOVASTATIN INSENSITIVE 1 Plays Regulatory Roles in Cytosolic and Plastidial Isoprenoid Biosynthesis through RNA Editing. *Plant J.* 61 (3), 456–466. doi:10.1111/j.1365-313x.2009.04082.x

Wang, Y., Liu, X.-Y., Yang, Y.-Z., Huang, J., Sun, F., Lin, J., et al. (2019). Empty Pericarp21 Encodes a Novel PPR-DYW Protein that Is Required for Mitochondrial RNA Editing at Multiple Sites, Complexes I and V Biogenesis, and Seed Development in maize. *Plos Genet.* 15 (8), e1008305. doi:10.1371/journal.pgen.1008305

Wang, Z., Zou, Y., Li, X., Zhang, Q., Chen, L., Wu, H., et al. (2006). Cytoplasmic Male Sterility of rice with Boro II Cytoplasm Is Caused by a Cytotoxic Peptide and Is Restored by Two Related PPR Motif Genes via Distinct Modes of mRNA Silencing. *The Plant Cell* 18 (3), 676–687. doi:10.1105/tpc.105.038240

Xiao, H., Zhang, Q., Qin, X., Xu, Y., Ni, C., Huang, J., et al. (2018). Rice PPS1 Encodes a DYW Motif-Containing Pentatricopeptide Repeat Protein Required for Five Consecutive RNA-Editing Sites of Nad3 in Mitochondria. *New Phytol.* 220 (3), 878–892. doi:10.1111/nph.15347

Xing, H., Fu, X., Yang, C., Tang, X., Guo, L., Li, C., et al. (2018). Genome-wide Investigation of Pentatricopeptide Repeat Gene Family in poplar and Their Expression Analysis in Response to Biotic and Abiotic Stresses. *Sci. Rep.* 8 (1), 2817–2819. doi:10.1038/s41598-018-21269-1

Xu, C., Song, S., Yang, Y. Z., Lu, F., Zhang, M. D., Sun, F., et al. (2020). DEK46 Performs C-to-U Editing of a Specific Site in Mitochondrial Nad7 Introns that Is Critical for Intron Splicing and Seed Development in maize. *Plant J.* 103 (5), 1767–1782. doi:10.1111/tpj.14862

Yang, Y. Z., Ding, S., Wang, H. C., Sun, F., Huang, W. L., Song, S., et al. (2017). The Pentatricopeptide Repeat Protein EMP 9 Is Required for Mitochondrial ccmB and Rps4 Transcript Editing, Mitochondrial Complex Biogenesis and Seed Development in maize. *New Phytol.* 214 (2), 782–795. doi:10.1111/nph.14424

Yuan, H., and Liu, D. (2012). Functional Disruption of the Pentatricopeptide Protein SLG1 Affects Mitochondrial RNA Editing, Plant Development, and Responses to Abiotic Stresses in Arabidopsis. *Plant J.* 70 (3), 432–444. doi:10.1111/j.1365-313x.2011.04883.x

Zhang, D., Chen, C., Wang, H., Niu, E., Zhao, P., Fang, S., et al. (2021). Cotton Fiber Development Requires the Pentatricopeptide Repeat Protein GhIm for Splicing of Mitochondrial Nad7 mRNA. *Genetics* 217 (1), 1–17. doi:10.1093/genetics/iyaa017

Zhang, J., Guo, Y., Fang, Q., Zhu, Y., Zhang, Y., Liu, X., et al. (2020a). The PPR-SMR Protein ATP4 Is Required for Editing the Chloroplast Rps8 mRNA in rice and maize. *Plant Physiol.* 184 (4), 2011–2021. doi:10.1104/pp.20.00849

Zhang, Q., Xu, Y., Huang, J., Zhang, K., Xiao, H., Qin, X., et al. (2020b). The rice Pentatricopeptide Repeat Protein PPR756 Is Involved in Pollen Development by Affecting Multiple RNA Editing in Mitochondria. *Front. Plant Sci.* 11, 749. doi:10.3389/fpls.2020.00749

Zheng, P., Liu, Y., Liu, X., Huang, Y., Sun, F., Wang, W., et al. (2021). OsPPR939, a Nad5 Splicing Factor, Is Essential for Plant Growth and Pollen Development in rice. *Theor. Appl. Genet.* 134 (3), 923–940. doi:10.1007/s00122-020-03742-6

Zhu, Q., Dugardeyn, J., Zhang, C., Mühlenbock, P., Eastmond, P. J., Valcke, R., et al. (2014). The *Arabidopsis thaliana* RNA Editing Factor SLO2, Which Affects the Mitochondrial Electron Transport Chain, Participates in Multiple Stress and Hormone Responses. *Mol. Plant* 7 (2), 290–310. doi:10.1093/mp/sst102

Zhu, Q., Dugardeyn, J., Zhang, C., Takenaka, M., Kühn, K., Craddock, C., et al. (2012). SLO2, a Mitochondrial Pentatricopeptide Repeat Protein Affecting Several RNA Editing Sites, Is Required for Energy Metabolism. *Plant J.* 71 (5), 836–849. doi:10.1111/j.1365-313x.2012.05036.x

Zsigmond, L., Rigo´, G., Szarka, A., Sze´kely, G., Otvös, K., Darula, Z., et al. (2008). Arabidopsis PPR40 Connects Abiotic Stress Responses to Mitochondrial Electron Transport. *Plant Physiol.* 146 (4), 1721–1737. doi:10.1104/pp.107.111260

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 29 October 2021
doi: 10.3389/fgene.2021.748279

Construction of Prognostic Risk Model of 5-Methylcytosine-Related Long Non-Coding RNAs and Evaluation of the Characteristics of Tumor-Infiltrating Immune Cells in Breast Cancer

*Zhidong Huang†, Junjing Li†, Jialin Chen and Debo Chen**

Department of Breast Surgery, Quanzhou First Hospital of Fujian Medical University, Quanzhou, China

OPEN ACCESS

Edited by:
Jia Meng,
Xi'an Jiaotong-Liverpool University, China

Reviewed by:
Qiqi Xie,
Affiliated Hospital of Qinghai University, China
Attila Patocs,
Semmelweis University, Hungary

***Correspondence:**
Debo Chen
debochensr@fjmu.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *27 July 2021*
Accepted: *12 October 2021*
Published: *29 October 2021*

Citation:
Huang Z, Li J, Chen J and Chen D (2021) Construction of Prognostic Risk Model of 5-Methylcytosine-Related Long Non-Coding RNAs and Evaluation of the Characteristics of Tumor-Infiltrating Immune Cells in Breast Cancer. Front. Genet. 12:748279. doi: 10.3389/fgene.2021.748279

Purpose: The role of 5-methylcytosine-related long non-coding RNAs (m5C-lncRNAs) in breast cancer (BC) remains unclear. Here, we aimed to investigate the prognostic value, gene expression characteristics, and correlation between m5C-lncRNA risk model and tumor immune cell infiltration in BC.

Methods: The expression matrix of m5C-lncRNAs in BC was obtained from The Cancer Genome Atlas database, and the lncRNAs were analyzed using differential expression analysis as well as univariate and multivariate Cox regression analysis to eventually obtain BC-specific m5C-lncRNAs. A risk model was developed based on three lncRNAs using multivariate Cox regression and the prognostic value, accuracy, as well as reliability were verified. Gene set enrichment analysis (GSEA) was used to analyze the Kyoto Encyclopedia of Genes and Genomes signaling pathway enrichment of the risk model. CIBERSORT algorithm and correlation analysis were used to explore the characteristics of the BC tumor-infiltrating immune cells. Finally, reverse transcription-quantitative polymerase chain reaction was performed to detect the expression level of three lncRNA in clinical samples.

Results: A total of 334 differential m5C-lncRNAs were identified, and three BC-specific m5C-lncRNAs were selected, namely AP005131.2, AL121832.2, and LINC01152. Based on these three lncRNAs, a highly reliable and specific risk model was constructed, which was proven to be closely related to the prognosis of patients with BC. Therefore, a nomogram based on the risk score was built to assist clinical decisions. GSEA revealed that the risk model was significantly enriched in metabolism-related pathways and was associated with tumor immune cell infiltration based on the analysis with the CIBERSORT algorithm.

Conclusion: The efficient risk model based on m5C-lncRNAs associated with cancer metabolism and tumor immune cell infiltration could predict the survival prognosis of patients, and AP005131.2, AL121832.2, and LINC01152 could be novel biomarkers and therapeutic targets for BC.

Keywords: m5C-related lncRNAs, breast cancer, risk model, nomogram, tumor-infiltrating immune cells

INTRODUCTION

Breast cancer (BC) is one of the leading causes of cancer-related deaths in women. According to the most recent data, in 2020, among the 19.3 million new cancer cases, female BC surpassed lung cancer for the first time by 11.7%, becoming the most commonly diagnosed cancer worldwide (Sung et al., 2021). Although the comprehensive treatment model of BC has been continuously improved in recent years, the prognosis of BC is not ideal due to recurrence and distant metastasis (Schwartz and Erban, 2017). In addition, statistics from the American Cancer Society have shown the mortality rate of breast cancer to be the highest among women aged 20–59 years (Sung et al., 2021). Early diagnosis combined with timely treatment is the key to improving the prognosis of patients with BC (DeSantis et al., 2019; Li S. et al., 2021; Lu et al., 2021); therefore, active exploration of novel prognostic biomarkers and building reliable models to further guide the diagnosis and treatment decisions of BC in future have become imperative.

In recent years, aided by the advancement of analytical chemistry tools and the rapid development of genome sequencing technology (Liu L. et al., 2020, 2021; Hasan et al., 2021a; Tang et al., 2021), epigenetic and non-coding RNA research have captured considerable attention. RNA modifications are jointly regulated by three types of effector proteins (writers, readers, and erasers) (Biswas and Rao, 2018). These form a complex protein network that regulates the gene expression process. In addition, RNA modifications, a reversible and dynamic way of regulating genetic information, play an important role in the control of genetic information, whether in coding or non-coding RNAs (Roundtree et al., 2017; Hasan et al., 2021b). Furthermore, RNA modifications have been proven to be closely related to human diseases such as cancer, metabolic diseases, vascular diseases, and neurological diseases (Jonkhout et al., 2017).

Genome-wide association studies (GWAS) have revealed that >93% of disease-linked susceptible areas are located in the non-coding region of the genome (Tak and Farnham, 2015), indicating that non-coding elements might be related to diseases. Long non-coding RNA (lncRNA), a major component of the non-coding genome (Cai et al., 2020), is a type of mRNA-like transcript with a length of >200 nucleotides (Lai et al., 2020). It plays a role in regulating gene expression at multiple levels, including epigenetics and transcriptional and post-transcriptional regulation, by interacting with other biomolecules (Wu et al., 2018). In the past decade, lncRNAs have emerged as potentially important regulators of pathological processes, including cancer. For example, lncRNA *XIST* sponges miR-34a suppresses the expression of WNT1 by binding to its 3′ untranslated regions (3′-UTR), thus inhibiting the proliferation of colon cancer (Sun et al., 2018). The lncRNA CTBP1-AS represses CTBP1 expression *via* a persisting-cell-stimulating factor (PSF)-dependent mechanism that affects prostate cancer progression (Takayama et al., 2013).

RNA modification-related lncRNAs have been confirmed to be associated with head and neck squamous cell carcinoma, colorectal cancer, hepatocellular carcinoma, and liver cancer (Tu et al., 2020; Xue et al., 2020; Zuo et al., 2020; Feng et al., 2021). However, the role of 5-methylcytosine-related lncRNAs (m5C-lncRNAs) in BC cells is still poorly defined. The current study aimed to screen differentially expressed and prognostic m5C-lncRNAs to build a risk model that can reliably predict the prognosis of patients. This could aid the built of a nomogram, which might help with clinical decisions. Lastly, we explored the potential signaling pathways of m5C-lncRNA-mediated tumor development through gene set enrichment analysis (GSEA) and the correlation between risk model and tumor-infiltrating immune cells (TIICs) (**Figure 1**).

MATERIALS AND METHODS

Data Collection and Identification of m5C-lncRNAs

RNA-sequencing, patient characteristics, and clinical information of BC (1109 samples) and normal breast tissue (113 samples) were obtained from The Cancer Genome Atlas (TCGA) (https://portal.gdc.cancer.gov/) database (**Table 1**). Genome names and genotypes were annotated using the Genome Reference Consortium Human Build 38 (GRCH38) data from the Ensembl database (https://asia.ensembl.org) to further screen for lncRNAs and mRNAs. The following 15 m5C regulatory factors were collected based on previous reports: NSUN1, NSUN2, NSUN3, NSUN4, NSUN5, NSUN6, NSUN7, DNMT1, DNMT2, DNMT3A, DNMT3B, TET2, ALYREF, TRDMT1, and YBX1. A total of 422 m5C-lncRNAs were screened using the limma package from R.

Screening and Validation of BC-Specific m5C-lncRNAs

We used a two-step screening method to identify the candidate BC-specific lncRNAs. First, differential m5C-lncRNA expression analysis between BC and normal breast tissue was performed using limma package, and 334 differentially expressed lncRNAs were screened out ($p < 0.05$). Five prognosis-related lncRNAs were identified using univariate Cox regression analysis, and the significant lncRNAs were subsequently included in a multivariate Cox regression analysis ($p < 0.05$). Three m5C-lncRNAs were selected. To further verify the tissue specificity of lncRNAs in BC, Kaplan–Meier (KM) survival analysis and subgroup analysis for the clinicopathological characteristics of the three lncRNAs were performed.

Construction of Prognostic Model and Nomogram

Based on BC-specific m5C-lncRNAs, we constructed a prognostic risk model of m5C-lncRNAs in patients with BC. Multivariate Cox regression was used to calculate their coefficients (Coefi). Then, the fragments per kb per million mapped reads value (*xi*) of each m5C-related lncRNA and Coefi were used to develop a formula for the risk score:

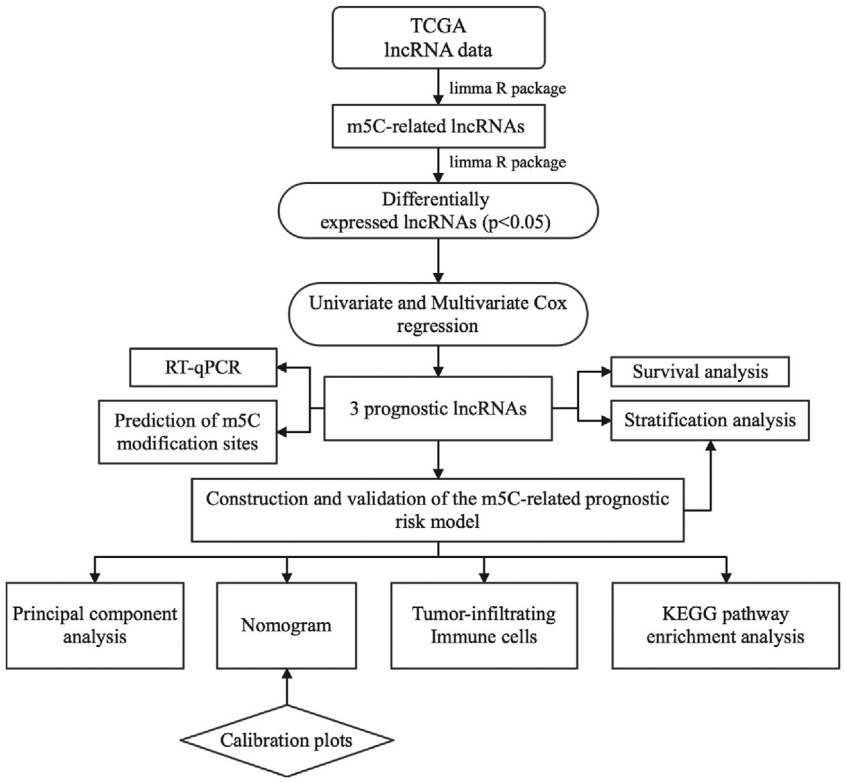

FIGURE 1 | Flowchart for establishing and evaluating prognostic risk model.

$$Risk\ score = \sum_{i=1}^{n} coef\,i \cdot xi$$

All patients with BC were divided into two subgroups (low- and high-risk groups) using the median of all patient risk scores as the cut-off. Stratified analysis of clinical characteristics and difference in expression of the three prognostic lncRNAs in the different risk groups were analyzed by creating a heat map. To test the prognostic value of the risk model, the correlation between patient survival and risk score was determined. Univariate and multivariate Cox analyses, receiver operating characteristic (ROC) curve, as well as principal component analysis (PCA) further proved the independent prognostic value, sensitivity, specificity, and reliability of the risk model. To enable the risk model to make clinical diagnosis and treatment decisions, we constructed a nomogram based on independent prognostic factors, and calibrated the same to evaluate the predicted probabilities of the nomogram.

KEGG Pathway Enrichment Analysis

We explored the potential signaling pathways implicated in the m5C-lncRNA risk model using KEGG pathway enrichment analysis and the GSEA software. The reference gene set was retrieved from c2.cp.kegg.v7.4. symbol files; pathways with | NES | >1; NOM p-value < 0.05, and FDR q-value <0.25 were defined as significantly enriched.

Correlation Analysis of TIICs

Using CIBERSORT L22 as the reference (Newman et al., 2015), we analyzed the m5C-lncRNAs expression matrix using CIBERSORT R script acquired from the CIBERSORT website (http://cibersort.stanford.edu/), and the relative proportions of 22 immune cells were calculated. For each tumor sample, the sum of relative proportion scores of all immune cells was 1. The immune cell subtypes with lower scores were removed. The Spearman correlation analysis was performed on the remaining 21 immune cell types. We identified the low- and high-risk groups to screen the differentially infiltrated immune cells and investigated whether there was a difference in infiltration across the different groups.

RNA Isolation and Quantitative Real Time Polymerase Chain Reaction

To further verify the differential expression of three m5C-lncRNAs in BC, we extracted RNA from fresh frozen tissues using TRIzol reagent (Invitrogen, Carlsbad, CA, United States) and tested the expression levels of all three lncRNAs. cDNA was synthesized using a reverse transcription kit (TaKaRa, Japan). The LightCycler 480 Real-Time PCR System was used to detect the relative expression of the target lncRNAs. GAPDH was used as an internal normalization control. We collected 31 pairs of fresh BC and paracarcinoma tissues from the Department of Breast Surgery, Quanzhou First Hospital, between 2020 and 2021

TABLE 1 | Clinical characteristics of patients with breast cancer.

Variables	No. of patients	Percentage (%)
Age (years)		
≤55	471	42.9
>55	626	57.1
Unknown	19	1.7
Gender		
Female	1085	98.9
Male	12	1.1
Pathological stage		
I	183	16.7
II	621	56.6
III	249	22.7
IV	20	1.8
Unknown	24	2.2
T stage		
T1	281	25.6
T2	635	57.9
T3	138	12.6
T4	40	3.6
Unknown	3	0.3
N stage		
N0	516	47.0
N1	364	33.2
N2	120	10.9
N3	77	7.0
Unknown	20	1.8
M stage		
M0	912	83.1
M1	22	2.0
Unknown	163	14.9

(**Table 2**). The primer sequences are listed in **Supplementary Table S1**.

Prediction of m5C Sites on the Three lncRNAs

RNAm5Cfinder (http://www.rnanut.net/rnam5cfinder/); (Li J. et al., 2018), iRNAm5C-PseDNC (http://www.jci-bioinfo.cn/iRNAm5C-PseDNC); (Qiu et al., 2017), and iRNA-m5C (http://lin-group.cn/server/iRNA-m5C/service.html); (Lv et al., 2020) databases were used to further verify the m5C modification sites on lncRNAs.

Statistical Analysis

Statistical significance was set at $p < 0.05$, and Kruskal-Wallis or Wilcoxon test was used for comparing the differential expression of m5C-lncRNA between different groups. The KM and life table method was used to estimate survival, and log-rank test was employed to compare the survival curves across the different groups. Pearson's correlation coefficient was used to determine the correlation between lncRNAs and infiltration of tumor immune cells.

RESULTS

Identification of BC-specific m5C-lncRNAs

We used differential expression and survival analyses to select BC-specific lncRNAs from 422 m5C-lncRNAs deposited in the TCGA dataset. A total of 334 differentially expressed lncRNAs between BC and normal breast tissues were screened ($p < 0.05$). Univariate Cox regression analysis was performed to analyze the prognostic factors of m5C-lncRNAs based on these lncRNAs ($p < 0.05$) (**Figure 2A**).

Construction of the m5C-lncRNA-Related Risk Model

The prognostic lncRNAs identified following univariate Cox regression analysis were included in the multivariate analysis. Three independent prognostic lncRNAs (AP005131.2, AL121832.2, and LINC01152) (**Supplementary Table S2**) were screened to build a three-gene risk model. Coefficients of m5C-lncRNAs from the multivariate Cox regression were used as coefficients of corresponding factors in the risk model (**Supplementary Table S3**). As shown in **Figure 2B**, the co-expression network that combined the function of m5C, m5C regulators, selected m5C-lncRNAs, and their prognostic roles indicated NSUN5, TET2, and TRDMT1 modified three lncRNAs to be protective factors.

According to the risk score, patients with BC were divided into high- and low-risk groups. **Figure 2C** shows the relationship between risk score and the corresponding survival status, suggesting that higher the risk scores, higher the number of deaths.

Correlation Between Risk Model, Three m5C-lncRNAs, and Clinical Variables

We further analyzed the relationship between the risk model, m5C-lncRNAs, and different clinicopathological features. The age

TABLE 2 | Clinical information of 31 patients with breast cancer in the qPCR cohort.

Variables	No. of patients	Percentage (%)
Age (years)		
≤55	4	12.9
>55	27	87.1
Gender		
Female	31	100.0
Male	0	0.0
Pathological stage		
I	6	19.4
II	15	48.4
III	10	32.3
IV	0	0.0
T stage		
T1	11	35.5
T2	18	58.1
T3	2	6.5
T4	0	0.0
N stage		
N0	13	41.9
N1	8	25.8
N2	4	12.9
N3	6	19.4
M stage		
M0	31	100.0
M1	0	0.0

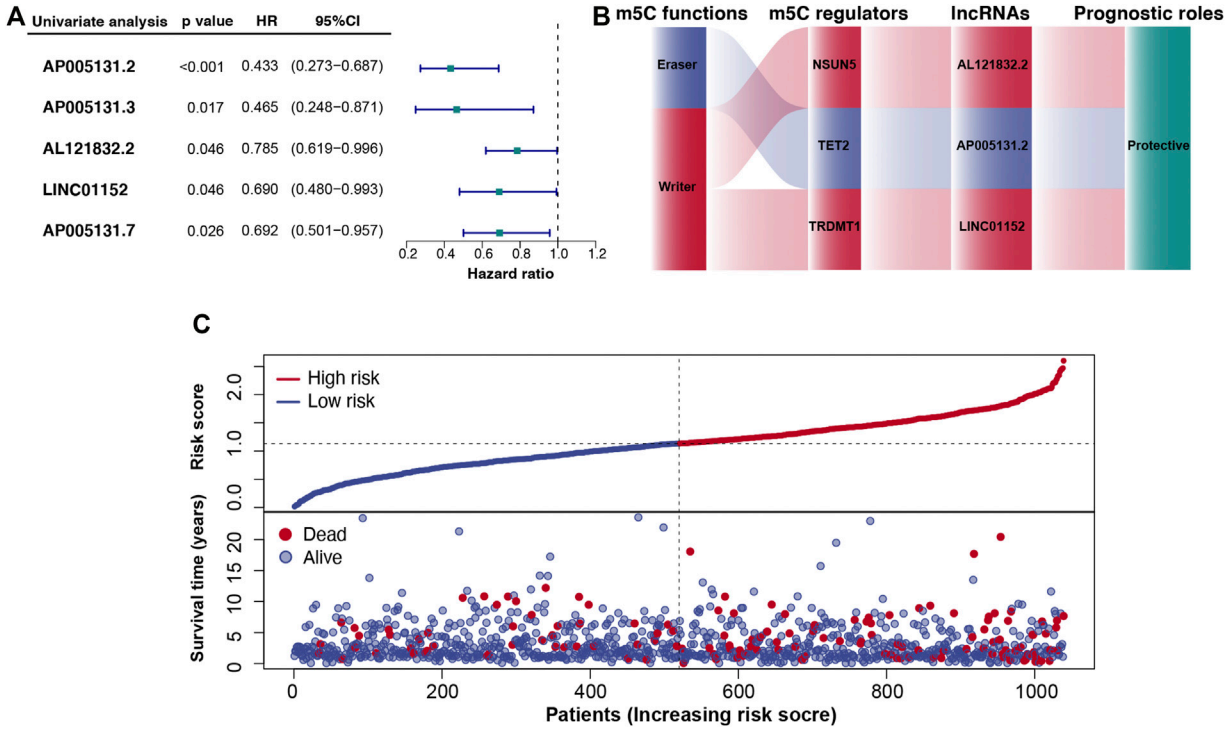

FIGURE 2 | Cancer-specific m5C-lncRNAs and their co-expression networks with m5C regulators in breast cancer (BC). **(A)** Forest diagram of univariate Cox regression analysis of m5C-lncRNAs. **(B)** Co-expression networks of the three m5C-lncRNAs screened using multivariate Cox regression analysis. **(C)** Distributions of risk scores and survival status of patients with BC based on m5C-lncRNA signature from the TCGA dataset.

($p < 0.01$) and survival status ($p < 0.01$) were significantly different between the high- and low-risk groups. Moreover, the three prognostic m5C-lncRNAs were differentially expressed in the high- and low-risk subgroups; high expression of AP005131.2, AL121832.2, and LINC01152 was associated with low-risk scores. However, there was no significant difference between the pathological stage and TNM stage (**Figure 3A**). The risk model prognostic value was assessed using survival analysis. Results indicated that the high-risk group was associated with worse outcomes compared to those observed in the low-risk group ($p < 0.05$) (**Figure 3B**). KM survival analysis showed a close relationship of the three lncRNAs with the overall survival (OS) of BC patients (**Figure 4A**). To further explore the value of risk model in BC, subgroup analysis of the three prognostic lncRNAs was performed based on molecular subtypes, histology stage, T stage, and N stage. As shown in **Figure 4B**, all lncRNAs showed a significant correlation with BC molecular subtypes ($p < 0.001$). In terms of histological stage, the expression of AP005131.2 in the early stage was significantly higher than that in the later stage ($p < 0.05$). For T stage, AP005131.2 and LINC01152 were differentially expressed in each group ($p < 0.05$). However, expression of AP005131.2, AL121832.2, and LINC01152 was not significantly different across groups in the N stage.

Validation of the m5C-lncRNA Risk Model

Univariate and multivariate Cox regression analyses were used to analyze the independent prognostic value of the risk model and clinical variables. The univariate and multivariate analyses revealed that age (continuous variable; HR = 1.033 and 1.031, respectively, 95% CI: 1.019–1.048 and 1.016–1.046, respectively; $p < 0.001$) and risk score (continuous variable; HR = 2.192 and 2.007, respectively, 95% CI: 1.526–3.149 and 1.372–2.936, respectively; $p < 0.001$) can be independent of gender, pathological stage, and TNM stage while being important prognostic factors (**Figure 5A**) Time-dependent ROC curves were used to evaluate the prognostic and predictive ability of age and risk score. Results revealed that, compared to other predictors, age and risk score had higher precision (AUC = 0.775, 0.741, respectively) (**Figure 5B**). PCA was used to show the different distribution patterns of m5C between the two risk groups across all genes, m5C-lncRNAs, as well as the risk gene expression profiles and found that patients in the risk gene group were significantly spread in different directions, showing the sensitivity, specificity, and positive and negative predictive values of the risk model (**Figures 5C–E**). The above results indicated that the m5C-lncRNA risk model is a significant independent prognostic indicator, applicable for clinical prognosis evaluation in patients with BC.

Construction of m5C-lncRNA Nomogram Based on the Risk Model

To establish a clinically available quantitative tool for predicting patient prognosis, we constructed a nomogram using the risk score and age (**Figure 6A**). Calibration plots showed that the

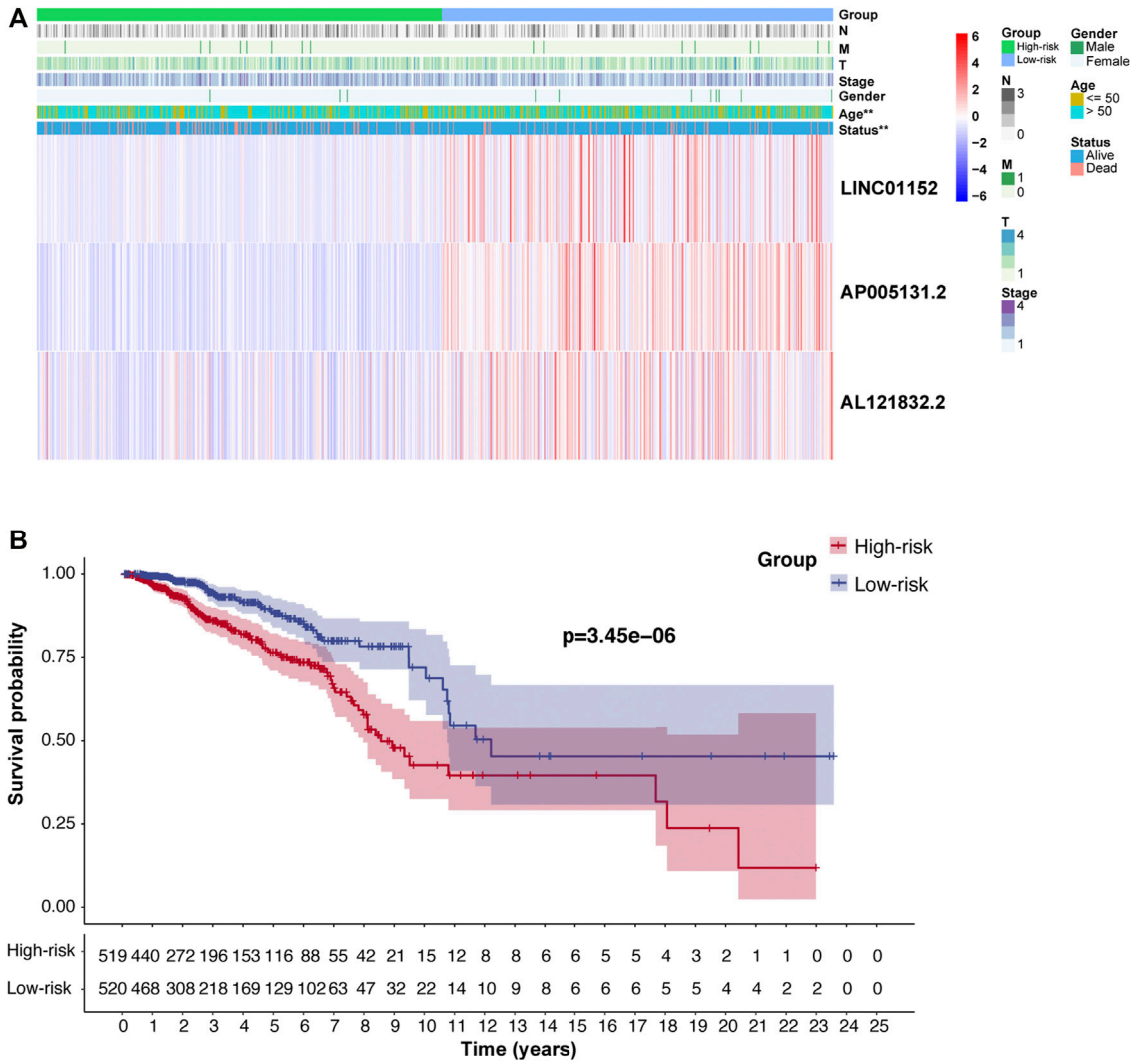

FIGURE 3 | The clinicopathological features and prognosis of patients in the high- and low-risk groups. **(A)** Heatmap showing the clinicopathological characteristics and expression levels of the three m5C-lncRNAs in the high- and low-risk groups. **(B)** Prognostic value of the model based on Kaplan–Meier survival analysis. (**$p < 0.01$).

observed *vs.* predicted rates of 3-, 5-, and 10-years overall survival (OS) were in perfect concordance (**Figure 6B**).

Verification of *in vitro* Expression Level and Prediction of m5C Modification Sites on Three m5C-lncRNAs

We identified AP005131.2, AL121832.2, and LINC01152 as the lncRNAs with the highest prognostic value. We further verified the expression of the m5C-lncRNAs in BC and paracancerous tissues. The expression levels of AP005131.2 and AL121832.2 were significantly lower in tumor tissues than in paracancerous tissues ($p < 0.05$). While there was no statistically significant difference in the expression of LINC01152, we observed a trend towards lower expression levels in tumor tissues (**Supplementary Figure S1**). The results were consistent with the TCGA cohort. We further investigated the m5C modification sites on three m5C-lncRNAs using three online databases, which are based on different algorithms (RNAm5Cfinder, iRNAm5C-PseDNC, and iRNAm5C) (**Supplementary Table S4**); the result confirmed that the three m5C-lncRNAs were lncRNA modified by m5C.

GSEA and the Correlation of TIICs

The alteration of risk model associated multiple cancer signaling pathways were identified through GSEA. The analysis showed that the high-risk group was mainly enriched in the following terms: citrate cycle (TCA cycle), glycan biosynthesis, sphingolipid metabolism, and steroid biosynthesis (**Figure 7A**). Based on the important role of TIICs in tumor occurrence and development, we analyzed the heterogeneity of TIICs in BC based on the risk score. First, we analyzed the correlation between 21 types of immune cells. Activated memory CD4 T cells had a strong

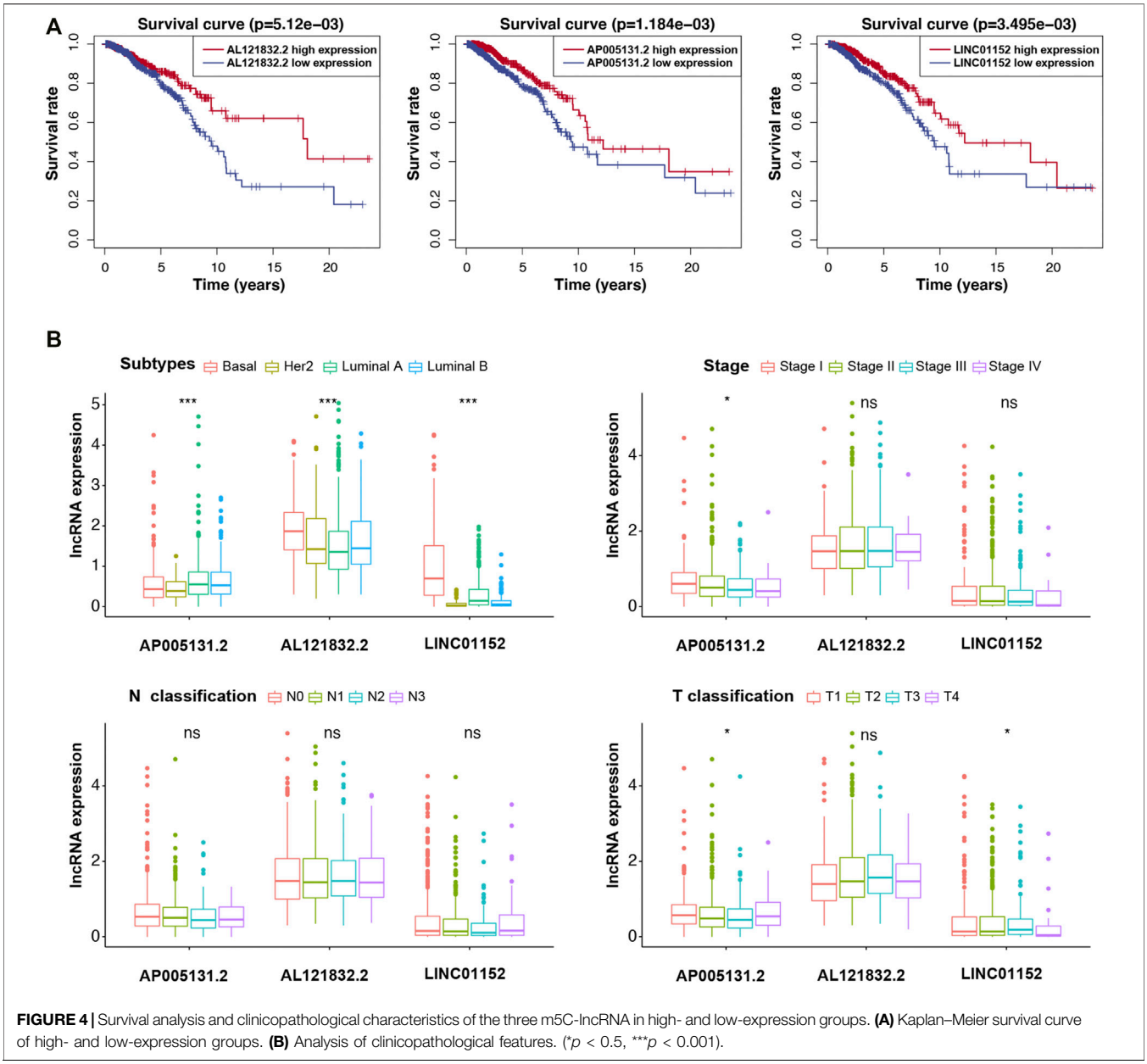

FIGURE 4 | Survival analysis and clinicopathological characteristics of the three m5C-lncRNA in high- and low-expression groups. **(A)** Kaplan–Meier survival curve of high- and low-expression groups. **(B)** Analysis of clinicopathological features. (*p < 0.5, ***p < 0.001).

positive correlation with CD8 T cells (r = 0.45), followed by the plasma cells and naive B cells (r = 0.37). Conversely, resting CD4 memory T cells had a strong negative correlation with macrophages M0 (r = −0.51) in BC tissues (**Figure 7B**). Next, we analyzed the expression levels of the 21 types of immune cells between high- and low-risk groups. The proportion of naive B cells and regulatory T cells in the low-risk group was higher than that in the high-risk group (p < 0.01). Additionally, the proportion of gamma delta T cells, macrophages M2, resting dendritic cells, resting mast cells, and neutrophils in the high-risk group was higher than that in the low-risk group (p < 0.05) (**Figure 7C**). These findings indicated that the risk model can distinguish the characteristics of TIICs in BC tissues.

DISCUSSION

BC is one of the most frequent malignancies with the highest mortality rate worldwide (DeSantis et al., 2019). Although a variety of therapeutic options (surgery, chemotherapy, radiotherapy, and immunotherapy, etc.) have made significant progress in recent years (Zhou et al., 2012; Pan et al., 2021), the prognosis of BC requires further improvement. Precise diagnosis and treatment are essential for improving the survival of patients with BC. Thus, comprehensive research would be required to explore the effective therapeutic targets and tailor precise treatment plans for each patient. Currently, more than 150 RNA post-transcriptional modification methods are in use, of

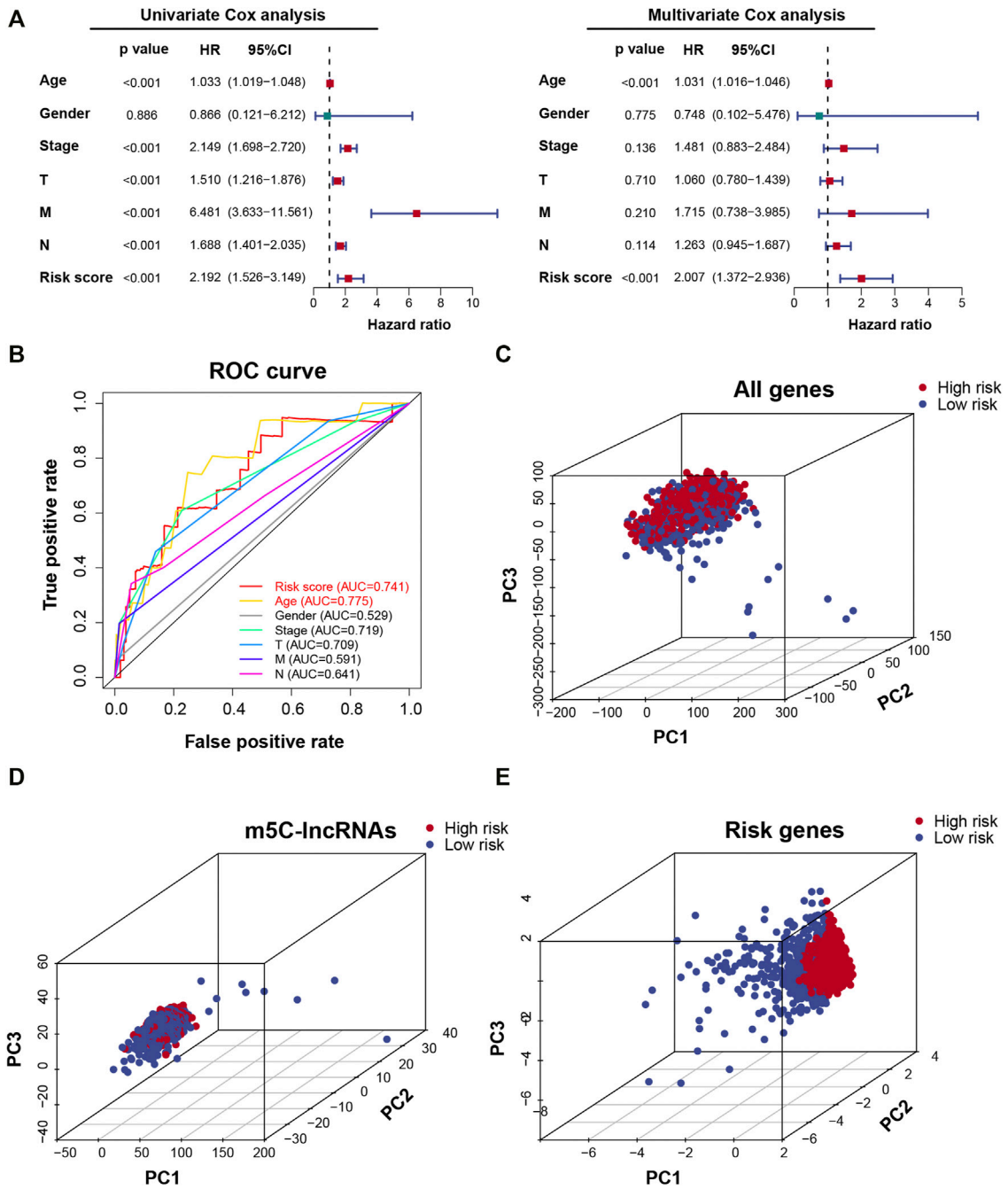

FIGURE 5 | Prognostic risk model verification. **(A)** Univariate and multivariate Cox regression analyses of risk scores combined with clinicopathological characteristics. **(B)** The receiver operating characteristic (ROC) curve verifies the sensitivity and reliability of the risk model. **(C–E)** Principal component analysis between low- and high-risk groups based on the signature of all genes, m5C-lncRNAs, and risk genes.

which m6A, m5C, and m1A are the most common modifications (Zhao et al., 2017) involved in cell proliferation regulation and disease progression. Previous studies have found that RNA post-transcriptional modification play an important role in lncRNA (Liu H. et al., 2020; Yi et al., 2020); however, the m5C modification of lncRNA is still poorly understood. In the present study, we screened m5C-lncRNAs for prognostic value as well as abnormal expression, and using multivariate Cox and risk scoring methods, we constructed a prognostic prediction model and combined the same with clinicopathological features and tumor immune cell infiltration to explore the role of m5C-lncRNAs in BC.

LncRNAs are transcripts of more than 200 nucleotides (Ma et al., 2016) and play an important role in the multi-stage

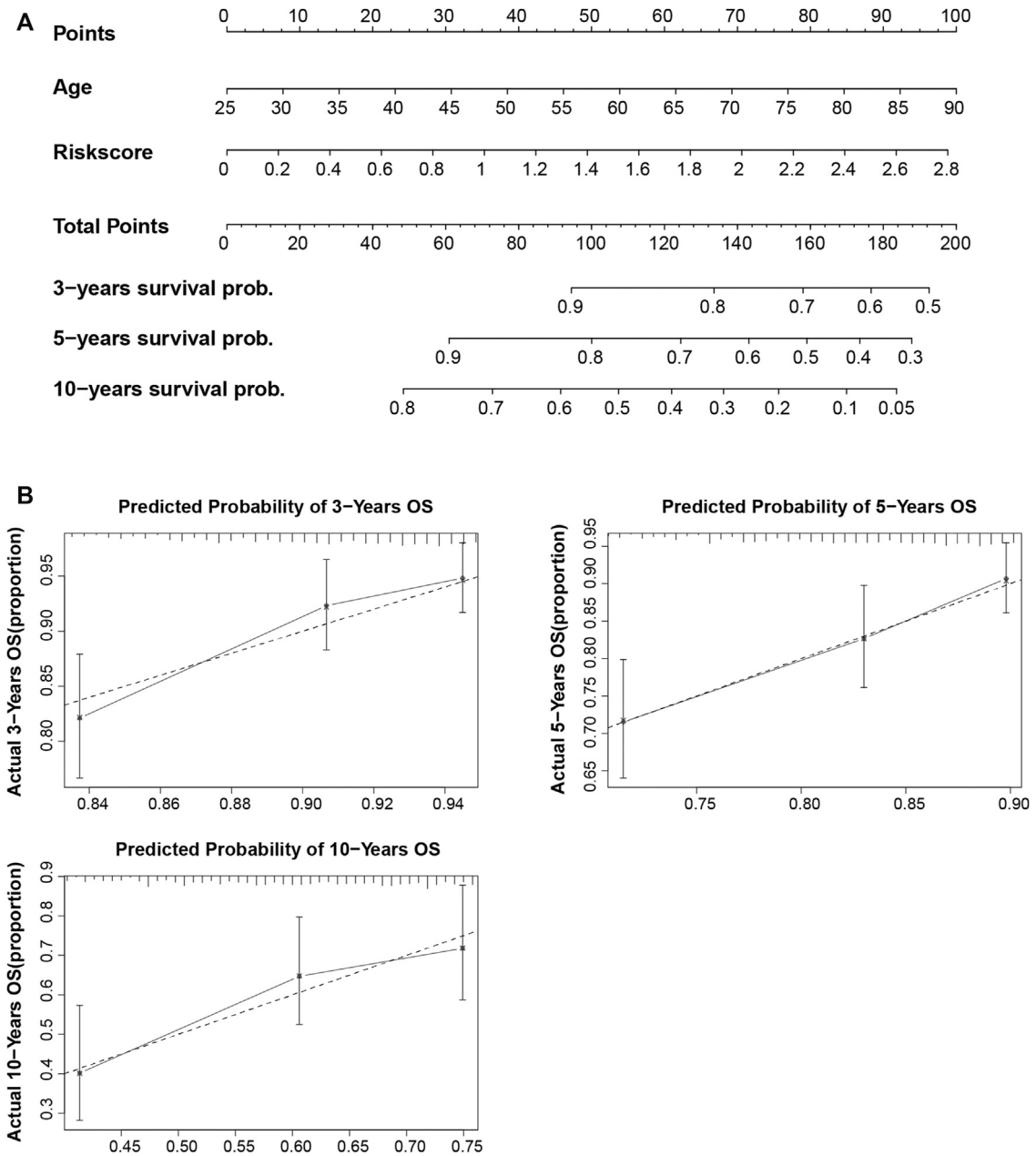

FIGURE 6 | Construction and calibration of the nomogram. **(A)** Construction of a nomogram based on age and risk score as independent prognostic factors. **(B)** The calibration curve revealed that the nomogram was well calibrated; the 3-, 5-, and 10-years overall survival showed an optimal agreement between the actual observation and nomogram prediction.

(epigenetics, transcription level, post-transcriptional level) occurrence and development of various malignancies (Chalei et al., 2014; Gong et al., 2015; Qin et al., 2016; Goyal et al., 2017) including mediating the interaction between DNA and protein, adsorbing miRNA, and regulating the expression of target proteins. Increasing evidence indicates that the expression level of lncRNA changes under pathological conditions, thus affecting cancer progression. For example, m6A methyltransferase-like 3 has been reported to stabilize the expression of LINC00958, which acts as a competitive endogenous RNA of *miR-378a-3p* to upregulate the expression of YY1; thereby facilitating BC progression (Rong et al., 2021). Studies have revealed that there are m6A modification sites on lncRNA NEAT1, which is related to bone metastasis in prostate cancer. The lncRNA NEAT1 acts as a bridge between CYCLINL1 and CDK19 to promote Pol II ser2 phosphorylation, which might represent a new target for the treatment and diagnosis of bone cancer metastasis (Wen S. et al., 2020). However, currently there are only few studies reporting the

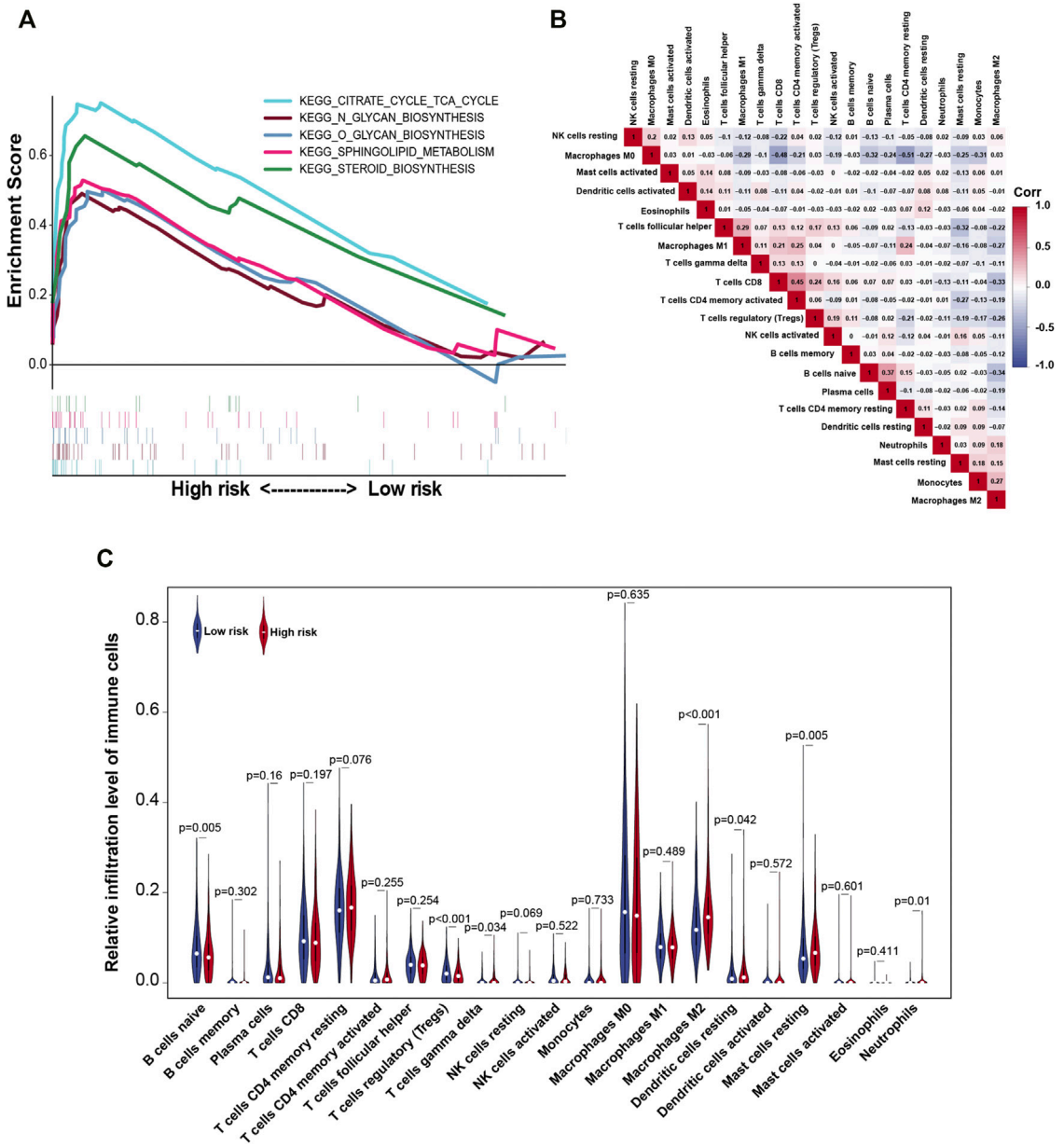

FIGURE 7 | Gene set enrichment analysis (GSEA) and the correlation between risk model and tumor-infiltrating immune cells. **(A)** Enrichment analysis of gene signaling pathways in risk models. **(B)** Correlation between the expressions of tumor-infiltrating immune cells in BC. **(C)** Violin plot of relative infiltration level of tumor immune cells in the high- and low-risk groups. Corr, Spearman correlation coefficient.

effects of m5C-lncRNAs. In hepatocellular carcinoma (HCC), m5C-related RNA methyltransferase *NSUN2* targets lncRNA *H19* and stabilizes its expression. LncRNA *H19* further binds to the known oncoprotein G3BP1, leading to downstream MYC accumulation, subsequently promoting the proliferation of HCC cells (Sun et al., 2020). In esophageal squamous cell carcinoma (ESCC), lncRNA *NMR* methylated by *NSUN2* combines with the chromatin regulator BPTF to promote the expression of MMP3 and MMP10 *via* the ERK1/2 pathway; thus, promoting the progression of ESCC (Li Y. et al., 2018). However, the regulatory role of m5C on lncRNAs in BC has not been reported yet.

In this study, we used BC-specific m5C-lncRNAs AP005131.2, AL121832.2, and LINC01152 to construct a risk model. Univariate Cox, multivariate Cox, and ROC curve analyses confirmed that the risk model can be independent of the traditional TNM stages and other clinical features and has good prognostic and predictive values. This is the first model based on m5C-related lncRNAs that is predictive of BC prognosis, offering new directions for research and clinical management of this common cancer. Besides, compared with other similar RNA methylation-related risk models, this model also has a high degree of prediction sensitivity and specificity

(Zhang et al., 2020; Wang et al., 2021; Zhang et al., 2021). To apply the prognostic model in clinical practice, we built a nomogram based on age, stage, and risk score, which can easily predict the prognosis of patients over 1, 5, and 10 years. Furthermore, PCA showed that risk gene can better illustrate the m5C characteristics. The results presented above indicated that AP005131.2, AL121832.2, and LINC01152 can be used as potential biomarkers, and a reliable risk model and risk-model-based nomograms are critical for providing the necessary evidence for clinical adoption and for driving continual improvement in patients with BC.

Currently, only LINC01152 has been reported among the three m5C-lncRNAs (Chen et al., 2019, 23; Wu et al., 2021). In glioblastoma multiforme, LINC01152 can upregulate the expression of MAML2 *via* the Notch signaling pathway to promote glioblastoma multiforme tumorigenesis. It can bind to the promoter region of IL-23, promote its transcriptional activity, and upregulate the levels of Stat3 and p-Stat3, resulting in the subsequent progression of HCC. Combined with bioinformatics analysis and tissue level verification, AP005131.2, AL121832.2, and LINC01152 generally showed low expression, which was related to better prognostic value in BC. BC is a highly heterogeneous tumor based on gene expression profiles, and can be divided into three subtypes: luminal (ER-and/or PR-positive), HER2 enriched (HER2 positive), and basal subtypes (ER, PR, and HER2 negative) (Yeo and Guan, 2017). Different subtypes have different pathological processes. Most of the autophagy-related lncRNAs had been previously reported to be significantly related to molecular subtypes of BC, indicating that autophagy-related lncRNAs may participate in the regulation of ER, PR, and HER2 status (Li X. et al., 2021). In this study, the expression of three m5C-lncRNAs was significantly correlated with different molecular subtypes, indicating that m5C-lncRNAs might participate in the regulation of ER, PR, and HER2 status.

We explored the biological processes that m5C-lncRNAs may participate in. GSEA demonstrated that in patients with BC and high-risk scores, cell signaling pathways were mainly enriched in the citrate cycle (TCA cycle), glycan biosynthesis, sphingolipid metabolism, and steroid biosynthesis. We found that the risk models were mainly enriched in metabolic-related pathways. Studies have shown that cholesterol-derived metabolites play a critical functional role in supporting cancer progression and suppressing immune responses (Huang et al., 2020). In BC, sphingosine kinase 1, which is necessary for the generation of S1P and its receptor S1PR1 can induce the release of proinflammatory cytokines, macrophage infiltration, and tumor progression (Nagahashi et al., 2018). The specific mechanism by which m5C-related lncRNAs participate in the regulation of metabolic pathways requires further *in vivo* and *in vitro* experimentation.

In recent years, research on the effects of lncRNAs in TIM has received widespread attention. Immune-related lncRNAs have important prognostic value in various cancers, such as lung adenocarcinoma, HCC, and low-grade glioma (Wen J. et al., 2020; Hong et al., 2020; Li et al., 2020). Recently, a prognostic model has been built based on four immune-related lncRNAs that can better predict the prognosis of patients with BC (Shen et al., 2020). Here, we provided an in-depth analysis of the relationship between the risk model and TIM cells. Our findings highlighted that the high- and low-risk groups significantly distinguished the characteristics of naive B cells, regulatory T cells, gamma delta T cells, macrophages M2, resting dendritic cells, resting mast cells, and neutrophils, among which naive B cells and regulatory T cells exhibited a higher degree of infiltration in the low-risk group than in the high-risk group. Alternatively, gamma delta T cells, macrophages M2, resting dendritic cells, resting mast cells, and neutrophils exhibited a higher degree of infiltration in the high-risk group than in the low-risk group. Altogether, the risk model could evaluate the tumor infiltration of immune cells to analyze the tumor immune characteristics, thereby determining the prognosis of patients with BC.

However, there are some limitations associated with our research. First, we screened m5C-related lncRNAs using the TCGA database and constructed a risk model based on the same database. Secondly, owing to the limitations of an imperfect annotation of lncRNAs and a dearth of complete lncRNA-seq data in cancer, the risk model could not be further verified. In addition, we examined the expression levels of the three m5C-lncRNAs at histological level but did not perform further *in vivo* and *in vitro* analyses. We did not perform further experiments, such as MeRIP-seq and RNA-BisSeq, to verify the m5C modification sites on lncRNAs. Further studies are warranted to validate our findings in future.

In conclusion, the risk model constructed based on the three m5C-lncRNAs (AP005131.2, AL121832.2, and LINC01152) has independent prognostic value and extremely high reliability in BC; thus, providing clues for in-depth research on m5C modification sites in lncRNAs. Moreover, the risk model for BC was translated into a nomogram, providing a convenient and quantitative prognostic prediction tool for clinicians, new insights for understanding immune cell-specific genes in tumor regulation, and possibly improving the ability to individualize treatment for patients with BC.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Medical Ethics Committee of Quanzhou First Hospital of Fujian Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DC designed the study. ZH analyzed the data and performed the RT-qPCR. JL wrote the manuscript and generated the figures and tables. JC checked and polished the language. JL reviewed the

manuscript. All authors contributed to this article and approved the submitted version.

FUNDING

This research was funded by the Natural Science Foundation of Fujian Province, PRC (No. 2019J01599), Science and Technology Innovation Joint Fund Project of Fujian Province (No. 2019Y9049), Science and Technology Project of Quanzhou (2020C047R) and Startup Fund for Scientific Research of Fujian Medical University (No.2018QH1185).

ACKNOWLEDGMENTS

The authors appreciate TCGA database for providing the original study data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.748279/full#supplementary-material

REFERENCES

Biswas, S., and Rao, C. M. (2018). Epigenetic Tools (The Writers, the Readers and the Erasers) and Their Implications in Cancer Therapy. *Eur. J. Pharmacol.* 837, 8–24. doi:10.1016/j.ejphar.2018.08.021

Cai, P., Otten, A. B. C., Cheng, B., Ishii, M. A., Zhang, W., Huang, B., et al. (2020). A Genome-Wide Long Noncoding RNA CRISPRi Screen Identifies PRANCR as a Novel Regulator of Epidermal Homeostasis. *Genome Res.* 30, 22–34. doi:10.1101/gr.251561.119

Chalei, V., Sansom, S. N., Kong, L., Lee, S., Montiel, J. F., Vance, K. W., et al. (2014). The Long Non-Coding RNA Dali Is an Epigenetic Regulator of Neural Differentiation. *Elife.* 3, e04530. doi:10.7554/eLife.04530

Chen, T., Pei, J., Wang, J., Luo, R., Liu, L., Wang, L., et al. (2019). HBx-Related Long Non-Coding RNA 01152 Promotes Cell Proliferation and Survival by IL-23 in Hepatocellular Carcinoma. *Biomed. Pharmacother.* 115, 108877. doi:10.1016/j.biopha.2019.108877

DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., et al. (2019). Breast Cancer Statistics, 2019. *CA A. Cancer J. Clin.* 69, 438–451. doi:10.3322/caac.21583

Feng, Z.-Y., Gao, H.-Y., and Feng, T.-D. (2021). Immune Infiltrates of m6A RNA Methylation-Related lncRNAs and Identification of PD-L1 in Patients With Primary Head and Neck Squamous Cell Carcinoma. *Front. Cel Dev. Biol.* 9, 672248. doi:10.3389/fcell.2021.672248

Gong, C., Li, Z., Ramanujan, K., Clay, I., Zhang, Y., Lemire-Brachat, S., et al. (2015). A Long Non-Coding RNA, LncMyoD, Regulates Skeletal Muscle Differentiation by Blocking IMP2-Mediated mRNA Translation. *Dev. Cel.* 34, 181–191. doi:10.1016/j.devcel.2015.05.009

Goyal, A., Myacheva, K., Groß, M., Klingenberg, M., Duran Arqué, B., and Diederichs, S. (2017). Challenges of CRISPR/Cas9 Applications for Long Non-Coding RNA Genes. *Nucleic Acids Res.* 45, gkw883. doi:10.1093/nar/gkw883

Hasan, M. M., Basith, S., Khatun, M. S., Lee, G., Manavalan, B., and Kurata, H. (2021a). Meta-i6mA: an Interspecies Predictor for Identifying DNA N6-Methyladenine Sites of Plant Genomes by Exploiting Informative Features in an Integrative Machine-Learning Framework. *Brief Bioinform.* 22, bbaa202. doi:10.1093/bib/bbaa202

Hasan, M. M., Shoombuatong, W., Kurata, H., and Manavalan, B. (2021b). Critical Evaluation of Web-Based DNA N6-Methyladenine Site Prediction Tools. *Brief. Funct. Genomics.* 20, 258–272. doi:10.1093/bfgp/elaa028

Hong, W., Liang, L., Gu, Y., Qi, Z., Qiu, H., Yang, X., et al. (2020). Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. *Mol. Ther. - Nucleic Acids.* 22, 937–947. doi:10.1016/j.omtn.2020.10.002

Huang, B., Song, B.-L., and Xu, C. (2020). Cholesterol Metabolism in Cancer: Mechanisms and Therapeutic Opportunities. *Nat. Metab.* 2, 132–141. doi:10.1038/s42255-020-0174-0

Jonkhout, N., Tran, J., Smith, M. A., Schonrock, N., Mattick, J. S., and Novoa, E. M. (2017). The RNA Modification Landscape in Human Disease. *RNA.* 23, 1754–1769. doi:10.1261/rna.063503.117

Lai, J., Chen, B., Zhang, G., Li, X., Mok, H., and Liao, N. (2020). Molecular Characterization of Breast Cancer: a Potential Novel Immune-Related lncRNAs Signature. *J. Transl Med.* 18, 416. doi:10.1186/s12967-020-02578-4

Li, J.-P., Li, R., Liu, X., Huo, C., Liu, T.-T., Yao, J., et al. (2020). A Seven Immune-Related lncRNAs Model to Increase the Predicted Value of Lung Adenocarcinoma. *Front. Oncol.* 10, 560779. doi:10.3389/fonc.2020.560779

Li, J., Huang, Y., Yang, X., Zhou, Y., and Zhou, Y. (2018a). RNAm5Cfinder: A Web-Server for Predicting RNA 5-Methylcytosine (m5C) Sites Based on Random Forest. *Sci. Rep.* 8, 17299. doi:10.1038/s41598-018-35502-4

Li, Y., Li, J., Luo, M., Zhou, C., Shi, X., Yang, W., et al. (2018b). Novel Long Noncoding RNA NMR Promotes Tumor Progression *via* NSUN2 and BPTF in Esophageal Squamous Cell Carcinoma. *Cancer Lett.* 430, 57–66. doi:10.1016/j.canlet.2018.05.013

Li, S., Wu, J., Huang, O., He, J., Zhu, L., Chen, W., et al. (2021a). Association of Molecular Biomarkers Heterogeneity and Treatment Pattern, Disease Outcomes in Multifocal or Multicentric Breast Cancer Patients. *Cancer Res.* 81, 18–22. doi:10.1158/0008-5472.CAN-09-3903

Li, X., Jin, F., and Li, Y. (2021b). A Novel Autophagy-Related lncRNA Prognostic Risk Model for Breast Cancer. *J. Cel. Mol. Med.* 25, 4–14. doi:10.1111/jcmm.15980

Liu, H., Xu, Y., Yao, B., Sui, T., Lai, L., and Li, Z. (2020a). A Novel N6-Methyladenosine (m6A)-Dependent Fate Decision for the lncRNA THOR. *Cell Death Dis.* 11, 613. doi:10.1038/s41419-020-02833-y

Liu, L., Lei, X., Fang, Z., Tang, Y., Meng, J., and Wei, Z. (2020b). LITHOPHONE: Improving lncRNA Methylation Site Prediction Using an Ensemble Predictor. *Front. Genet.* 11, 545. doi:10.3389/fgene.2020.00545

Liu, Q., Chen, J., Wang, Y., Li, S., Jia, C., Song, J., et al. (2021). DeepTorrent: a Deep Learning-Based Approach for Predicting DNA N4-Methylcytosine Sites. *Brief Bioinform.* 22, bbaa124. doi:10.1093/bib/bbaa124

Lu, Y., Tong, Y., Chen, X., and Shen, K. (2021). Association of Biomarker Discrepancy and Treatment Decision, Disease Outcome in Recurrent/Metastatic Breast Cancer Patients. *Front. Oncol.* 11, 638619. doi:10.3389/fonc.2021.638619

Lv, H., Zhang, Z.-M., Li, S.-H., Tan, J.-X., Chen, W., and Lin, H. (2020). Evaluation of Different Computational Methods on 5-Methylcytosine Sites Identification. *Brief Bioinform.* 21, 982–995. doi:10.1093/bib/bbz048

Ma, Y., Yang, Y., Wang, F., Moyer, M.-P., Wei, Q., Zhang, P., et al. (2016). Long Non-Coding RNA CCAL Regulates Colorectal Cancer Progression by Activating Wnt/β-Catenin Signalling Pathway *via* Suppression of Activator Protein 2α. *Gut.* 65, 1494–1504. doi:10.1136/gutjnl-2014-308392

Nagahashi, M., Yamada, A., Katsuta, E., Aoyagi, T., Huang, W.-C., Terracina, K. P., et al. (2018). Targeting the SphK1/S1P/S1PR1 Axis that Links Obesity, Chronic Inflammation, and Breast Cancer Metastasis. *Cancer Res.* 78, 1713–1725. doi:10.1158/0008-5472.CAN-17-1423

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods.* 12, 453–457. doi:10.1038/nmeth.3337

Pan, H., Qian, M., Chen, H., Wang, H., Yu, M., Zhang, K., et al. (2021). Precision Breast-Conserving Surgery With Microwave Ablation Guidance: A Pilot Single-Center, Prospective Cohort Study. *Front. Oncol.* 11, 680091. doi:10.3389/fonc.2021.680091

Qin, W., Li, X., Xie, L., Li, S., Liu, J., Jia, L., et al. (2016). A Long Non-Coding RNA,APOA4-AS, regulates APOA4 Expression Depending on HuR in Mice. *Nucleic Acids Res.* 44, 6423–6433. doi:10.1093/nar/gkw341

Qiu, W.-R., Jiang, S.-Y., Xu, Z.-C., Xiao, X., and Chou, K.-C. (2017). iRNAm5C-PseDNC: Identifying RNA 5-Methylcytosine Sites by Incorporating Physical-

Chemical Properties Into Pseudo Dinucleotide Composition. *Oncotarget.* 8, 41178–41188. doi:10.18632/oncotarget.17104

Rong, D., Dong, Q., Qu, H., Deng, X., Gao, F., Li, Q., et al. (2021). m6A-Induced LINC00958 Promotes Breast Cancer Tumorigenesis *via* the miR-378a-3p/YY1 axis. *Cell Death Discov.* 7, 27. doi:10.1038/s41420-020-00382-z

Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell* 169, 1187–1200. doi:10.1016/j.cell.2017.05.045

Schwartz, R. S., and Erban, J. K. (2017). Timing of Metastasis in Breast Cancer. *N. Engl. J. Med.* 376, 2486–2488. doi:10.1056/NEJMcibr1701388

Shen, Y., Peng, X., and Shen, C. (2020). Identification and Validation of Immune-Related lncRNA Prognostic Signature for Breast Cancer. *Genomics.* 112, 2640–2646. doi:10.1016/j.ygeno.2020.02.015

Sun, N., Zhang, G., and Liu, Y. (2018). Long Non-Coding RNA XIST Sponges miR-34a to Promotes Colon Cancer Progression *via* Wnt/β-Catenin Signaling Pathway. *Gene.* 665, 141–148. doi:10.1016/j.gene.2018.04.014

Sun, Z., Xue, S., Zhang, M., Xu, H., Hu, X., Chen, S., et al. (2020). Aberrant NSUN2-Mediated m5C Modification of H19 lncRNA Is Associated With Poor Differentiation of Hepatocellular Carcinoma. *Oncogene.* 39, 6906–6919. doi:10.1038/s41388-020-01475-w

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., et al. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA A. Cancer J. Clin.* 71, 209–249. doi:10.3322/caac.21660

Tak, Y. G., and Farnham, P. J. (2015). Making Sense of GWAS: Using Epigenomics and Genome Engineering to Understand the Functional Relevance of SNPs in Non-Coding Regions of the Human Genome. *Epigenetics & Chromatin.* 8, 57. doi:10.1186/s13072-015-0050-4

Takayama, K.-I., Horie-Inoue, K., Katayama, S., Suzuki, T., Tsutsumi, S., Ikeda, K., et al. (2013). Androgen-Responsive Long Noncoding RNA CTBP1-AS Promotes Prostate Cancer. *EMBO J.* 32, 1665–1680. doi:10.1038/emboj.2013.99

Tang, Y., Chen, K., Song, B., Ma, J., Wu, X., Xu, Q., et al. (2021). m6A-Atlas: a Comprehensive Knowledgebase for Unraveling the N6-Methyladenosine (m6A) Epitranscriptome. *Nucleic Acids Res.* 49, D134–D143. doi:10.1093/nar/gkaa692

Tu, Z., Wu, L., Wang, P., Hu, Q., Tao, C., Li, K., et al. (2020). N6-Methylandenosine-Related lncRNAs Are Potential Biomarkers for Predicting the Overall Survival of Lower-Grade Glioma Patients. *Front. Cel Dev. Biol.* 8, 642. doi:10.3389/fcell.2020.00642

Wang, S., Zou, X., Chen, Y., Cho, W. C., and Zhou, X. (2021). Effect of N6-Methyladenosine Regulators on Progression and Prognosis of Triple-Negative Breast Cancer. *Front. Genet.* 11, 580036. doi:10.3389/fgene.2020.580036

Wen, J., Wang, Y., Luo, L., Peng, L., Chen, C., Guo, J., et al. (2020a). Identification and Verification on Prognostic Index of Lower-Grade Glioma Immune-Related LncRNAs. *Front. Oncol.* 10, 578809. doi:10.3389/fonc.2020.578809

Wen, S., Wei, Y., Zen, C., Xiong, W., Niu, Y., and Zhao, Y. (2020b). Long Non-Coding RNA NEAT1 Promotes Bone Metastasis of Prostate Cancer Through N6-Methyladenosine. *Mol. Cancer.* 19, 171. doi:10.1186/s12943-020-01293-4

Wu, J., Wang, N., Yang, Y., Jiang, G., Zhan, H., Li, F., et al. (2021). LINC01152 Upregulates MAML2 Expression to Modulate the Progression of Glioblastoma Multiforme *via* Notch Signaling Pathway. *Cel Death Dis.* 12, 115. doi:10.1038/s41419-020-03163-9

Wu, X., Tudoran, O. M., Calin, G. A., and Ivan, M. (2018). The Many Faces of Long Noncoding RNAs in Cancer. *Antioxid. Redox Signaling.* 29, 922–935. doi:10.1089/ars.2017.7293

Xue, M., Shi, Q., Zheng, L., Li, Q., Yang, L., and Zhang, Y. (2020). Gene Signatures of m5C Regulators May Predict Prognoses of Patients With Head and Neck Squamous Cell Carcinoma. *Am. J. Transl Res.* 12, 6841–6852.

Yeo, S. K., and Guan, J.-L. (2017). Breast Cancer: Multiple Subtypes Within a Tumor? *Trends Cancer.* 3, 753–760. doi:10.1016/j.trecan.2017.09.001

Yi, Y.-C., Chen, X.-Y., Zhang, J., and Zhu, J.-S. (2020). Novel Insights Into the Interplay Between m6A Modification and Noncoding RNAs in Cancer. *Mol. Cancer.* 19, 121. doi:10.1186/s12943-020-01233-2

Zhang, B., Gu, Y., and Jiang, G. (2020). Expression and Prognostic Characteristics of m6A RNA Methylation Regulators in Breast Cancer. *Front. Genet.* 11, 604597. doi:10.3389/fgene.2020.604597

Zhang, X., Shen, L., Cai, R., Yu, X., Yang, J., Wu, X., et al. (2021). Comprehensive Analysis of the Immune-Oncology Targets and Immune Infiltrates of N6-Methyladenosine-Related Long Noncoding RNA Regulators in Breast Cancer. *Front. Cel Dev. Biol.* 9, 1661. doi:10.3389/fcell.2021.686675

Zhao, B. S., Roundtree, I. A., and He, C. (2017). Post-Transcriptional Gene Regulation by mRNA Modifications. *Nat. Rev. Mol. Cel Biol.* 18, 31–42. doi:10.1038/nrm.2016.132

Zhou, W., Zha, X., Liu, X., Ding, Q., Chen, L., Ni, Y., et al. (2012). US-Guided Percutaneous Microwave Coagulation of Small Breast Cancers: a Clinical Study. *Radiology.* 263, 364–373. doi:10.1148/radiol.12111901

Zuo, X., Chen, Z., Gao, W., Zhang, Y., Wang, J., Wang, J., et al. (2020). M6A-Mediated Upregulation of LINC00958 Increases Lipogenesis and Acts as a Nanotherapeutic Target in Hepatocellular Carcinoma. *J. Hematol. Oncol.* 13, 5. doi:10.1186/s13045-019-0839-x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 11 November 2021
doi: 10.3389/fgene.2021.735454

Comprehensive Analysis of N^6-Methyladenosine (m^6A) Methylation in Neuromyelitis Optica Spectrum Disorders

Hong Yang[1†], Yi-Fan Wu[2†], Jie Ding[2], Wei Liu[1], De-Sheng Zhu[2], Xia-Feng Shen[1] and Yang-Tai Guan[2,3]**

[1]Department of Neurology, The First Rehabilitation Hospital of Shanghai, Tongji University School of Medicine, Shanghai, China, [2]Department of Neurology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [3]Department of Neurology, Tongji Hospital, Tongji University School of Medicine, Shanghai, China

OPEN ACCESS

Edited by:
Jia Meng,
Xi'an Jiaotong-Liverpool University, China

Reviewed by:
Omid Mirmosayyeb,
Isfahan University of Medical Sciences, Iran
Piyush Khandelia,
Birla Institute of Technology and Science, India

***Correspondence:**
Xia-Feng Shen
shenxiafeng@aliyun.com
Yang-Tai Guan
yangtaiguan@hotmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 02 July 2021
Accepted: 25 October 2021
Published: 11 November 2021

Citation:
Yang H, Wu Y-F, Ding J, Liu W, Zhu D-S, Shen X-F and Guan Y-T (2021) Comprehensive Analysis of N^6-Methyladenosine (m^6A) Methylation in Neuromyelitis Optica Spectrum Disorders. Front. Genet. 12:735454. doi: 10.3389/fgene.2021.735454

Background: N^6-Methyladenosine (m^6A) methylation is the most prevalent internal posttranscriptional modification on mammalian mRNA. But its role in neuromyelitis optica spectrum disorders (NMOSD) is not known.

Aims: To explore the mechanism of m^6A in NMOSD patients.

Methods: This study assessed the m^6A methylation levels in blood from two groups: NMOSD patients and healthy controls. Methylated RNA immunoprecipitation Sequencing (MeRIP-seq) and RNA-seq were performed to assess differences in m^6A methylation between NMOSD patients and healthy controls. Ultra-high performance liquid chromatography coupled with triple quadruple mass spectrometry (UPLC-QQQ-MS) method was performed to check m^6A level. Differential m^6A methylation genes were validated by MeRIP-qPCR.

Results: Compared with that in the control group, the total m6A level was decreased in the NMOSD group. Genes with upregulated methylation were primarily enriched in processes associated with RNA splicing, mRNA processing, and innate immune response, while genes with downregulated methylation were enriched in processes associated with the regulation of transcription, DNA-templating, and the positive regulation of I-kappa B kinase/NF-kappa B signalling.

Conclusion: These findings demonstrate that differential m^6A methylation may act on functional genes to regulate immune homeostasis in NMOSD.

Keywords: neuromyelitis optica spectrum disorder, immune homeostasis, N^6-methyladenosine (m^6A) methylation, differential methylation peaks, MeRIP-seq, UPLC-QQQ-MS

INTRODUCTION

N^6-Methyladenosine (m^6A) is one of the most abundant internal modifications of eukaryotic messenger RNA (mRNA) and plays an important role in gene expression regulatory processes, including the maintenance of stability, splicing and translation (Chen et al., 2019a). Approximately 25% of mRNAs are estimated to contain at least one m^6A nucleotide site, and

most of them occur at the sixth position of adenosine. m^{6}A was shown to be specifically catalysed at the consensus gene sequence DRACH (D = A/G/U, R = A/G, H = A/C/U). In addition to m^{6}A, other types of posttranscriptional modifications exist, including 5-methylcytosine (m^{5}C) and pseudouridine (ψ). To detect regions containing m^{6}A peaks, researchers have developed an immunoprecipitation-based method combined with high-throughput sequencing, termed methylated RNA immunoprecipitation sequencing(MeRIP-seq) (Dominissini et al., 2012). To determine m^{6}A level, researchers have developed the UPLC-QQQ-MS method, which has the advantages of short analysis time, good resolution and high sensitivity (Su et al., 2014).

m^{6}A modification is dynamic and reversible and regulated by m^{6}A methyltransferases (writers), demethylases (erasers) and reading proteins (readers). Writers include Methyltransferase Like 3 (METTL3), methyltransferase-like 14 (METTL4) and Wilms tumour 1-associated protein (WTAP). METTL3 catalyses the production of m6A; METTL4 forms a complex with METTL3, participates in interactions with target mRNA and recruits RNA, and WTAP stabilizes the complex (Liu et al., 2014; Ping et al., 2014; Lin et al., 2016). Erasers, including fat and obesity-associated protein (FTO) and alk B homologue 5 (ALKBH5), catalyse the demethylation of bases that have been modified by m^{6}A (Jia et al., 2011; Zheng et al., 2013). Readers, including YTH N6 methyladenosine RNA-binding protein 1–3 (YTHDF1-3) and insulin-like growth factor 2 mRNA-binding protein 1–3 (IGF2BP1-3), are also necessary for this process (Wang et al., 2014; Wang et al., 2015; Hanniford et al., 2020). These proteins recognize sites of mRNA methylation and affect RNA metabolism directly or indirectly. Overall, m6A-modifying enzymes play important roles in tumorigenesis and in the development of the human central nervous system (Lin et al., 2017; Zhang et al., 2018; Heck et al., 2020; Xue et al., 2021). In recent years, studies have revealed that m^{6}A participates in regulating the immune microenvironment (Chen et al., 2019b; Tang et al., 2020; Xu et al., 2020), but its involvement in NMOSD is unknown.

NMOSD is a kind of recurrent antibody-mediated neuroimmune disease, in which the optic nerve, spinal cord, and area postrema of the medulla are damaged, often causing severe symptoms (Wingerchuk et al., 2015). The specific diagnostic and pathogenic biomarker for NMOSD is aquaporin 4-immunoglobulin IgG (AQP4-IgG), which binds to the foot processes of astrocytes, activating complement and thereby recruiting neutrophils and eosinophils; this leads to astrocyte death, followed by axonal degeneration and oligodendrocyte injury (Lennon et al., 2004). The prevalence and incidence of NMOSD are 0.52–4.4 and 0.05–0.40 per 100,000, respectively (Pandit et al., 2015). Approximately 60% of NMOSD patients relapse within 1 year and approximately 90% within 5 years. Furthermore, approximately 50% of NMOSD patients require a wheelchair within 5 years of their diagnosis due to limb paralysis, and 62% become blind, affecting their quality of life and imposing a heavy burden on patients, families and society (Pittock and Lucchinetti, 2016). Currently, methylprednisolone and plasma exchange used in the acute stage as well as rituximab, mycophenolate mofetil and azathioprine used in the remission stage are only partially effective and do not substantially reduce the high disability associated with the disease or its recurrence. Therefore, it is necessary to further understand the pathogenesis of NMOSDs and explore more effective drugs.

Regardless, the m^{6}A status has not been assessed in NMOSD. To investigate differences in m^{6}A modification patterns between healthy controls (HCs) and NMOSD patients, we performed UPLC-QQQ-MS to assess the m^{6}A level and MeRIP-seq to identify the first known transcriptome-wide m^{6}A sites in NMOSD patients. In general, the total m6A level was decreased in patients compared to HCs. We also detected 3,371 differentially methylated peaks within mRNAs and 95 within long noncoding RNAs (lncRNCs, $p < 0.05$). Intriguingly, mRNAs harbouring differentially methylated peaks were shown to be involved in many important biological pathways associated with immune homeostasis.

MATERIALS AND METHODS

Blood Collection From NMOSD Patients and HCs

(1.1) Inclusion criteria: 1) aged 18–75 years old; 2) meeting the diagnostic criteria for NMOSD (Wingerchuk et al., 2015); 3) within 30 days of onset in the acute phase and not receiving relevant drug treatment; and 4) provision of signed informed consent.

(1.2) Exclusion criteria: 1) severe mental symptoms precluding cooperation with treatment; 2) no informed consent signed; 3) receiving immunosuppressive therapy within 30 days of new clinical symptoms; 4) other autoimmune diseases or tumours; 5) new clinical symptoms and not receiving relevant treatment but with clinical symptoms for more than 30 days before seeing a doctor.

(1.3) On the second day after admission, 10 ml of blood was collected into a BD Paxgene blood RNA tube; the samples were stored at −4°C for 24 h and then at −80°C for preparation of RNA.

(1.4) In total, six healthy individuals with no acute or chronic illness from the Health Care Centre of the First Rehabilitation Hospital of Shanghai were enrolled in the study as a control group.

RNA Preparation and Quality Control

Total RNA was extracted from whole blood stored in Paxgene blood RNA tubes by using a Paxgene Blood RNA Extraction Kit (PreAnalytiX, Qiagen). A Qubit fluorometer was used to quantify RNA at 260/280 nm, and an Agilent 2,100 Bioanalyzer was employed to determine the RNA integrity number (RIN) and 28S/18S values for quality control (RIN ≥ 6.0 and 28S/18S ≥ 0.7) and confirmed by electrophoresis on a denaturing agarose gel.

UPLC-QQQ-MS Method

Analysis of m6A level was performed on the UPLC-QQQ-MS system, consisting of a Waters Acquity UPLC (Waters

Corporation, Massachusetts, United States) and an AB SCIEX 5500 QQQ-MS instrument (AB Sciex LLC, Framingham, United States). As previously described (Su et al., 2014), we prepared a mixture containing all of the cofactors and enzymes needed for hydrolysis of the RNA into nucleosides. Then, 5 μl of the mixture was added to each sample of RNA, and the final volume was brought to 25 μl with RNase-free water. The samples were incubated at 37°C for 4 h, after which 75 μL of methanol was added to the system. Different dilutions were used for detection. UPLC separation was performed on an Acquity UPLC BEH amide column (1.7 μm, 2.1 mm × 100 mm, Waters Corporation, Massachusetts, United States) with a flow rate of 0.35 ml/min at 40°C. Formic acid in water (0.1%, v/v, solvent A, Aladdin, Shanghai) and methyl cyanides (v/v, solvent B, Sigma-Aldrich, Shanghai) were employed as the mobile phase. Mass spectrometry detection was performed in positive electrospray ionization mode, and multiple reaction monitoring (MRM) was used to monitor target nucleosides by using the following mass transitions: (precursor ions → product ions) of m^6A (282.1 → 150.1), A (268 → 136), U (245 → 113), G (284 → 152), C (244 → 112), and I (269 →137). Quantification was performed using a standard curve originating from A and m^6A. The calculated m^6A/A ratio was used to indicate the m^6A level. To achieve maximal detection sensitivity, we optimized the MRM parameters of all nucleosides.

RNA MeRIP-Seq Library Construction and Sequencing

MeRIP-seq was performed by Biotechnology Corporation (Shanghai, China). As previously described (Dominissini et al., 2012), GenSeq® m6A MeRIP KIT (GenSeq, China) was performed, and the protocol involved the following four steps. Step 1: RNA fragmentation. The RNA was randomly divided into approximately 200-nt segments by using fragmentation buffer and stop buffer. The sizes and concentrations of the RNA fragments were detected by an Agilent Bioanalyzer and an Agilent RNA 6000 Pico kit (Bioptic Inc., Taiwan, China). Three micrograms of fragmented RNA was used as the input group and stored at −80°C. The remaining fragmented RNA was used for subsequent immunoprecipitation experiments. Step 2: Preparation of immunoprecipitated magnetic beads. The porcine gastric mucine (PGM) magnetic beads were gently blown with a pipette to promote full suspension, after which they were washed with 1 × IP buffer, and an m^6A antibody was added. Step 3: Immunoprecipitation. The MeRIP reaction solution, which included 50 μl of fragmented RNA, 150 μl of nuclease-free water, and 50 μl of 5 × IP buffer, was prepared. Next, 250 μL of the reaction solution was added to the magnetic beads prepared in step 2. The magnetic beads were washed with 1 × IP buffer, LB buffer, and HS buffer separately and repeatedly. Step 4: RNA purification. The magnetic beads were resuspended in RLT buffer, washed with 75% ethanol, and then completely resuspended in 11.25 μl of nuclease-free water. The eluted RNA was transferred to a new centrifuge tube and then immediately used for subsequent experiments or stored at −80°C. Both the input samples without immunoprecipitation and the m^6A IP samples were used for RNA-seq library generation. The library quality was evaluated with a Bioptic Qsep100 Analyser (Bioptic Inc., Taiwan, China). Library sequencing was performed on an Illumina NovaSeq 6,000 (LC-Bio Technology) instrument with 150 bp paired-end reads.

FastQC was applied to analyse the quality of the sequencing data by assessing the sequencing quality distribution, base content distribution, and proportion of repeated sequencing fragments. HISAT2 software was used to compare the filtered clean reads with the human reference genome (GRCh38/hg38) and thereby obtain unique mapped reads for further analysis. Peak calling was carried out with exomePeak software. After obtaining the peak, the gene structure locations of the peaks and the overall distribution characteristics were determined, and peak annotation analysis was performed with HOMER software (http://homer.ucsd.edu/homer/ngs/peakMotifs.html).GO analysis is used to assess the function of genes from various aspects and is divided into three main categories: biological process (BP), molecular function (MF) and cellular component (CC). The mRNAs of genes modified by m^6A were evaluated based on GO annotations in the BP, MF, and CC categories in the database, and Fisher's test was used to determine the significance level (p-value) of each category to screen for significant GO terms. KEGG analysis was based on the gene annotations, and selected genes with mRNA m^6A modification were annotated according to their associated KEGG pathways. The significance level (p-value) was determined by Fisher's test to screen significant pathway terms related to m^6A gene enrichment.

Methylated RNA ImmunoprecipitationeRIP-qPCR

MeRIP-qPCR was performed by ChoudSeq Biotech Inc. (Shanghai, China). Its analysis was used according to a previously reported method (Dominissini et al., 2012), as shown in the RNA MeRIP-seq library construction and sequencing section. m6A enrichment was determined by qPCR analysis. Reverse transcription was performed using a SuperScriptTM III Reverse Transcriptase kit (Invitrogen) for mRNA. For relative qRT-PCR, qPCR SYBR Green master mix (CloudSeq)was used to generate mRNA cDNA according to the manufacturer's instructions. The reactions were performed on a QuantStudio five Real-Time PCR System (Thermo Fisher). The expression levels of mRNAs normalized to those of the endogenous control were calculated using the $2^{-\Delta\Delta CT}$ method and are presented as the fold change relative to the control group. % (IP/Input) was used to calculate the differences.The sequences of the primers utilized in this study are listed in Additional file 1: **Supplementary Table S1**.

Data Analysis

Data are presented as the mean ± SD. Significance differences between groups were determined by Student's t-test using GraphPad Prism six sofeware. Significance was established at $p < 0.05$.

TABLE 1 | Characteristics of subjects and m^{6}A levels in NMOSD patients and healthy controls.

Item	NMOSD (n = 3) for MeRIP-seq	HC (n = 3) for MeRIP-seq	*p*-value	NMOSD (n = 3) for MeRIP-qPCR	HC(n = 3) for MeRIP-qPCR	*p*-value
Mean age (mean ± SD; years)	47.33 ± 2.082	47.33 ± 3.055	$p > 0.05$	59.00 ± 6.083	59.67 ± 6.087	*p > 0.05*
Sex (female/male, %)	3/0 (100,0)	3/0 (100,0)	$p > 0.05$	3/0 (100,0)	3/0 (100,0)	*p > 0.05*
Anti-AQP4 antibody (positive/negative, n,%)	3/0 (100,0)			3/0 (100,0)		
Current disease duration (mean ± SD; days)	5 ± 2			5 ± 2		
m^{6}A level (mean ± SD; %)	0.09531 ± 0.009259	0.1399 ± 0.02533	$p < 0.05$	-	-	-

Note:SD,standard deviation; Anti-AQP4, antibody,anti-aquaporin-4 antibody.

RESULTS

Demographic and Clinical Features of the Patients With NMOSD

We enrolled three NMOSD patients and three HCs for MeRIP-seq, with mean ages of 47.33 ± 2.082 and 47.33 ± 3.055, respectively ($p > 0.05$). The female: male ratio was 3:0 in both groups ($p > 0.05$). The current disease duration at the time of recruitment was 5 ± 2 days (**Table 1**). We enrolled another three NMOSD patients and three HCs for MeRIP-qPCR validation, with mean ages of 59.00 ± 6.083 and 59.67 ± 6.807, respectively ($p > 0.05$). The female:male ratio was 3:0 in both groups ($p > 0.05$). The current disease duration at the time of recruitment was 5 ± 2 days (**Table 1**).

Levels of m^{6}A in NMOSD Patients and HCs

The levels of m^{6}A differed significantly between the NMOSD patients and HCs (0.09531 ± 0.009259 (%), 0.1399 ± 0.02533 (%); $p < 0.05$) (**Table 1**; **Figure 1A**). Additionally, R^2 was greater than 0.99, indicating that the linear relationship of the standard curve is good (**Figure 1B**).

Sequencing Statistics and Quality Control

Some joint and low-quality sequences were abtained in the original data, and adaptor and low-quality data were removed to abtain clean reads. In the MeRIP-seq library, 6,308,338,650, 7,779,927,600 and 6,841,024,050 total bases and 6,125,183,549, 7,628,452,162 and 6,635,546,539 valid total bases were abtained, and the effective bases accounted for 97.10, 98.10 and 97.00% of the totals, respectively. In the HC blood samples, 6,495,933,300, 6,730,312,650 and 6,736,672,200 total bases and 6,292,021,712, 6,453,732,389 and 6,604,071,380 valid total bases were obtained, with the effective bases accounting for 96.90, 95.90 and 98.00% of the totals, respectively. In the RNA-seq library, the effective bases accounted for more than 90% of the total bases in both the NMOSD and HC groups. The results are shown in **Table 2**.

Mapping Reads to the Reference Genome

We used HISAT2 to map reads to the GRCh38/hg38 genome with default parameters. Detailed statistical analyses were performed by comparing reads with reference sequences. In the MeRIP-seq library, the mapping ratios were 94.37, 87.19 and 91.55% in NMOSD patients, and 92.64, 93.87, 89.08% in HCs. In the RNA-seq library, the mapping ratios were 91.07, 89.26 and 86.16% in NMOSD patients and 89.31, 90.42, 88.04% in HCs. The ratios of unique mapped reads are shown in **Table 3**.

General Features of m^{6}A Methylation in NMOSD Patients and HCs

MeRIP-seq analysis of RNA derived from whole blood revealed 14,444 nonoverlapping m^{6}A peaks within 7,034 coding transcripts (mRNAs) and 440 nonoverlapping m^{6}A peaks within 330 lncRNAs in the NMOSD group; there were three biological replicates. In the HC group, there were 12,806 nonoverlapping m^{6}A peaks within 6,480 coding transcripts (mRNAs) and 369 nonoverlapping m^{6}A peaks within 265 lncRNAs, with three biological replicates. Of these, 5,568 coding transcripts (70.1%) within mRNAs (**Figure 1C**) and 199 long noncoding transcripts (50.3%) within lncRNAs (**Figure 1D**) overlapped between the NMOSD patients and HCs. Of these, there were 3,600 differential peaks ($p < 0.05$), including 3,371 differentially methylated peaks within mRNAs and 95 within lncRNAs ($p < 0.05$) (**Figure 1I**).

To analyse the distribution profiles of m^{6}A peaks within mRNAs, the peaks were categorized into five transcript segments: the 5'UTR; start codon (400 nucleotides centred on the start codon); coding sequence (CDS); stop codon (400 nucleotides centred on the stop codon); and 3'UTR. In our study, m^{6}A was most often detected in the CDS, with some sites being observed near the 3'UTR and stop codon (**Figures 1E–G**).

Motif analysis of peaks within mRNAs with the highest scores (*p*-value = 1e-141) obtained from three biological replicates revealed a consensus sequence (GGAC) as well as other motifs in the NMOSD and HC samples (**Figure 1H**), indicating the reliability of the data.

Distribution of Differentially Methylated m^{6}A Sites

In total, we identified 3,371 differentially methylated m^{6}A sites (DMMSs) within mRNAs, of which 68.5% (2,310/3,371) showed significantly increased methylation and 31.5% (1,061/3,371) significantly reduced methylation (NMOSD vs. HCs; **Table 4**). We also identified 95 DMMSs for lncRNAs, of which 68.5% (71/95) exhibited significant increased methylation levels and 31.5% (24/95) exhibited significantly decreased methylation levels (**Table 4**). **Tables 5**, **6** provide the top ten m^{6}A sites

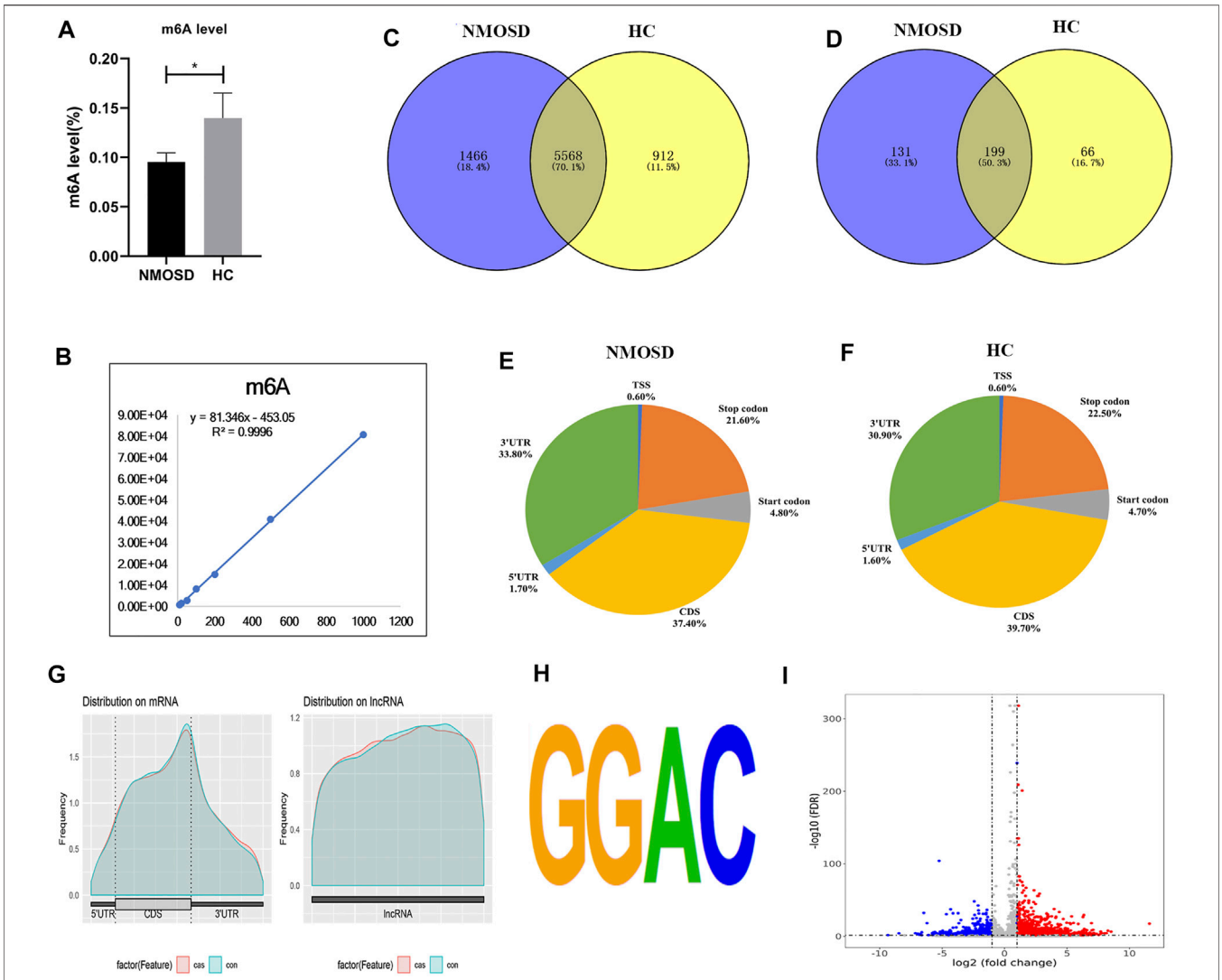

FIGURE 1 | Overview of N^6-methyladenosine methylation in NMOSD patients and HCs. **(A)** m^{6}A levels in NMOSD and HCs,*$p < 0.05$,Student's *t*-test. **(B)** R^2 was greater than 0.99, indicating that the linear relationship of the standard curve is good. **(C)** Venn diagram showing the overlap of m^{6}A peaks with mRNA in NMOSD patients and HCs. **(D)** We used a Venn diagram to depict the overlap of m^{6}A peaks with lncRNAs in NMOSD and HCs. **(E)** We used pie charts to show the accumulation of m^{6}A peaks along transcripts in NMOSD. **(F)** We used pie charts to display the accumulation of m^{6}A peaks along transcripts in HCs. **(G)** Metagene plot of peak distribution in RNA structures (NMOSD and HCs). **(H)** The consensus motif for m^{6}A modification is "GGAC".**(I)** Volcano map of differential m^{6}A modification. Diff.log2 FC < 0 indicates hypomethylation, indicating that the site was demethylated under treatment conditions; Diff.log2 FC > 0 indicates hypermethylation, indicating hypermethylation of this site under treatment conditions. Diff.lg.fdr: log10 (FDR) of differential methylation analysis is the log10 conversion value of the FDR value obtained from the differential analysis. Red indicates hypermethylation, and blue indicates demethylation. Red and blue represent genes with a more than 2-fold difference in m^{6}A modification.

displaying increased and reduced methylation levels with the highest fold change values (**Tables 5**, **6**).

Differentially Methylated RNAs Are Involved in Important Biological Pathways

To elucidate the functions of m^{6}A in NMOSD, we selected protein-coding genes containing DMMSs for GO and KEGG pathway analyses. In the BP category, genes with increased m^{6}A site methylation were significantly ($p < 0.05$) enriched in RNA splicing, mRNA processing, and gene expression (**Figure 2A**); genes with decreased m^{6}A methylation were highly enriched in the regulation of transcription, transcription, DNA templating, and viral processes (**Figure 2B**). In the MF category, poly(A) RNA binding and protein binding were enriched among mRNAs with increased m^{6}A methylation (**Figure 2A**), and those with lower levels of methylation were enriched in protein binding and DNA binding (**Figure 2B**). With respect to the CC category, the nucleus and nucleolus showed enrichment for both increased (**Figure 2A**) and decreased m^{6}A methylation (**Figure 2B**). We also used a GO bubble chart to illustrate the top twenty enriched GO terms, mainly involving transcription, DNA templating,

TABLE 2 | Summary of reads quality control analysis.

SampleName	TotalReads_Before	Totalbase_Before	Totalbase_After	BaseFilter% (%)
Patient1-IP	42,055,591	6,308,338,650bp	6,125,183,549bp	97.10
Patient2-IP	51,866,184	7,779,927,600bp	7,628,452,162bp	98.10
Patient3-IP	45,606,827	6,841,024,050bp	6,635,546,539bp	97.00
Control1-IP	43,306,222	6,495,933,300bp	6,292,021,712bp	96.90
Control2-IP	44,868,751	6,730,312,650bp	6,453,732,389bp	95.90
Control3-IP	44,911,148	6,736,672,200bp	6,604,071,380bp	98.00
Patient1-Input	44,180,024	6,627,003,600bp	6,300,789,080bp	95.10
Patient2-Input	57,064,051	8,559,607,650bp	8,145,517,911bp	95.20
Patient3-Input	39,713,253	5,956,987,950bp	5,475,127,085bp	91.90
Control1-Input	43,211,484	6,481,722,600bp	6,080,268,117bp	93.80
Control2-Input	43,081,347	6,462,202,050bp	6,047,838,238bp	93.60
Control3-Input	40,372,667	6,055,900,050bp	5,664,859,110bp	93.50

TABLE 3 | Summary of reads mapped to the GRCh38/hg38 reference genome.

Item	All	UnMapped	UniqueMapped	UniqueMappedRate (%)	RepeatMapped	Mapped	MappedRate (%)
Patient1-IP	42,055,591	2,367,986	38,154,965	90.73	1,532,640	39,687,605	94.37
Patient2-IP	51,866,184	6,643,849	41,866,190	80.72	3,356,145	45,222,335	87.19
Patient3-IP	45,606,827	3,854,600	39,952,635	87.60	1,799,592	41,752,227	91.55
Control1-IP	43,306,222	3,188,802	38,005,282	87.76	2,112,138	40,117,420	92.64
Control2-IP	44,868,751	2,751,102	40,569,248	90.42	1,548,401	42,117,649	93.87
Control3-IP	44,911,148	4,905,279	36,597,922	81.49	3,407,947	40,005,869	89.08
Patient1-Input	44,180,024	3,943,392	36,724,619	83.12	3,512,013	40,236,632	91.07
Patient2-Input	57,064,051	6,131,507	40,183,337	70.42	10,749,207	50,932,544	89.26
Patient3-Input	39,713,253	5,497,001	31,232,336	78.64	2,983,916	34,216,252	86.16
Control1-Input	43,211,484	4,620,990	34,806,849	80.55	3,783,645	38,590,494	89.31
Control2-Input	43,081,347	4,127,518	35,559,483	82.54	3,394,346	38,953,829	90.42
Control3-Input	40,372,667	4,830,090	31,472,902	77.96	4,069,675	35,542,577	88.04

TABLE 4 | General numbers of differentially methylated peaks and genes.

Item	Increased methylation peak	Increased methylation gene	Decreased methylation peak	Increased methylation gene
mRNA	2,310	1,855	1,061	974
lncRNA	71	63	24	24

TABLE 5 | Top ten increased methylation peaks.

Chromosome	thickStart	thickEnd	Gene name	Fold change	Class
14	74,736,350	74,736,560	FCF1	11.6	three_prime_utr
17	62,551,671	6,2,551,852	AC008026.3	8.53	exon
12	6,593,496	6,594,518	CHD4	8.28	CDS
12	68,670,970	68,671,151	RAP1B	8.15	three_prime_utr
1	16,574,534	16,574,684	NBPF1	8.02	CDS
17	28,042,829	28,043,009	NLK	7.99	CDS
1	211,574,019	211,574,200	SLC30A1	7.75	three_prime_utr
1	1,340,169	1,341,701	DVL1	7.75	CDS
6	33,663,772	33,665,055	ITPR3	7.7	CDS
15	34,980,730	34,980,881	ZNF770	7.67	three_prime_utr

Note: thickStart/thickEnd: start/end position of the differentially methylated RNA peak; CDS, coding sequence.

TABLE 6 | Top ten reduced methylation peaks.

Chromosome	thickStart	thickEnd	Gene name	Fold change	Class
18	10,454,687	10,474,202	APCDD1	−9.34	CDS
4	186,708,548	186,708,998	FAT1	−8.46	CDS
9	38,397,946	38,398,155	ALDH1B1	−7.14	three_prime_utr
2	237,366,914	237,367,035	COL6A3	−6.98	CDS
15	100,900,720	100,907,186	ALDH1A3	−6.94	CDS
16	81,045,709	81,045,920	ATMIN	−6.83	three_prime_utr
16	67,828,216	67,828,456	CENPT	−6.78	exon
16	15,107,433	15,110,752	PKD1P6-NPIPP1	−6.75	exon
5	16,699,540	16,701,140	MYO10	−6.64	CDS
2	187,484,874	187,488,374	TFPI	−6.63	exon

Note: thickStart/thickEnd: start/end position of differentially methylated RNA peak; CDS, coding sequence.

regulation of transcription, and gene expression, among others (**Figure 2C**).

We performed KEGG pathway analysis of DMMSs and observed that genes with increased m^{6}A methylation were significantly ($p < 0.05$) enriched in the mRNA surveillance pathway, neurotrophin signalling pathway, and Epstein-Barr virus infection, among others (**Figure 3A**). Genes with decreased m^{6}A methylation were significantly ($p < 0.05$) enriched in the TNF signalling pathway, oestrogen signalling pathway, and focal adhesion (**Figure 3B**). We also used a KEGG bubble chart to show the top twenty enriched pathways, including Epstein-Barr virus infection and the mRNA surveillance pathway (**Figure 3C**). These results suggest that m^{6}A may have many key roles in NMOSD.

Association Analysis of Differential RNA Modification and Differential Gene Expression

We drew a four-quadrant map based on two sets of data: RNA-seq, all gene differential expression table (including non-significant results); and RNA modification, all differential peak table (including nonsignificant results). The top 10 genes (the largest absolute values of diff.log2. fc are shown in purple) were determined to be MT-RNR1, C60rf203, XK, CD22, SEMA3A, HECW2, TIGD5, VISIG4, HLA-DQB1 and NOC2L. (**Figure 4A**).

Validation of Differential RNA Modifications by MeRIP-qPCR

In order to verify the reliability of MeRIP-seq results, MeRIP-qPCR was used to verify the modification levels of modification sites on the target genes.We chose the top four hypomethylated genes and the top four hypermethylated genes as target genes (**Figure 4A**). Among the eight target genes, MeRIP-qPCR results of MT-RNR1, XK, CD22, SEMA3A, HECW2 and TIGD5 were consistent with the sequencing results, and there were significant differences ($p < 0.05$) (**Figure 4B**). MeRIP-qPCR results of VISIG4 and C60rf203 were opposite with the sequencing results, although there were significant differences ($p < 0.05$) (**Figure 4B**). Nevertheless, it suggested that our sequencing results were credible.

DISCUSSION

m^{6}A modification is a dynamic and reversible RNA modification occurring in eukaryotes that is carried out by "writers", "erasers", and "readers". In recent years, studies have found that m^{6}A is involved in the regulation of the immune microenvironment (Chen et al., 2019b; Tang et al., 2020; Xu et al., 2020). NMOSD is a kind of autoimmune disease that is triggered by disruption of immune homeostasis. In this study, we assessed the m^{6}A status in NMOSD patients and HCs, revealing differences between the groups that support the dynamic features of m^{6}A modification. Indeed, we found that the m^{6}A level was much lower in patients than in HCs, demonstrating that m^{6}A modification is very important in the process of NMOSD. In addition, Zhang found that altering m^{6}A levels might be a novel way for improving the efficacy of platinum in cancer treatment (Zhang et al., 2021a), and Zhao reported that m^{6}A regulation could influence the progression of lung adenocarcinoma (Zhao and Xie, 2021). We hypothesized that increasing the level of m^{6}A might relieve the symptoms of NMOSD, but further studies are needed to confirm this hypothesis.

In the present study, many important biological pathways were found to be related to differentially methylated mRNAs. Because of the construction of strand-specific libraries, we also predicted lncRNAs harbouring m^{6}A peaks and identified DMMSs. While lncRNAs are known to play important roles in mediating transcriptional and posttranscriptional regulation (Cantile et al., 2021; Moison et al., 2021), their biological functions in regards to m^{6}A modification remain unknown. In general, the m^{6}A modification of lncRNAs is very common in cancers. For example, Wu found that m^{6}A-induced lncRNA RP11 could cause the dissemination of colorectal cancer cells (Wu et al., 2019). In our study, we investigated the m^{6}A modification of lncRNAs between NMOSD patients and HCs, which highlighted immune homeostasis. However, further analysis is needed to validate these results.

Evidence supports a strong relationship between immune homeostasis and m^{6}A modification, with m^{6}A modification playing important biological roles in NMOSD (Chen et al., 2020; Chong et al., 2021; Zhang et al., 2021b; Zhong et al., 2021). In our study, the GO and KEGG analyses of mRNAs harbouring DMMSs showed that those with increased methylation were mostly enriched in RNA splicing, mRNA

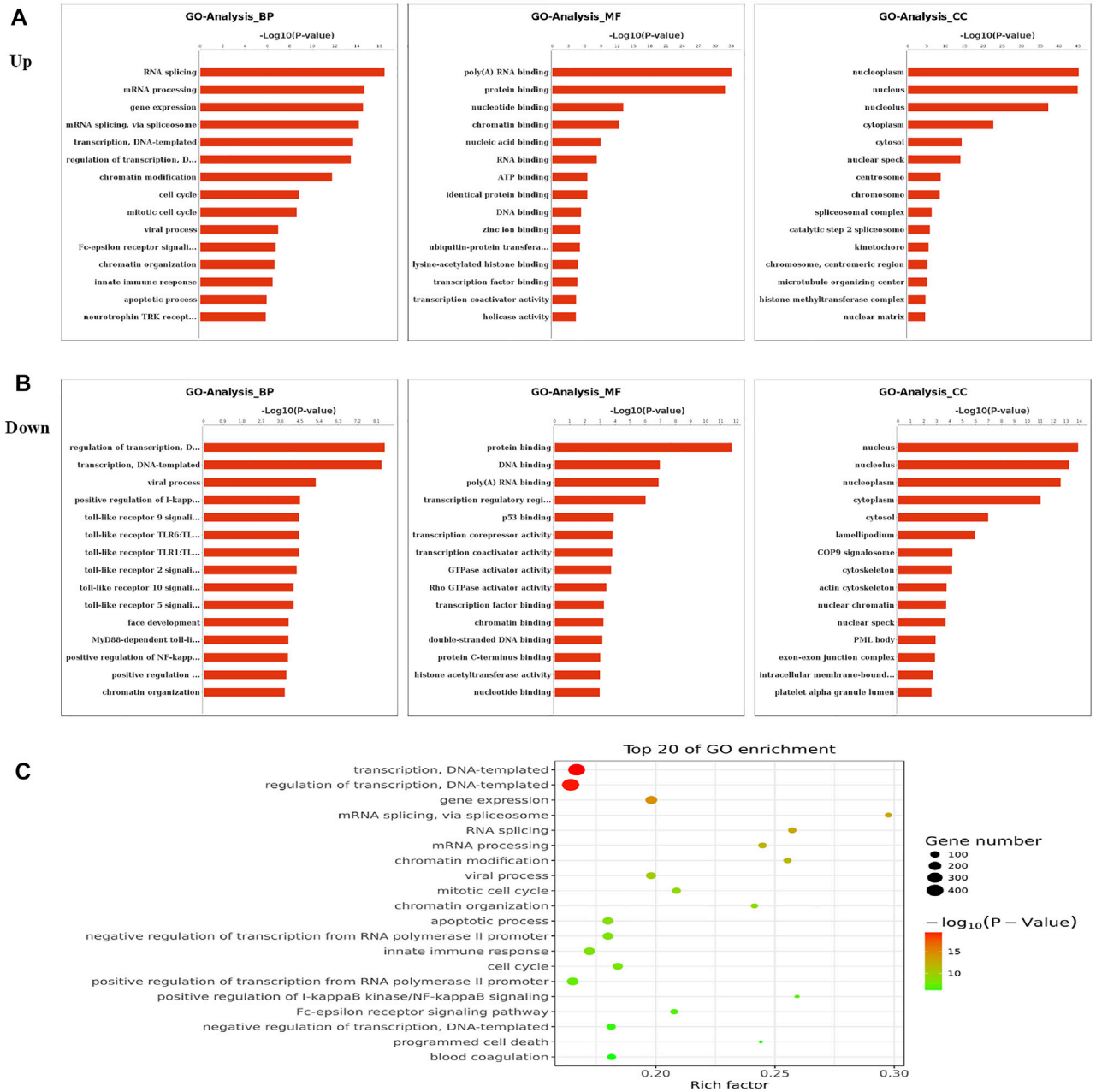

FIGURE 2 | Gene ontology and Enrichment analysis of differential m^{6}A peaks. **(A)** The top ten Gene Ontology terms related to mRNAs with increased methylation; terms in the biological process (BP), molecular function (MF), and cellular component (CC) categories are shown; (–log10p) for significance. The log-transformed results were used for visualization. **(B)** The top ten Gene Ontology terms related to mRNAs with decreased methylation, terms in the BP, MF and CC categories are shown; (–log10p) for significance. The log-transformed results were used for visualization. **(C)** The GO bubble shows the top 20 Gene Ontology terms corresponding to the methylated genes.

processing and gene expression, which supports the importance of m^{6}A modification in NMOSD. For example, the increased expression of phosphatase and tensin homolog deleted on chromosome Ten (PTEN) and an approximately twofold enhancement methylation were observed in NMOSD patients compared with the HCs, indicating PTEN serves as a key enzyme in NMOSD (**Figure 5A**) (Han et al., 2019). Furthermore, PTEN is involved in inositol phosphate metabolism. Gene expression is important in NMOSD, and our results suggest a link between NMOSD and the regulation of gene expression, subsequently resulting the disruption of immune homeostasis.

In contrast to m^{6}A sites with increased methylation, those with decreased methylation were mostly enriched in regulation of transcription, DNA templating, and positive regulation of

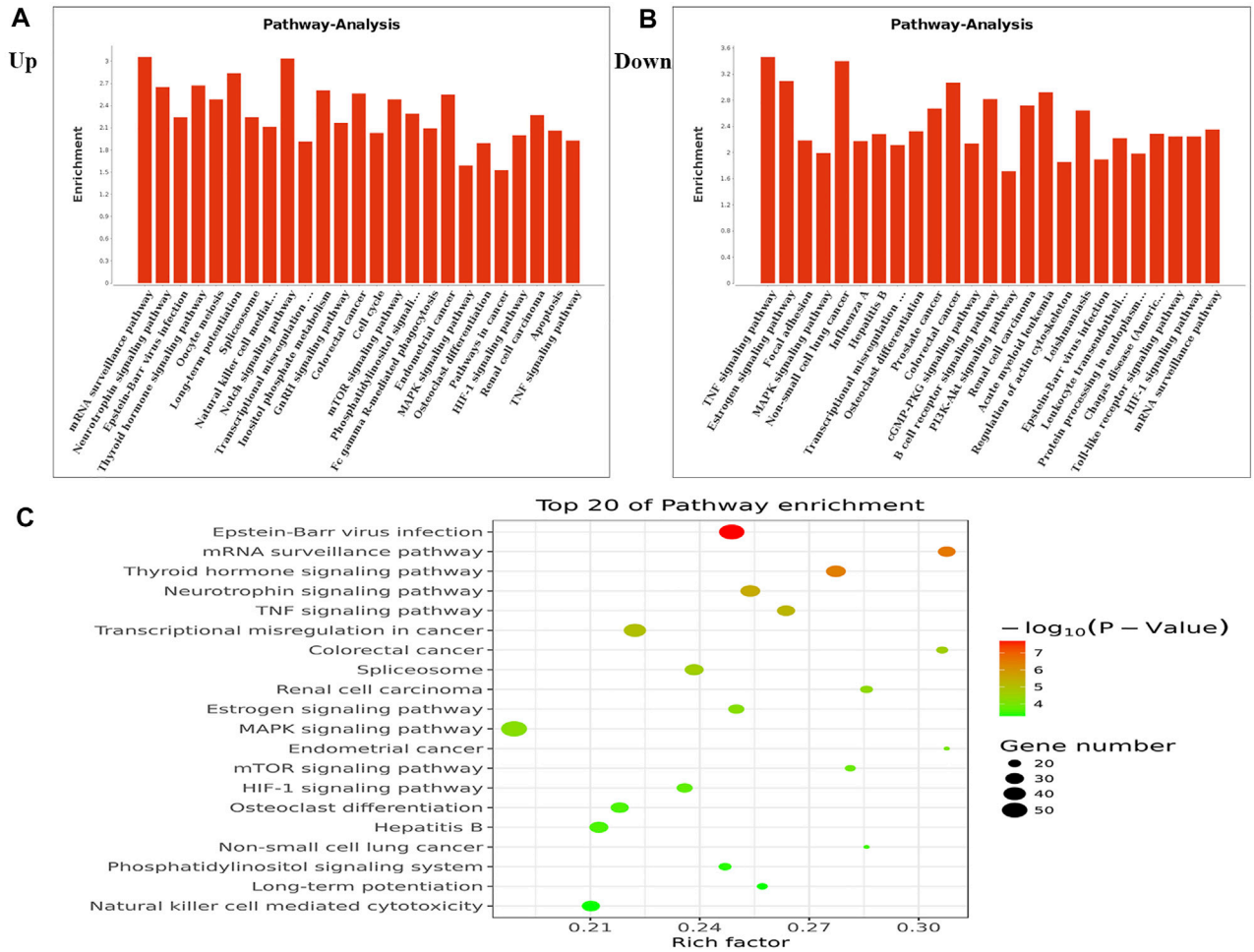

FIGURE 3 | KEGG pathway and Enrichment analysis of differential m^{6}A peaks. **(A)** The top ten enrichment pathway terms related to mRNAs with increased methylation are shown; (–log10p) for significance. The log-transformed results were used for visualization. **(B)** The top ten enrichment pathway terms related to mRNAs with decreased methylation; (–log10p) for significance. The log-transformed results were used for visualization. **(C)** The KEGG bubble shows the top 20 enrichment pathway terms related to methylated mRNAs.

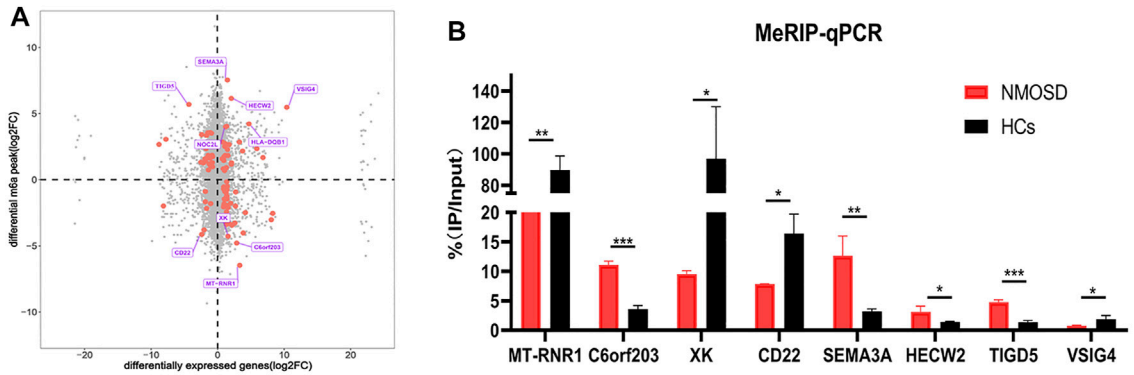

FIGURE 4 | Differential RNA modification and validation. **(A)** Analysis of differential RNA modification and differential gene expression. The X-axis represents the RNA-seq differential genes, and the Y-axis represents the RNA-modified differential peak genes; the top 10 genes (largest absolute values of diff.log2. fc) are shown in purple. The red dots represent significant differences (both RNA-seq and RNA modification data meet $p < 0.05$); the grey dots indicate that the data did not meet the condition required for the red dots. **(B)** Validation of the top four hypomethylated genes and the top four hypermethylated genes by using MeRIP-qPCR.

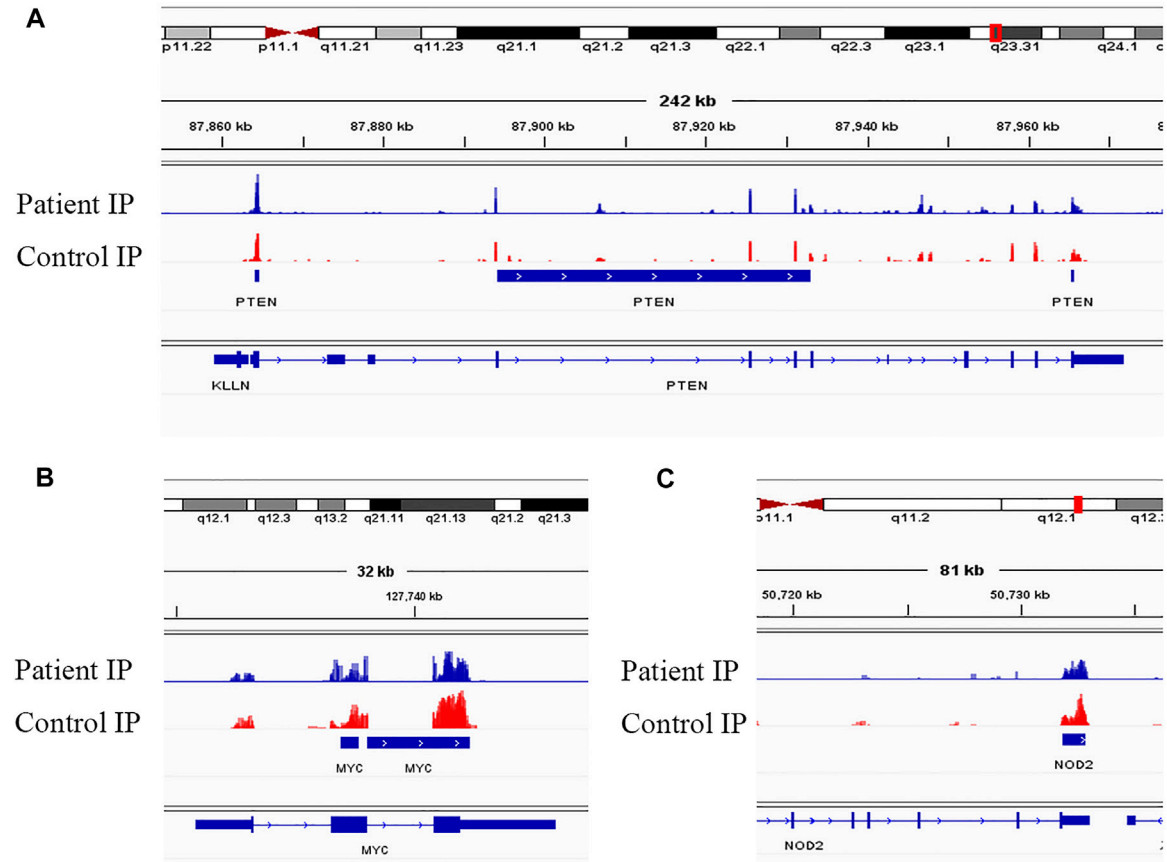

FIGURE 5 | Visualization of representative genes. **(A)** Visualization of the m^{6}A-modified gene PTEN in NMOSD patients and HCs. **(B)** Visualization of the m^{6}A-modified gene MYC in NMOSD patients and HCs. **(C)** Visualization of the m^{6}A-modified gene NOD2 in NMOSD patients and HCs.

I-kappa B kinase/NF-kappa B signalling pathways, which include genes such as v-myc avian myelocytomatosis viral oncogene homologue (MYC) (**Figure 5B**) and nucleotide-binding oligomerization domain containing 2 (NOD2) (**Figure 5C**). "Readers", which include YTHDF1-3, recognize and bind to m^{6}A modifications on mRNAs (Wang et al., 2014; Li et al., 2017; Jiang et al., 2019), thereby playing an important role in regulating translation. According to recent studies, we hypothesize that the methylation of mRNAs is associated with their translation, as it may influence their expression, leading to subsequent changes in global translation.

In our association analysis of differential RNA modification and differential gene expression, the top 10 genes included SEMA3A and CD22 (**Figure 4A**). Zhang also reported that SEMA3A may negatively regulate axonal regeneration in retinal ganglion cells (Zhang et al., 2020), and CD22 is an important inhibitory molecule on the surface of B cells that negatively regulates B cell activation (Gross et al., 2009). All these findings are very beneficial for future studies.

Although few studies have explored the roles of m^{6}A modification in NMOSD, there are still some limitations to our study. For example, we may need to expand the number of enrolled patients. In addition, data from the same patient should be obtained before and after treatment.

CONCLUSION

Our study demonstrates the differential m^{6}A methylome in NMOSD patients compared to healthy controls, suggesting a strong association between m^{6}A methylation and the regulation of immune homeostasis in NMOSD. The findings fundamentally contribute to future studies on immune homeostasis in NMOSD.

DATA AVAILABILITY STATEMENT

The raw data have been made publicly available under SRA accession number PRJNA737585.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of Shanghai First

Rehabilitation Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HY contributed to the conception of the study and wrote the manuscript. Y-FW and WL contributed to the data collection. JD and D-SZ contributed to the data analyses. X-FS and Y-TG helped to guide this study.

FUNDING

This study was supported by the National Natural Science Foundation of China (81771295) and the Shanghai Municipal Health Commission Foundation (202140414).

ACKNOWLEDGMENTS

The author would like to thank the NMOSD patients and HCs who provided blood for this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.735454/full#supplementary-material

REFERENCES

Cantile, M., Di Bonito, M., Tracey De Bellis, M., and Botti, G. (2021). Functional Interaction Among lncRNA HOTAIR and MicroRNAs in Cancer and Other Human Diseases. *Cancers* 13, 570. doi:10.3390/cancers13030570

Chen, J., Fang, X., Zhong, P., Song, Z., and Hu, X. (2019a). N6-Methyladenosine Modifications: Interactions with Novel RNA-Binding Proteins and Roles in Signal Transduction. *RNA Biol.* 16, 991–1000. doi:10.1080/15476286.2019.1620060

Chen, Y. G., Chen, R., Ahmad, S., Verma, R., Kasturi, S. P., Amaya, L., et al. (2019b). N6-Methyladenosine Modification Controls Circular RNA Immunity. *Mol. Cell* 76, 96–109. doi:10.1016/j.molcel.2019.07.016

Chen, S., Zhang, N., Zhang, E., Wang, T., Jiang, L., Wang, X., et al. (2020). A Novel m6A Gene Signature Associated with Regulatory Immune Function for Prognosis Prediction in Clear-Cell Renal Cell Carcinoma. *Front. Cell Dev. Biol.* 8, 616972. doi:10.3389/fcell.2020.616972

Chong, W., Shang, L., Liu, J., Fang, Z., Du, F., Wu, H., et al. (2021). m6A Regulator-Based Methylation Modification Patterns Characterized by Distinct Tumor Microenvironment Immune Profiles in Colon Cancer. *Theranostics* 11, 2201–2217. doi:10.7150/thno.52717

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., et al. (2012). Topology of the Human and Mouse m6A RNA Methylomes Revealed by m6A-Seq. *Nature* 485, 201–206. doi:10.1038/nature11112

Gross, A. J., Lyandres, J. R., Panigrahi, A. K., Prak, E. T. L., and Defranco, A. L. (2009). Developmental Acquisition of the Lyn-CD22-SHP-1 Inhibitory Pathway Promotes B Cell Tolerance. *J. Immunol.* 182, 5382–5392. doi:10.4049/jimmunol.0803941

Han, J., Wang, J.-z., Yang, X., Yu, H., Zhou, R., Lu, H.-C., et al. (2019). METTL3 Promote Tumor Proliferation of Bladder Cancer by Accelerating Pri-miR221/222 Maturation in m6A-Dependent Manner. *Mol. Cancer* 18, 110. doi:10.1186/s12943-019-1036-9

Hanniford, D., Ulloa-Morales, A., Karz, A., Berzoti-Coelho, M. G., Moubarak, R. S., Sánchez-Sendra, B., et al. (2020). Epigenetic Silencing of CDR1as Drives IGF2BP3-Mediated Melanoma Invasion and Metastasis. *Cancer Cell* 37, 55–70. doi:10.1016/j.ccell.2019.12.007

Heck, A. M., Russo, J., Wilusz, J., Nishimura, E. O., and Wilusz, C. J. (2020). YTHDF2 Destabilizes m6A-Modified Neural-specific RNAs to Restrain Differentiation in Induced Pluripotent Stem Cells. *RNA* 26, 739–755. doi:10.1261/rna.073502.119

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., et al. (2011). N6-methyladenosine in Nuclear RNA Is a Major Substrate of the Obesity-Associated FTO. *Nat. Chem. Biol.* 7, 885–887. doi:10.1038/nchembio.687

Jiang, Q., Sun, B., Liu, Q., Cai, M., Wu, R., Wang, F., et al. (2019). MTCH2 Promotes Adipogenesis in Intramuscular Preadipocytes via an M 6 A-YTHDF1-Dependent Mechanism. *FASEB j.* 33, 2971–2981. doi:10.1096/fj.201801393RRR

Lennon, V. A., Wingerchuk, D. M., Kryzer, T. J., Pittock, S. J., Lucchinetti, C. F., Fujihara, K., et al. (2004). A Serum Autoantibody Marker of Neuromyelitis Optica: Distinction from Multiple Sclerosis. *Lancet* 364, 2106–2112. doi:10.1016/s0140-6736(04)17551-x

Li, A., Chen, Y.-S., Ping, X.-L., Yang, X., Xiao, W., Yang, Y., et al. (2017). Cytoplasmic m6A Reader YTHDF3 Promotes mRNA Translation. *Cell Res.* 27, 444–447. doi:10.1038/cr.2017.10

Lin, S., Choe, J., Du, P., Triboulet, R., and Gregory, R. I. (2016). The M(6) A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* 62, 335–345. doi:10.1016/j.molcel.2016.03.021

Lin, Z., Hsu, P. J., Xing, X., Fang, J., Lu, Z., Zou, Q., et al. (2017). Mettl3-/Mettl14-Mediated mRNA N6-Methyladenosine Modulates Murine Spermatogenesis. *Cell Res.* 27, 1216–1230. doi:10.1038/cr.2017.117

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., et al. (2014). A METTL3-METTL14 Complex Mediates Mammalian Nuclear RNA N6-Adenosine Methylation. *Nat. Chem. Biol.* 10, 93–95. doi:10.1038/nchembio.1432

Moison, M., Pacheco, J. M., Lucero, L., Fonouni-Farde, C., Rodríguez-Melo, J., Mansilla, N., et al. (2021). The lncRNA APOLO Interacts with the Transcription Factor WRKY42 to Trigger Root Hair Cell Expansion in Response to Cold. *Mol. Plant* 14 (6), 937–948. doi:10.1016/j.molp.2021.03.008

Pandit, L., Asgari, N., Apiwattanakul, M., Palace, J., Paul, F., Leite, M., et al. (2015). Demographic and Clinical Features of Neuromyelitis Optica: A Review. *Mult. Scler.* 21, 845–853. doi:10.1177/1352458515572406

Ping, X.-L., Sun, B.-F., Wang, L., Xiao, W., Yang, X., Wang, W.-J., et al. (2014). Mammalian WTAP Is a Regulatory Subunit of the RNA N6-Methyladenosine Methyltransferase. *Cel Res.* 24, 177–189. doi:10.1038/cr.2014.3

Pittock, S. J., and Lucchinetti, C. F. (2016). Neuromyelitis Optica and the Evolving Spectrum of Autoimmune Aquaporin-4 Channelopathies: a Decade Later. *Ann. N.Y. Acad. Sci.* 1366, 20–39. doi:10.1111/nyas.12794

Su, D., Chan, C. T. Y., Gu, C., Lim, K. S., Chionh, Y. H., Mcbee, M. E., et al. (2014). Quantitative Analysis of Ribonucleoside Modifications in tRNA by HPLC-Coupled Mass Spectrometry. *Nat. Protoc.* 9, 828–841. doi:10.1038/nprot.2014.047

Tang, R., Zhang, Y., Liang, C., Xu, J., Meng, Q., Hua, J., et al. (2020). The Role of m6A-Related Genes in the Prognosis and Immune Microenvironment of Pancreatic Adenocarcinoma. *PeerJ* 8, e9602. doi:10.7717/peerj.9602

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., et al. (2014). N6-methyladenosine-Dependent Regulation of Messenger RNA Stability. *Nature* 505, 117–120. doi:10.1038/nature12730

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., et al. (2015). N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell* 161, 1388–1399. doi:10.1016/j.cell.2015.05.014

Wingerchuk, D. M., Banwell, B., Bennett, J. L., Cabre, P., Carroll, W., Chitnis, T., et al. (2015). International Consensus Diagnostic Criteria for Neuromyelitis Optica Spectrum Disorders. *Neurology* 85, 177–189. doi:10.1212/wnl.0000000000001729

Wu, Y., Yang, X., Chen, Z., Tian, L., Jiang, G., Chen, F., et al. (2019). m6A-induced lncRNA RP11 Triggers the Dissemination of Colorectal Cancer Cells via Upregulation of Zeb1. *Mol. Cancer* 18, 87. doi:10.1186/s12943-019-1014-2

Xu, S., Tang, L., Dai, G., Luo, C., and Liu, Z. (2020). Expression of m6A Regulators Correlated with Immune Microenvironment Predicts Therapeutic Efficacy and Prognosis in Gliomas. *Front. Cell Dev. Biol.* 8, 594112. doi:10.3389/fcell.2020.594112

Xue, L., Li, J., Lin, Y., Liu, D., Yang, Q., Jian, J., et al. (2021). m 6 A Transferase METTL3-induced lncRNA ABHD11-AS1 Promotes the Warburg Effect of Non-small-cell Lung Cancer. *J. Cell Physiol.* 236, 2649–2658. doi:10.1002/jcp.30023

Zhang, Z., Wang, M., Xie, D., Huang, Z., Zhang, L., Yang, Y., et al. (2018). METTL3-mediated N6-Methyladenosine mRNA Modification Enhances Long-Term Memory Consolidation. *Cell Res* 28, 1050–1061. doi:10.1038/s41422-018-0092-9

Zhang, J., Liu, W., Zhang, X., Lin, S., Yan, J., and Ye, J. (2020). Sema3A Inhibits Axonal Regeneration of Retinal Ganglion Cells via ROCK2. *Brain Res.* 1727, 146555. doi:10.1016/j.brainres.2019.146555

Zhang, M., Bai, M., Wang, L., Lu, N., Wang, J., Yan, R., et al. (2021a). Targeting SNHG3/miR-186-5p Reverses the Increased m6A Level Caused by Platinum Treatment through Regulating METTL3 in Esophageal Cancer. *Cancer Cell Int.* 21, 114. doi:10.1186/s12935-021-01747-9

Zhang, X., Zhang, S., Yan, X., Shan, Y., Liu, L., Zhou, J., et al. (2021b). m6A Regulator-mediated RNA Methylation Modification Patterns Are Involved in Immune Microenvironment Regulation of Periodontitis. *J. Cell Mol Med.* 25, 3634–3645. doi:10.1111/jcmm.16469

Zhao, W., and Xie, Y. (2021). KIAA1429 Promotes the Progression of Lung Adenocarcinoma by Regulating the m6A Level of MUC3A. *Pathol. - Res. Pract.* 217, 153284. doi:10.1016/j.prp.2020.153284

Zheng, G., Dahl, J. A., Niu, Y., Fedorcsak, P., Huang, C.-M., Li, C. J., et al. (2013). ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol. Cell* 49, 18–29. doi:10.1016/j.molcel.2012.10.015

Zhong, J., Liu, Z., Cai, C., Duan, X., Deng, T., and Zeng, G. (2021). m6A Modification Patterns and Tumor Immune Landscape in clear Cell Renal Carcinoma. *J. Immunother. Cancer* 9, e001646. doi:10.1136/jitc-2020-001646

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 06 December 2021
doi: 10.3389/fgene.2021.771847

Identification of the Pyroptosis-Related Gene Signature and Risk Score Model for Colon Adenocarcinoma

Bixian Luo†, *Jianwei Lin*†, *Wei Cai** and *Mingliang Wang**

Department of General Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

The prognosis of advanced colon adenocarcinoma (COAD) remains poor. However, existing methods are still difficult to assess patient prognosis. Pyroptosis, a lytic and inflammatory process of programmed cell death caused by the gasdermin protein, is involved in the development and progression of various tumors. Moreover, there are no related studies using pyroptosis-related genes to construct a model to predict the prognosis of COAD patients. Thus, in this study, bioinformatics methods were used to analyze the data of COAD patients downloaded from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases to construct a risk model for the patient prognosis. TCGA database was used as the training set, and GSE39582 downloaded from GEO was used as the validation set. A total of 24 pyroptosis-related genes shown significantly different expression between normal and tumor tissues in COAD and seven genes (CASP4, CASP5, CASP9, IL6, NOD1, PJVK, and PRKACA) screened by univariate and LASSO cox regression analysis were used to construct the risk model. The receiver operating characteristic (ROC) and Kaplan–Meier (K–M curves) curves showed that the model based on pyroptosis-related genes can be used to predict the prognosis of COAD and can be validated by the external cohort well. Then, the clinicopathological factors were combined with the risk score to establish a nomogram with a C-index of 0.774. In addition, tissue validation results also showed that CASP4, CASP5, PRKACA, and NOD1 were differentially expressed between tumor and normal tissues from COAD patients. In conclusion, the risk model based on the pyroptosis-related gene can be used to assess the prognosis of COAD patients well, and the related genes may become the potential targets for treatment.

Keywords: pyroptosis, colon adenocarcinoma, risk score, gene signature, nomogram

OPEN ACCESS

Edited by:
Dongyu Zhao,
Peking University, China

Reviewed by:
Antonio Russo,
University of Palermo, Italy
Yanchun Zhang,
Boston Children's Hospital and Harvard Medical School, United States

***Correspondence:**
Mingliang Wang
wml_2902@163.com
Wei Cai
caiwei@shsmu.edu.cn

†*These authors have contributed equally to this work and share first authorship*

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *07 September 2021*
Accepted: *12 November 2021*
Published: *06 December 2021*

Citation:
Luo B, Lin J, Cai W and Wang M (2021) Identification of the Pyroptosis-Related Gene Signature and Risk Score Model for Colon Adenocarcinoma. Front. Genet. 12:771847. doi: 10.3389/fgene.2021.771847

INTRODUCTION

Colon cancer is one of the most common cancers in the world (Bray et al., 2018). Colon adenocarcinoma (COAD) is the most common type of colon cancer (Fleming et al., 2012). The main treatment for colorectal adenocarcinoma is surgery, but the 5-year survival of patients is not satisfactory due to postoperative recurrence and metastasis (Zhai et al., 2017). Biomarkers have been used to aid in identifying patients at high risk of tumor progression or recurrence, such as the RAS mutation state, BRAF mutation state, and microsatellite instability (MSI) state (Sepulveda et al., 2017;

Koncina et al., 2020). Therefore, the determination of molecular changes in COAD patients has become a focus of COAD research.

Pyroptosis is a lytic and inflammatory process of programmed cell death caused by the gasdermin protein (Wang et al., 2017). The members of gasdermin families include GSDMA, GSDMB, GSDMC, GSDMD, GSDME (also known as DFNA5), and PJVK (also known as DFNB59) (Broz et al., 2020). In contrast to apoptosis, pyroptosis can cause plasma membrane rupture, pore formation, cytoplasmic swelling, and chromatin condensation (Fink and Cookson, 2006). After cell rupture, proinflammatory cytokines and immunogenic substances are released to promote immune cell activation and infiltration, which may result in a strong inflammatory response and significant tumor regression (Wang et al., 2017; Loveless et al., 2021). Increasing studies indicated that pyroptosis may play important roles in the development of many cancers (Al Mamun et al., 2021). In COAD, pyroptosis may participate in the tumorigenesis of cancer (Tan et al., 2020; Tang et al., 2020). Meanwhile, pyroptosis induction can increase the chemosensitivity of COAD (Guo et al., 2021). Hence, pyroptosis-related genes may become the potential biomarkers to predict the prognosis of COAD and provide guidance for treatment.

In this study, bioinformatics was used to determine the expression levels of relevant genes between normal tissues and tumor tissues in COAD to explore the prognostic value of these genes. Then, a risk model based on pyroptosis-related genes was constructed by univariate and LASSO cox regression analysis. Moreover, a nomogram established by the clinicopathological features and risk model was used to further improve the prognostic ability in COAD. Finally, the tumor tissue and paired normal tissue from 13 patients with COAD were used to verify the gene expression in the model.

METHODS

Data Collection

The mRNA expression data (Workflow Type: HT seq-FPKM) and relevant clinical information for COAD patients downloaded from TCGA website (https://portal.gdc.cancer.gov/repository) on 2 August, 2021, were used as the training set. There are 437 samples (39 normal tissues and 398 COAD tissues) and 384 cases of COAD patients being collected. The data of GSE39582 from GEO (http://www.ncbi.nlm.nih.gov/geo/) was used as the validation set. A total of 579 cases of COAD patients were collected. The GEO samples were analyzed using the Affymetrix Human Genome U133 Plus 2.0 Array platform. Patients from TCGA and GEO with missing clinical information were deleted in subsequent studies. Clinicopathological characteristics of patients (age, gender, stage, T stage, N stage, and M stage) were recorded.

Identification of Differentially Expressed Genes

The pyroptosis-related genes were obtained from the research about ovarian cancer (Ye et al., 2021). The expression data of these pyroptosis-related genes were downloaded from TCGA databases. The "limma" package of R language was regarded as the method to identify differentially expressed genes (DEGs) with the P-value <0.05. The significance of DEGs was marked as follows: * for P-value < 0.05, ** for P-value < 0.01, and *** for P-value <0.001. The STRING database (http://string-db.org/) was set for searching the online for possible interactions between related genes. The PPI network of DEGs was constructed by this database.

Construction of the Risk Score Model by Univariate Cox and LASSO Cox Regression Analyses

Univariate regression analysis was used to screen pyroptosis-related genes associated with prognosis. The cutoff P-value was set to 0.2 to prevent omissions. LASSO analysis with the "glmnet" R package was utilized to construct the risk score model after univariate regression analysis. Each COAD patient risk score was calculated by this model and used to divide patients into two groups (low-risk and high-risk groups) by the median value. The receiver operating characteristic (ROC) and Kaplan–Meier (K–M curves) curves were used to evaluate the prognostic ability of the risk model. PCA was analyzed by R language with the "stats" package. GSE39582 was regarded as the validation set to verify the predictive ability of the prognostic risk model based on TCGA database. In addition, univariate and multivariate Cox regression analyses were also used for verifying whether the risk model was associated with prognosis and could be used as an independent prognostic risk factor for COAD.

Functional Enrichment Analysis and Alteration of Seven Genes in the Model

The cBioPortal dataset (https://www.cbioportal.org/), which contains genomic data from 104 different tumors, was used to study genetic variations of the genes in our model. GeneMANIA (http://www.genemania.org)contains genetic information, analysis of gene lists, and functional analysis of prioritized genes, with high predictive algorithms (Warde-Farley et al., 2010). Thus, GeneMANIA was used to analyze genes interacting with gene models and their enrichment function.

Construction of the Nomogram to Estimate the Clinical Outcome of COAD Patients

The "rms" R package was used to construct a nomogram. Calibration curves were used to test association between the predicted outcome and the actual situation in 1-, 3-, and 5 -year survival rate. GSE39582 was the validation dataset for the nomogram.

Correlation of the Genes and Risk Score with Clinicopathological Features and Immune Cells as well as Immune Signaling Pathways

The "beeswarm" R package was used to assess the correlation of the genes and risk score with clinicopathological features. The

TABLE 1 | Primer sequences of genes in the risk model used for qPCR.

Gene	Primer sequence
β-actin	F:GACCTGTACGCCAACACAGT
	R:CTCAGGAGGAGCAATGATCT
CASP4	F:TCTGCGGAACTGTGCATGATG
	R:TGTGTGATGAAGATAGAGCCCAT
CASP5	F:TTCAACACCACATAACGTGTCC
	R:GTCAAGGTTGCTCGTTCTATGG
CASP9	F:CTGTCTACGGCACAGATGGAT
	R:GGGACTCGTCTTCAGGGGAA
IL6	F:ACTCACCTCTTCAGAACGAATTG
	R:CCATCTTTGGAAGGTTCAGGTTG
NOD1	F:ACTGAAAAGCAATCGGGAACTT
	R:CACACACAATCTCCGCATCTT
PJVK	F:CTGGATCAGATTCCATTGCAGT
	R:GTGGTTCGGATGCTCTCCAT
PRKACA	F:CAAGGAGACCGGGAACCACTA
	R:CATTCAGGGTGTGTTCGATCTG

qPCR, quantitative real-time polymerase chain reaction.

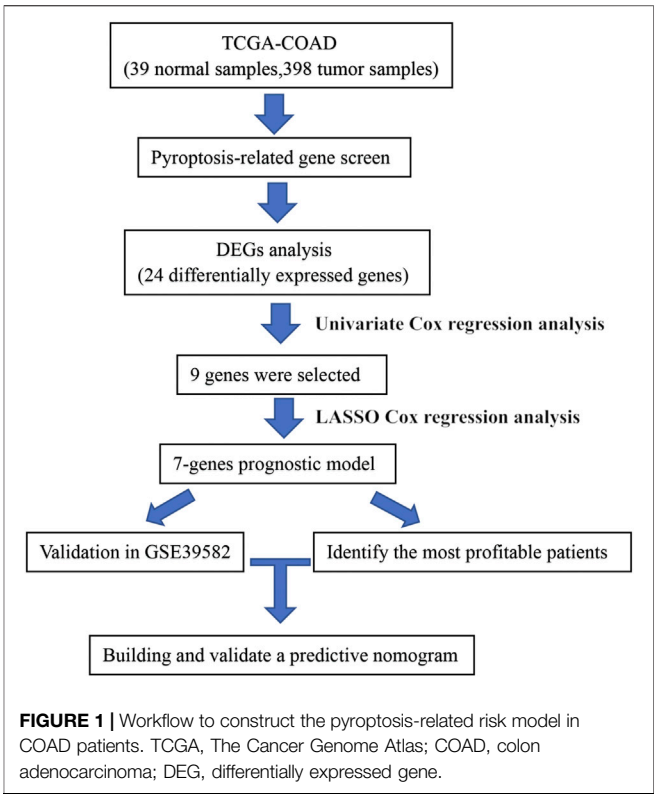

FIGURE 1 | Workflow to construct the pyroptosis-related risk model in COAD patients. TCGA, The Cancer Genome Atlas; COAD, colon adenocarcinoma; DEG, differentially expressed gene.

tumor immune microenvironment is the key to tumor–anti-tumor immunity. The "gsva" R package was utilized to conduct the ssGSEA to calculate the scores of infiltrating immune cells and to evaluate the activity of immune-related pathways.

Tissue Sample Collection

A total of thirteen pairs of fresh tumor tissues from COAD patients and their paired normal tissues were collected from Ruijin Hospital of Shanghai Jiao Tong University, Shanghai Jiao Tong University School of Medicine, which was approved by the Human Research Ethics Committee of this hospital. All fresh samples were stored at −80°C for the following experiments.

Ethics Statement

This study was approved by the Human Ethics Committee of Ruijin Hospital. Informed consent was obtained from all enrolled patients and healthy donors.

Quantitative Real-Time PCR

The total RNAs from tissues were extracted with the Trizol reagent (Invitrogen, CA, USA). The NanoDrop 2000 spectrophotometer (Thermo) was used to quantify RNA, and cDNA was generated by the PrimeScript™ RT Reagent Kit (Takara, China) and then analyzed using RT-qPCR with the TB Green® Premix Ex Taq™ II (Takara, China) on the 7,500 Fast Real-Time PCR System (Applied Biosystems, CA, USA). β-actin was exploited as an internal reference. The mRNA relative expression of individual genes was detected by $2^{-\Delta Ct}$ methods. The primer sequences used for analysis are listed in **Table 1**.

Statistical Analysis

R version 4.0.5, Perl version 5.28, and Graphpad Prism 8.0.2.263 were used for statistical analysis. TCGA and GEO data were organized by Excel Office 2019. Except that the P-value < 0.2 was set as the condition for screening prognostic genes in univariate Cox regression analysis, the P-value < 0.05 was used as the significant condition for others without special explanation.

TABLE 2 | Clinical information of COAD patients from TCGA and GEO.

		TCGA	GEO
		n = 384	n = 579
Age	≤65	159	227
	>65	225	351
	Unknown	0	1
Gender	Female	180	260
	Male	204	319
Stage	Stage 0	0	4
	Stage I–II	216	306
	Stage III–IV	157	269
	Unknown	11	0
T	T0	0	1
	T1–2	77	60
	T3–4	306	495
	Tis	1	3
	Unknown	0	20
M	M0	285	496
	M1	54	61
	Mx	39	2
	Unknown	6	20
N	N0	230	311
	N1–3	154	242
	Unknown	0	26

T, T stage; N, N stage; M, M stage; Stage, TNM, stage.

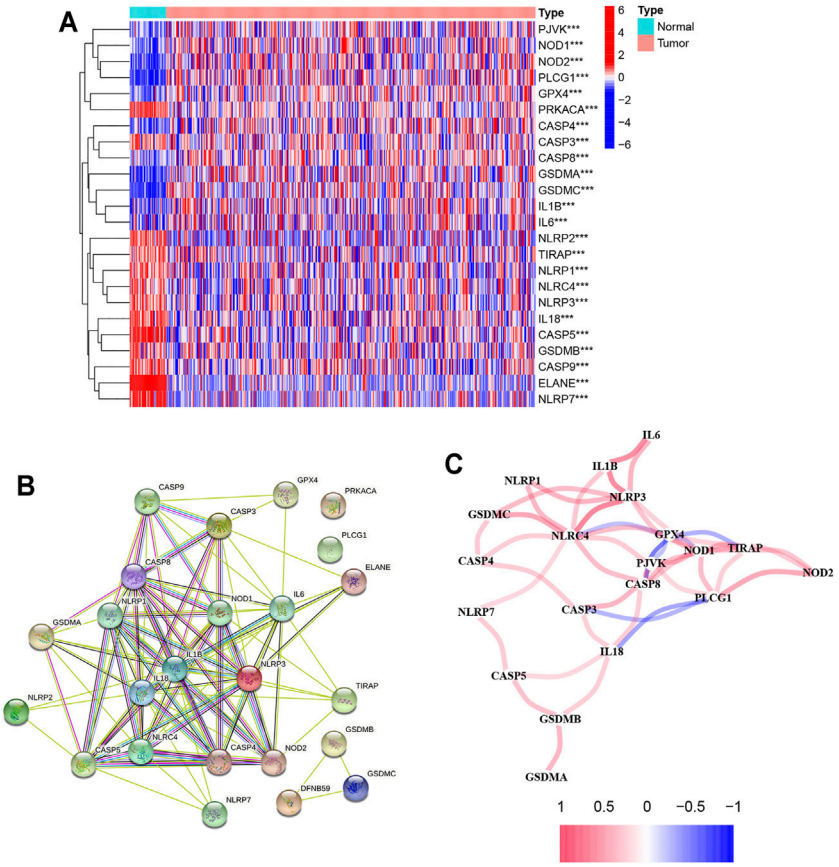

FIGURE 2 | Results of differential gene analysis. **(A)** Heatmap of differentially expressed pyroptosis-related genes. The vertical axis refers to genes, the horizontal axis refers to differences in the gene expression between tissues, the orange means high expression, and the blue means low expression. P-values are shown as: $^{**}p < 0.01$; $^{***}p < 0.001$; **(B)** PPI network showing the interactions of differentially expressed pyroptosis-related genes; **(C)** correlation of the differentially expressed pyroptosis-related genes (red line: positive correlation; blue line: negative correlation. The depth of the colors reflects the strength of the relevance).

RESULTS

Overall Design of the Study

The flow chart of this study is shown in **Figure 1**. The relevant clinical information of patients from TCGA and GEO is shown in **Table2**. There were 24 pyroptosis-related genes being screened. A total of seven genes were selected after univariate and LASSO Cox regression analysis. The ROC and K–M curves were used to evaluate the prognostic ability of the risk model based on the seven pyroptosis-related genes. The GSE39582 from GEO was used as the external cohort to validate the model and nomogram. The calibration and C-index verified the predictive ability of the nomogram.

Identification of Differentially Expressed Genes

There were 24 pyroptosis-related genes screened by analyzing the data from TCGA database which is shown in **Figure 2A**. Of these genes, 11 genes were upregulated (CASP8, NOD1, GPX4, CASP4, PJVK, IL6, IL1B, PLCG1, NOD2, GSDMA, and GSDMC) and 13 genes were downregulated (ELANE, CASP5, NLRP7, IL18, NLRP3, NLRC4, PRKACA, NLRP1, GSDMB, CASP9, CASP3, TIRAP, and NLRP2) between normal and tumor tissues. The PPI network based on the STRING database showed IL18, IL1B, IL6, NLRP3, NLRP1, NLRC4, CASP5, NOD1, CASP8, NOD2, CASP4, and CASP3 were the key genes which interacted with more other genes (**Figure 2B**). The correlation network containing 24 pyroptosis-related genes is shown in **Figure 2C**.

Tumor Classification Based on the Differentially Expressed Pyroptosis-Related Genes

To explore the relationship between the 24 pyroptosis-related genes and COAD subtypes, the consensus clustering analysis was used to analyze the patients from TCGA database. The patients whose follow-up time was less than 30 days were excluded from the consensus clustering analysis. As the clustering variable (k) increased (from 2 to 10), intragroup connections were the highest and intergroup connections were the lowest when k = 2, indicating that the COAD patients could be well divided into

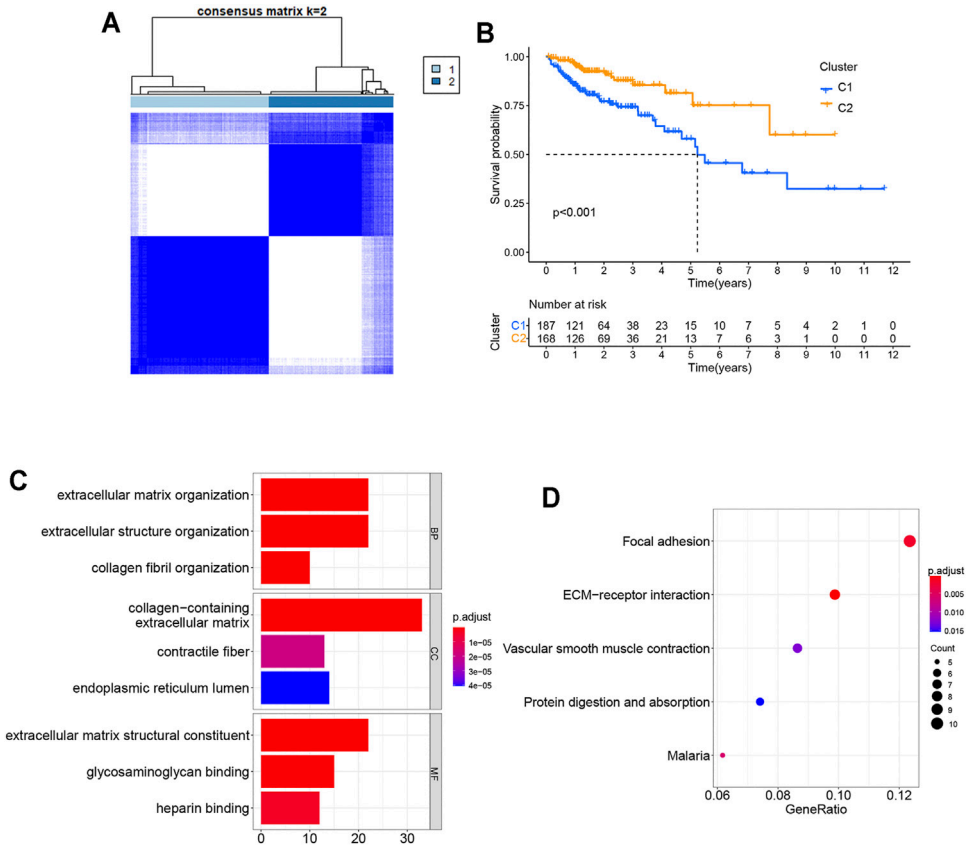

FIGURE 3 | Tumor classification based on the pyroptosis-related DEGs. **(A)** COAD patients were grouped into two clusters according to the consensus clustering matrix (k = 2). **(B)** Kaplan–Meier OS curves for the two clusters. **(C)** GO bar graph for genes in BP, CC, and MF. **(D)** Bubble graph of top 5 KEGG pathways with the most enriched genes; the vertical axis refers to names of the pathway; and the horizontal axis refers to the number of genes. GO, Gene Ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

two clusters based on the 24 differentially expressed genes (DEGs) (**Figure 3A**). The overall survival (OS) time was also compared between two clusters and showed great differences (**Figure 3B**). DEGs and clinical features between two groups are shown in **Supplementary Figure S1**, indicating that 207 genes and stages M stage and N stage were significantly different. Gene Ontology (GO) analysis for 207 genes showed that these genes participated in extracellular matrix organization, extracellular structure organization, collagen fibril organization, and so on (**Figure 3C**). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for 207 genes is shown in **Figure 3D**.

Risk Score Model

Univariate Cox regression analysis was performed on 24 differentially expressed pyroptosis-related genes in TCGA database with 384 COAD patients. When the P-value was set to 0.2, nine genes were identified (**Figure 4A**). Then, LASSO Cox regression analysis showed that seven genes could be used to construct prognostic risk models in the optimum λ value (**Figures 4B,C**). The risk score was calculated as follows:

$$\begin{aligned} risk\ score = {} & (CASP4 * 0.34668) + (CASP5 * -0.15059) \\ & + (CASP9 * -0.89296) + (IL6 * 0.11284) \\ & + (NOD1 * 0.55807) + (PJVK * 0.34451) \\ & + (PRKACA * 0.45435). \end{aligned}$$

Based on the median risk score, the patients from TCGA and GEO databases were divided into two groups (low-risk and high-risk groups). As shown in **Figure 4D**, the prognosis was worse in the high-risk group. In addition, the validation set proved that the risk model based on pyroptosis-related genes could well predict the prognosis of patients (**Figure 4E**). The principal component analysis (PCA) showed that patients with different risks were well separated into two clusters (**Figures 4F,G**).

The AUC of ROC curves in 1-year, 3-year, and 5-year were 0.630, 0.685, and 0.756, respectively (**Figure 4H**). In TCGA database, the expression levels of seven genes in different groups of tumor patients and corresponding clinical information are shown in **Figure 5A**. The univariate and multivariate Cox regression were used to evaluate whether the risk model in our study was the independent prognostic factor in

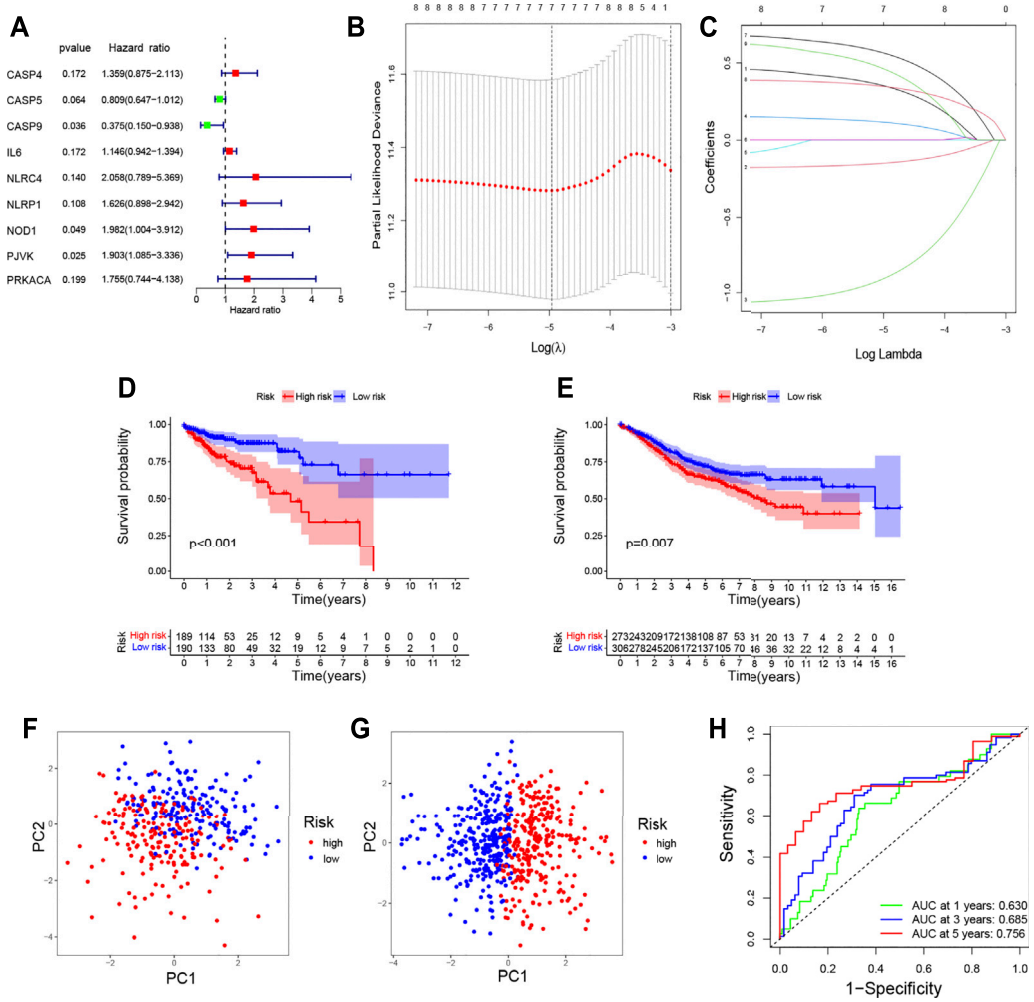

FIGURE 4 | Construction of the prognostic model for COAD. **(A)** Hazard ratio of univariate Cox analysis for pyroptosis-related DEGs; **(B)** distribution of LASSO coefficients for seven genes. Two vertical lines represent lambda. min and lambda. lse; **(C)** coefficients for seven genes analyzed by LASSO; survival analysis to verify the prognostic model in TCGA **(D)** and GEO **(E)**; PCA plot for COAD based on the risk score in TCGA **(F)** and GEO **(G)**; **(H)** time-dependent ROC curves for COAD.

COAD patients. As shown in **Figures 5B,C**, the risk score model was the independent prognostic factor for COAD.

In order to further understand patients who were more suitable for our risk model, patients were grouped according to different clinicopathological features and then prognostic analysis was performed. As shown in **Supplementary Figures S2A–F**, the patients with parameters such as age < = 65, male, stage I–II, T3–4, N0, and M0 seem more suitable for our risk model. This result suggested that our model was more suitable for *in situ* invasive COAD.

Functional Enrichment Analysis and Alteration of Seven Genes in the Model

The cBioPortal database was used to analyze the mutations of seven genes. Of the 524 patients with colorectal cancer, 247 had mutations in the genes in the seven-gene model (**Supplementary Figure S3A**). Of 332 colon cancer patients, 45.78% had the mutation (**Supplementary Figure S3B**). Functional analysis of 7 pyroptosis-related genes by GeneMANIA showed that the function of these genes focused on the inflammasome complex, positive regulation of cysteine-type endopeptidase activity, positive regulation of cysteine-type endopeptidase activity involved in the apoptotic process, and so on (**Supplementary Figure S3C**). These results indicated that the set of seven genes was able to greatly expand the level of difference detected in COAD patients. At the same time, the functions of the seven genes and their interactions showed that the genes in our gene model were mainly involved in inflammatory processes and cell death processes.

Construction and Validation of the Nomogram Based on Clinicopathological Features and Risk Score

The nomogram in this research was established by the clinicopathological features (age and stage) and risk score

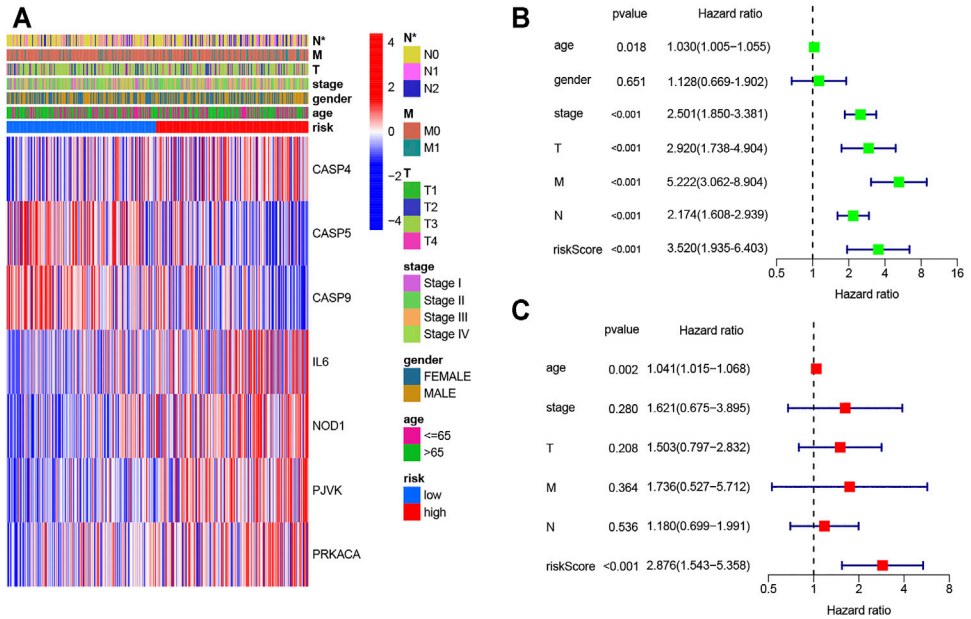

FIGURE 5 | Univariate and multivariate Cox regression analyses for the risk score. **(A)** Heatmap (blue: low expression; red: high expression) for the connections between clinicopathological features and the risk groups (*$p < 0.05$); **(B)** univariable Cox regression analysis for the risk score; **(C)** multivariable Cox regression analysis for the risk score.

based on the data from TCGA (**Figure 6A**). The C-index of the nomogram in predicting the survival rate was 0.774. The calibration curves for internal validation are shown in **Figure 6B**. The calibration curves for external validation are shown in **Figure 6C**. These results showed that the ability in predicting the prognosis of COAD patients can be improved by combining clinicopathological features with the risk score. At the same time, our verification also proved that this nomogram can well predict the prognosis of patients.

Correlation of the Genes and Risk Score with Clinicopathological Features and Immune Cells as well as Immune Signaling Pathways

Based on the risk score, the patients from TCGA were divided into two groups (low-risk and high-risk groups). Enrichment scores of 16 types of immune cells and the activity of 13 functional analyses of immune-related pathways were compared between two groups by the single-sample gene set enrichment analysis (ssGSEA). In TCGA cohort, the high-risk group was associated with higher levels of infiltration of activated dendritic cells (aDCs), macrophages, neutrophils, T-helper cells, and tumor infiltrating lymphocytes (TILs) (**Figure 7A**). In addition, the immune-related function enrichment of the high-risk group is shown in **Figure 7B**. Other immune cells and the immune-related function enrichment that did not show significant difference were shown in **Supplementary Figures S4A–B**. The risk score was greatly associated with clinicopathological factors of TCGA-COAD ($p < 0.05$; **Figures 7C,D**). The high-risk score was associated with the advanced TNM stage and N stage. The association between genes in risk models and clinicopathological features is shown in **Supplementary Figure S5**.

The mRNA Relative Expression of Seven Genes

Based on the verification of the tissues from COAD patients by the method of qPCR, the results showed that CASP4, CASP5, PRKACA, and NOD1 were expressed differentially between normal and tumor tissues (**Figures 8A–D**). The corresponding *p*-values of CASP4, CASP5, PRKACA, and NOD1 were 0.0384, 0.0392, 0.0173, and 0.0288, respectively. In addition, as shown in **Supplementary Figures S6A–C**, the expression of CASP9, IL6, and PJVK did not show significant difference, but the trend of the gene expression between tumor and normal tissue can still be seen, which may require us to expand the sample size to further prove. In conclusion, relevant validation results suggested that most of the genes in the model have different expressions, and more samples may be required for further validation.

DISCUSSION

To our knowledge, this was the first study to use pyroptosis-related genes to construct a risk model to predict prognosis of patients with COAD. In this study, the results indicated that pyroptosis-related gene risk model can be used to predict the prognosis of COAD. Meanwhile, the risk model was associated with clinicopathological factors, immune cells, and immune-related functions.

Pyroptosis is associated with many diseases (Al Mamun et al., 2020; Al Mamun et al., 2021; Chu et al., 2016; Wang et al., 2018).

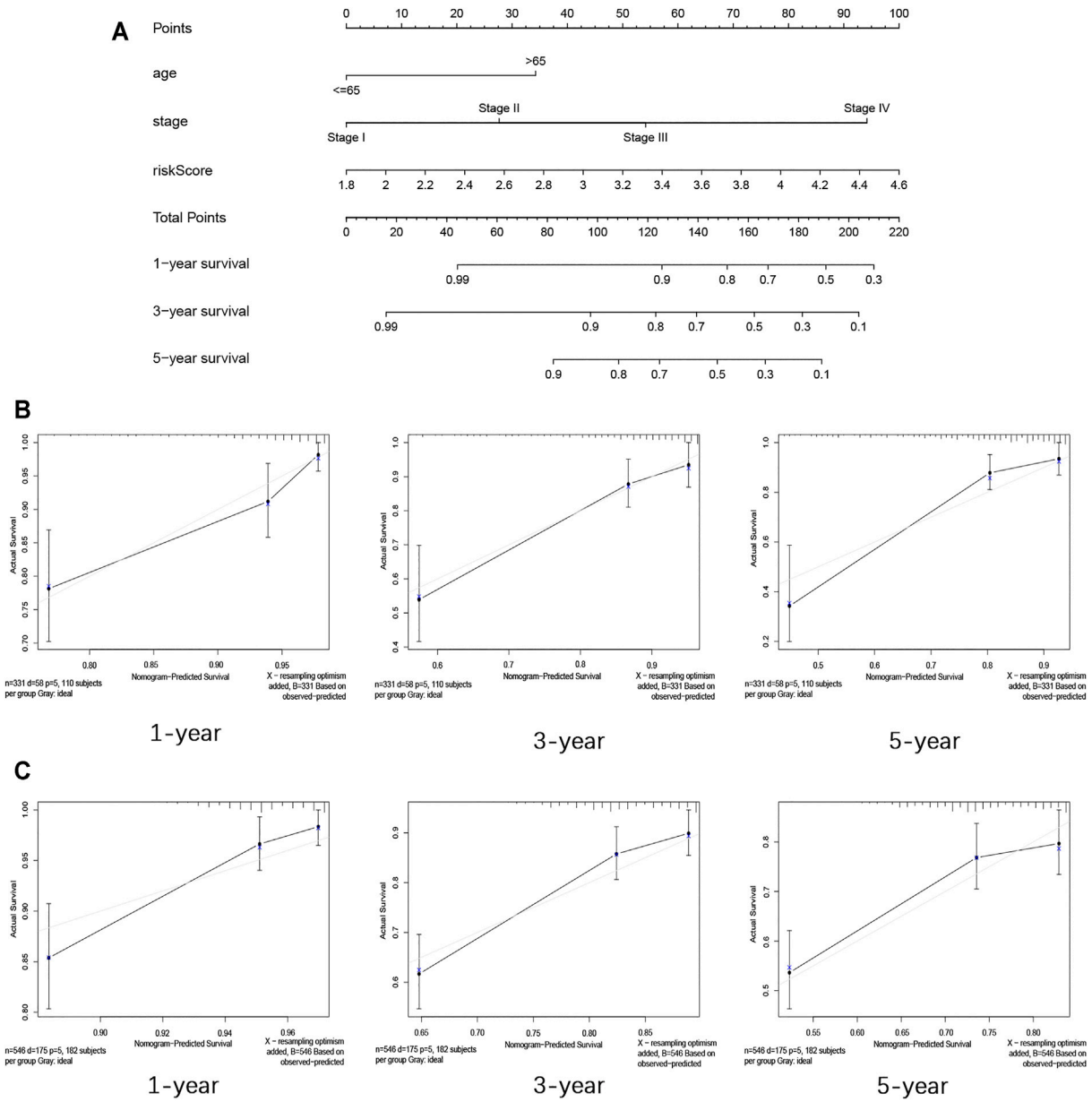

FIGURE 6 | Construction and validation of the nomogram for COAD. **(A)** Nomogram for predicting the 1-year, 3-year, and 5-year survival rate by age, stage, and risk score. **(B)** 1-, 3-, and 5-year calibration curves of TCGA dataset. **(C)** 1-, 3-, and 5-year calibration curves of GEO dataset.

Meanwhile, pyroptosis may mediate cell death by Caspase-1–dependent, Caspase-4/5/11–dependent, and Caspase-3/8–dependent inflammasome signaling pathways (Antonopoulos et al., 2015; Yu et al., 2019; Chen et al., 2020; Liu et al., 2021; L.; Wang et al., 2021). Recently, increasing studies show that pyroptosis is associated with the development of cancer (Al Mamun et al., 2021). Wang et al. indicted that 5-FU can induce pyroptosis in gastric cancer cell by Caspase-3 signaling pathways (Wang et al., 2018). Chu et al. showed that the signaling pathway of pyroptosis participated in the development of hepatocellular carcinoma and may become the target for treatment (Chu et al., 2016). As for COAD, Hu et al. found that pyroptosis was involved in the tumor development of colitis-related COAD (Hu et al., 2010). Chen et al. showed that pyroptosis played important roles in growth and metastasis of COAD (Tang et al., 2020). These studies suggested that pyroptosis-related genes may be the potential biomarkers for cancer. However, no studies have used pyroptosis-related genes to construct a risk model in COAD. In our study, data from TCGA were used as a training set, and data from GEO were used as a validation set to construct a seven-pyroptosis-related risk model by using univariable and LASSO Cox regression analyses. The prognostic model can well predict the prognosis of patients, which will be helpful to clinical evaluation and provide new therapeutic targets.

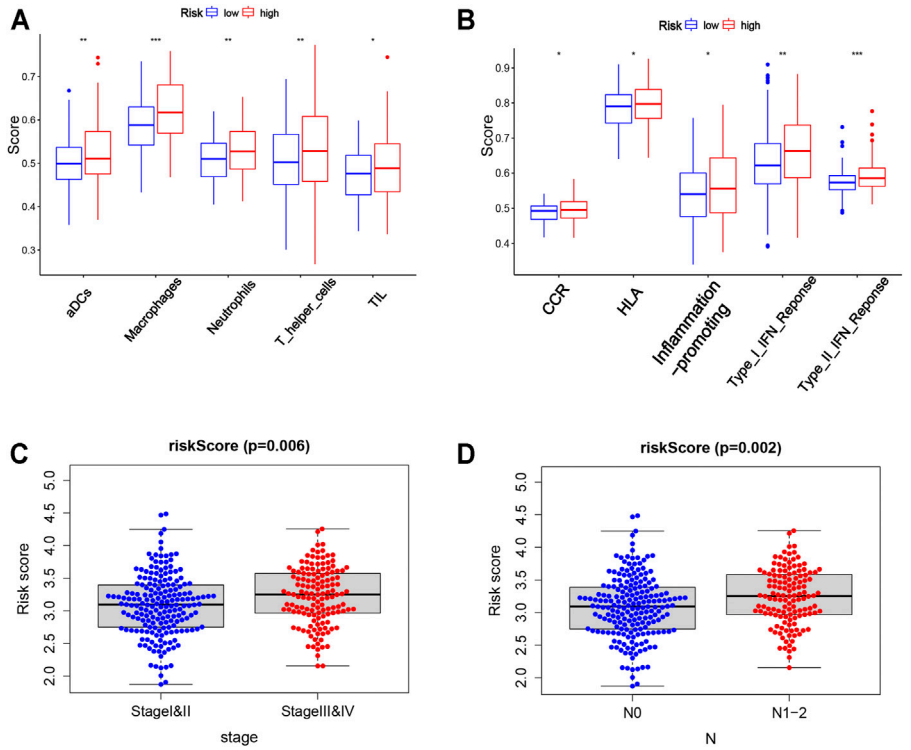

FIGURE 7 | Correlation of the risk score with clinicopathological features and immune cells as well as immune function. **(A,B)** Comparison of the enrichment scores of 5 types of immune cells and 5 immune-related functions between low- (blue box) and high-risk (red box) groups in the TCGA cohort. **(C,D)** Correlation between the risk score and clinicopathological features. Stage, tumor–node–metastasis (TNM) stage; N, node.

Genes in this risk model have been reported in studies of cancer. The Caspase 4 (CASP4) gene encodes a protein involved in immunity and inflammation (McIlwain et al., 2013). As the tumor-suppressor gene, CASP4 is associated with the poor outcome of esophageal squamous cell carcinoma (Shibamoto et al., 2017). Meanwhile, a study suggested that CASP4 may become the potential biomarker for diagnosis and treatment of tumors (Flood et al., 2015). Caspase 5 (CASP5) is an acknowledged frameshift target in the microsatellite instability gastrointestinal tract (Schwartz et al., 1999; Trojan et al., 2004). Caspase 9 (CASP9) can target colorectal cancer stem cells by inducible CASP9 to decrease the tumor size (Kemper et al., 2012). Interleukin 6 (IL6) is an important mediator of inflammatory responses and contributes to the development of inflammatory diseases (Koper-Lenkiewicz et al., 2021). IL6 plays an important role in the autophagy and chemotherapy resistance of COAD (Hu et al., 2021). Nucleotide oligomerization domain receptor 1 (NOD1) is a cytoplasmic pattern recognition receptor (Jiang et al., 2020). Some studies show that NOD1 can promote the carcinogenesis and metastasis of COAD, which may be the target to reduce postoperative recurrence (Jiang et al., 2020; Maisonneuve et al., 2021). Pejvakin (PJVK) is related to various auditory phenotypes in patients (Cheng et al., 2020). In serous ovarian cancer, the PJVK expression is downregulated, but its significance in tumors is unclear and needs further investigation (Berkel and Cacan, 2021). Protein kinase cAMP–activated catalytic subunit alpha (PRKACA) can interact with protein kinase cAMP–dependent type I regulatory subunit alpha (PRKAR1A) to inhibit the activity of protein kinase A, which may participate in the growth of COAD (Tseng et al., 2017; Zhao et al., 2021). In phosphorylation, PRKACA can help C16-ceramide to induce EMD phosphorylation so that it enhances the autophagosomal formation in COAD (Deroyer et al., 2014). These studies suggested that seven pyroptosis-related genes are associated with tumor. Our risk model was constructed by the tumor-related genes.

Survival analysis revealed that the risk model based on pyroptosis-related genes was more suitable for young males with early-stage cancer. In addition, the risk model was associated with immune cells (macrophages, neutrophils, and so on). These results indicated that our risk model was related to the tumor microenvironment.

Though our risk model can well predict the prognosis of COAD, there are still a few limitations. On the one hand, our risk model needs to collect data for further validation. On the other hand, some of the genes in the model need more samples for further validation and further study on their mechanism of action in COAD.

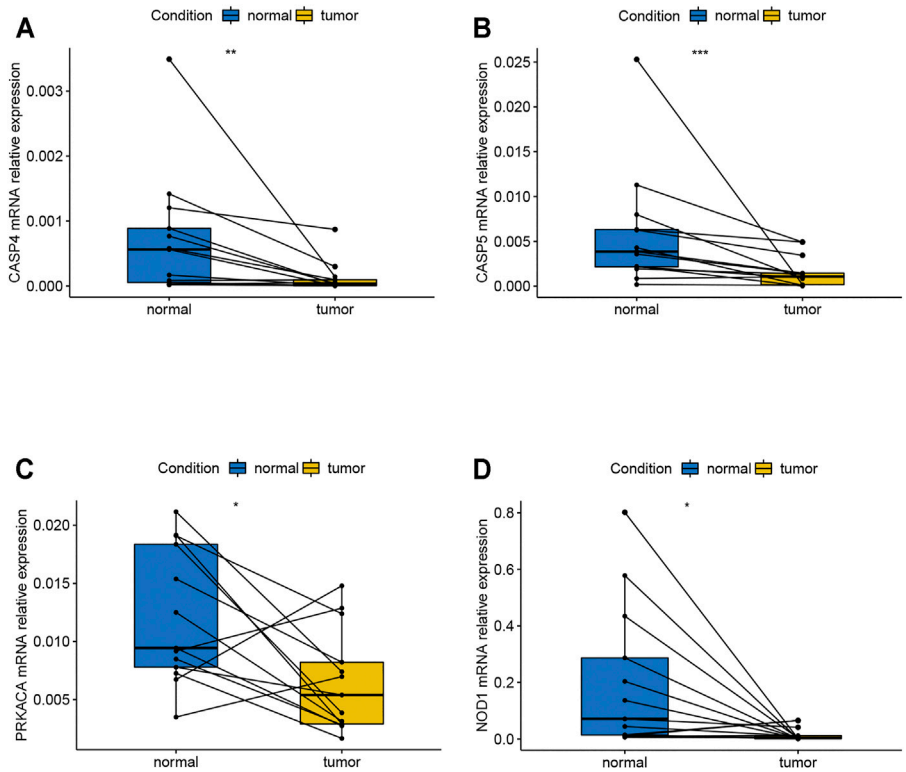

FIGURE 8 | mRNA relative expression of genes in the risk model by the method of qPCR. **(A)** mRNA relative expression of CASP4; **(B)** mRNA relative expression of CASP5; **(C)** mRNA relative expression of PRCAKA; **(D)** mRNA relative expression of NOD1. qPCR, quantitative real-time polymerase chain reaction.

CONCLUSION

In summary, our study revealed that pyroptosis-related genes showed great difference between normal and tumor tissues in COAD, and some of the genes in the risk model were validated in 13 patients with COAD. Moreover, the risk score based on seven pyroptosis-related genes was the independent factors for COAD and can well predict the prognosis. In addition, our model was more suitable for the early-stage patients, which may be regarded as the method to perform early diagnosis for tumor patients. Therefore, we thought our study can help identify patients in the early stages and may provide potentially effective new targets for the treatment of cancer patients.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

This article was written by BL. JL helped to modify the relevant R language codes and related pictures. MW and WC provided guidance to the manuscript preparation and research ideas for the writing of the manuscript. All authors have approved the final version of the editorial.

FUNDING

This research was supported by the Shanghai Municipal Science and Technology Commission (19441905400).

ACKNOWLEDGMENTS

We thank all the authors who contributed to this topic and thanks to the TCGA and GEO databases for providing data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.771847/full#supplementary-material

REFERENCES

Al Mamun, A., Mimi, A. A., Aziz, M. A., Zaeem, M., Ahmed, T., Munir, F., et al. (2021). Role of Pyroptosis in Cancer and its Therapeutic Regulation. *Eur. J. Pharmacol.* 910, 174444. doi:10.1016/j.ejphar.2021.174444

Al Mamun, A., Mimi, A. A., Zaeem, M., Wu, Y., Monalisa, I., Akter, A., et al. (2021). Role of Pyroptosis in Diabetic Retinopathy and its Therapeutic Implications. *Eur. J. Pharmacol.* 904, 174166. doi:10.1016/j.ejphar.2021.174166

Al Mamun, A., Wu, Y., Jia, C., Munir, F., Sathy, K. J., Sarker, T., et al. (2020). Role of Pyroptosis in Liver Diseases. *Int. Immunopharmacology* 84, 106489. doi:10.1016/j.intimp.2020.106489

Antonopoulos, C., Russo, H. M., El Sanadi, C., Martin, B. N., Li, X., Kaiser, W. J., et al. (2015). Caspase-8 as an Effector and Regulator of NLRP3 Inflammasome Signaling. *J. Biol. Chem.* 290 (33), 20167–20184. doi:10.1074/jbc.M115.652321

Berkel, C., and Cacan, E. (2021). Differential Expression and Copy Number Variation of Gasdermin (GSDM) Family Members, Pore-Forming Proteins in Pyroptosis, in Normal and Malignant Serous Ovarian Tissue. *Inflammation*. doi:10.1007/s10753-021-01493-0

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Broz, P., Pelegrín, P., and Shao, F. (2020). The Gasdermins, a Protein Family Executing Cell Death and Inflammation. *Nat. Rev. Immunol.* 20 (3), 143–157. doi:10.1038/s41577-019-0228-2

Chen, K. W., Demarco, B., and Broz, P. (2020). Beyond Inflammasomes: Emerging Function of Gasdermins during Apoptosis and NETosis. *EMBO J.* 39 (2), e103397. doi:10.15252/embj.2019103397

Cheng, Y.-F., Tsai, Y.-H., Huang, C.-Y., Lee, Y.-S., Chang, P.-C., Lu, Y.-C., et al. (2020). Generation and Pathological Characterization of a Transgenic Mouse Model Carrying a Missense PJVK Mutation. *Biochem. Biophysical Res. Commun.* 532 (4), 675–681. doi:10.1016/j.bbrc.2020.07.101

Chu, Q., Jiang, Y., Zhang, W., Xu, C., Du, W., Tuguzbaeva, G., et al. (2016). Pyroptosis Is Involved in the Pathogenesis of Human Hepatocellular Carcinoma. *Oncotarget* 7 (51), 84658–84665. doi:10.18632/oncotarget.12384

Deroyer, C., Rénert, A.-F., Merville, M.-P., and Fillet, M. (2014). New Role for EMD (Emerin), a Key Inner Nuclear Membrane Protein, as an Enhancer of Autophagosome Formation in the C16-Ceramide Autophagy Pathway. *Autophagy* 10 (7), 1229–1240. doi:10.4161/auto.28777

Fink, S. L., and Cookson, B. T. (2006). Caspase-1-dependent Pore Formation during Pyroptosis Leads to Osmotic Lysis of Infected Host Macrophages. *Cell Microbiol* 8 (11), 1812–1825. doi:10.1111/j.1462-5822.2006.00751.x

Fleming, M., Ravula, S., Tatishchev, S. F., and Wang, H. L. (2012). Colorectal Carcinoma: Pathologic Aspects. *J. Gastrointest. Oncol.* 3 (3), 153–173. doi:10.3978/j.issn.2078-6891.2012.030

Flood, B., Oficjalska, K., Laukens, D., Fay, J., O'Grady, A., Caiazza, F., et al. (2015). Altered Expression of Caspases-4 and -5 during Inflammatory Bowel Disease and Colorectal Cancer: Diagnostic and Therapeutic Potential. *Clin. Exp. Immunol.* 181 (1), 39–50. doi:10.1111/cei.12617

Guo, J., Zheng, J., Mu, M., Chen, Z., Xu, Z., Zhao, C., et al. (2021). GW4064 Enhances the Chemosensitivity of Colorectal Cancer to Oxaliplatin by Inducing Pyroptosis. *Biochem. Biophysical Res. Commun.* 548, 60–66. doi:10.1016/j.bbrc.2021.02.043

Hu, B., Elinav, E., Huber, S., Booth, C. J., Strowig, T., Jin, C., et al. (2010). Inflammation-induced Tumorigenesis in the colon Is Regulated by Caspase-1 and NLRC4. *Proc. Natl. Acad. Sci.* 107 (50), 21635–21640. doi:10.1073/pnas.1016814108

Hu, F., Song, D., Yan, Y., Huang, C., Shen, C., Lan, J., et al. (2021). IL-6 Regulates Autophagy and Chemotherapy Resistance by Promoting BECN1 Phosphorylation. *Nat. Commun.* 12 (1), 3651. doi:10.1038/s41467-021-23923-1

Jiang, H. Y., Najmeh, S., Martel, G., MacFadden-Murphy, E., Farias, R., Savage, P., et al. (2020). Activation of the Pattern Recognition Receptor NOD1 Augments colon Cancer Metastasis. *Protein Cell* 11 (3), 187–201. doi:10.1007/s13238-019-00687-5

Kemper, K., Rodermond, H., Colak, S., Grandela, C., and Medema, J. P. (2012). Targeting Colorectal Cancer Stem Cells with Inducible Caspase-9. *Apoptosis* 17 (5), 528–537. doi:10.1007/s10495-011-0692-z

Koncina, E., Haan, S., Rauh, S., and Letellier, E. (2020). Prognostic and Predictive Molecular Biomarkers for Colorectal Cancer: Updates and Challenges. *Cancers* 12 (2), 319. doi:10.3390/cancers12020319

Koper-Lenkiewicz, O. M., Dymicka-Piekarska, V., Milewska, A. J., Zińczuk, J., and Kamińska, J. (2021). The Relationship between Inflammation Markers (CRP, IL-6, sCD40L) and Colorectal Cancer Stage, Grade, Size and Location. *Diagnostics* 11 (8), 1382. doi:10.3390/diagnostics11081382

Liu, B., He, R., Zhang, L., Hao, B., Jiang, W., Wang, W., et al. (2021). Inflammatory Caspases Drive Pyroptosis in Acute Lung Injury. *Front. Pharmacol.* 12, 631256. doi:10.3389/fphar.2021.631256

Loveless, R., Bloomquist, R., and Teng, Y. (2021). Pyroptosis at the Forefront of Anticancer Immunity. *J. Exp. Clin. Cancer Res.* 40 (1), 264. doi:10.1186/s13046-021-02065-8

Maisonneuve, C., Tsang, D. K. L., Foerster, E. G., Robert, L. M., Mukherjee, T., Prescott, D., et al. (2021). Nod1 Promotes Colorectal Carcinogenesis by Regulating the Immunosuppressive Functions of Tumor-Infiltrating Myeloid Cells. *Cel Rep.* 34 (4), 108677. doi:10.1016/j.celrep.2020.108677

McIlwain, D. R., Berger, T., and Mak, T. W. (2013). Caspase Functions in Cell Death and Disease. *Cold Spring Harbor Perspect. Biol.* 5 (4), a008656. doi:10.1101/cshperspect.a008656

Schwartz, S., Jr., Yamamoto, H., Navarro, M., Maestro, M., Reventós, J., and Perucho, M. (1999). Frameshift Mutations at Mononucleotide Repeats in Caspase-5 and Other Target Genes in Endometrial and Gastrointestinal Cancer of the Microsatellite Mutator Phenotype. *Cancer Res.* 59 (12), 2995–3002.

Sepulveda, A. R., Hamilton, S. R., Allegra, C. J., Grody, W., Cushman-Vokoun, A. M., Funkhouser, W. K., et al. (2017). Molecular Biomarkers for the Evaluation of Colorectal Cancer. *J. Mol. Diagn.* 19 (2), 187–225. doi:10.1016/j.jmoldx.2016.11.001

Shibamoto, M., Hirata, H., Eguchi, H., Sawada, G., Sakai, N., Kajiyama, Y., et al. (2017). The Loss of CASP4 Expression Is Associated with Poor Prognosis in Esophageal Squamous Cell Carcinoma. *Oncol. Lett.* 13 (3), 1761–1766. doi:10.3892/ol.2017.5646

Tan, G., Huang, C., Chen, J., and Zhi, F. (2020). HMGB1 Released from GSDME-Mediated Pyroptotic Epithelial Cells Participates in the Tumorigenesis of Colitis-Associated Colorectal Cancer through the ERK1/2 Pathway. *J. Hematol. Oncol.* 13 (1), 149. doi:10.1186/s13045-020-00985-0

Tang, Z., Ji, L., Han, M., Xie, J., Zhong, F., Zhang, X., et al. (2020). Pyroptosis Is Involved in the Inhibitory Effect of FL118 on Growth and Metastasis in Colorectal Cancer. *Life Sci.* 257, 118065. doi:10.1016/j.lfs.2020.118065

Trojan, J., Brieger, A., Raedle, J., Weber, N., Kriener, S., Kronenberger, B., et al. (2004). BAX and Caspase-5 Frameshift Mutations and Spontaneous Apoptosis in Colorectal Cancer with Microsatellite Instability. *Int. J. Colorectal Dis.* 19 (6), 538–544. doi:10.1007/s00384-004-0597-1

Tseng, I.-C., Huang, W.-J., Jhuang, Y.-L., Chang, Y.-Y., Hsu, H.-P., and Jeng, Y.-M. (2017). Microinsertions inPRKACAcause Activation of the Protein Kinase A Pathway in Cardiac Myxoma. *J. Pathol.* 242 (2), 134–139. doi:10.1002/path.4899

Wang, L., Qin, X., Liang, J., and Ge, P. (2021). Induction of Pyroptosis: A Promising Strategy for Cancer Treatment. *Front. Oncol.* 11, 635774. doi:10.3389/fonc.2021.635774

Wang, Y., Gao, W., Shi, X., Ding, J., Liu, W., He, H., et al. (2017). Chemotherapy Drugs Induce Pyroptosis through Caspase-3 Cleavage of a Gasdermin. *Nature* 547 (7661), 99–103. doi:10.1038/nature22393

Wang, Y., Yin, B., Li, D., Wang, G., Han, X., and Sun, X. (2018). GSDME Mediates Caspase-3-dependent Pyroptosis in Gastric Cancer. *Biochem. Biophysical Res. Commun.* 495 (1), 1418–1425. doi:10.1016/j.bbrc.2017.11.156

Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA Prediction Server: Biological Network Integration for Gene Prioritization and Predicting Gene Function. *Nucleic Acids Res.* 38, W214–W220. Web Server issue. doi:10.1093/nar/gkq537

Ye, Y., Dai, Q., and Qi, H. (2021). A Novel Defined Pyroptosis-Related Gene Signature for Predicting the Prognosis of Ovarian Cancer. *Cell Death Discov.* 7 (1), 71. doi:10.1038/s41420-021-00451-x

Yu, J., Li, S., Qi, J., Chen, Z., Wu, Y., Guo, J., et al. (2019). Cleavage of GSDME by Caspase-3 Determines Lobaplatin-Induced Pyroptosis in colon Cancer Cells. *Cell Death Dis* 10 (3), 193. doi:10.1038/s41419-019-1441-4

Zhai, X., Xue, Q., Liu, Q., Guo, Y., and Chen, Z. (2017). Colon Cancer Recurrence-Associated Genes Revealed by WGCNA Co-expression Network Analysis. *Mol. Med. Rep.* 16 (5), 6499–6505. doi:10.3892/mmr.2017.7412

Zhao, Y., Wang, Y., Zhao, J., Zhang, Z., Jin, M., Zhou, F., et al. (2021). PDE2 Inhibits PKA-Mediated Phosphorylation of TFAM to Promote Mitochondrial Ca2+-Induced Colorectal Cancer Growth. *Front. Oncol.* 11, 663778. doi:10.3389/fonc.2021.663778

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 09 December 2021
doi: 10.3389/fgene.2021.746666

Genomics and Prognosis Analysis of N^6-Methyladenosine Regulators in Lung Adenocarcinoma

Yanpin Ma[1] *and Huping Zhang*[2]*

[1]Department of Oncology, The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China, [2]Department of Infectious Diseases, The First Affiliated Hospital, and College of Clinical Medicine of Henan University of Science and Technology, Luoyang, China

Objective: N^6-methyladenosine (m^6A) modification is involved in modulating various biological processes in human cancers. But the implication of m^6A modification in lung adenocarcinoma (LUAD) is still unclear. Hence, this study conducted a comprehensive analysis of the expression and clinical implication of m^6A regulators in LUAD.

Methods: Consensus clustering analysis of 502 LUAD samples in the TCGA dataset was presented based on the expression profiles of 20 m^6A regulators using ConsensusClusterPlus package. Overall survival (OS), activation of signaling pathways and tumor immunity (immune/stromal score, tumor purity, expression of HLA and immune checkpoints, and immune cell infiltration) were compared between m^6A modification patterns. The m^6A-related genes between patterns were identified and prognostic m^6A-related genes were imported into LASSO-cox regression analysis. The m^6A risk score was developed and its prognostic implication was evaluated and externally verified in the GSE30219 and GSE72094 dataset. Furthermore, a nomogram that contained independent prognostic indicators was established, followed by external verification.

Results: Two m^6A modification patterns were clustered across LUAD based on the expression similarity of the m^6A regulators via consensus clustering analysis, with distinct OS, activation of signaling pathways and tumor immunity. Totally, 213 m^6A-related genes that were identified by comparing two patterns were significantly related to LUAD prognosis. By LASSO method, we constructed the m^6A risk score that was a reliable and independent prognostic factor for LUAD. Patients with low m^6A risk score displayed a prominent survival advantage. After incorporating independent clinical features, we developed the prognostic nomogram that exhibited high predictive accuracy and the best clinical net benefit for OS.

OPEN ACCESS

Edited by:
Yanqiang Li,
Boston Children's Hospital and Harvard Medical School, United States

Reviewed by:
Xinlei Gao,
Boston Children's Hospital and Harvard Medical School, United States
Chongming Jiang,
Baylor College of Medicine, United States

***Correspondence:**
Huping Zhang
149697708@qq.com

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *24 July 2021*
Accepted: *11 October 2021*
Published: *09 December 2021*

Citation:
Ma Y and Zhang H (2021) Genomics and Prognosis Analysis of N^6-Methyladenosine Regulators in Lung Adenocarcinoma. Front. Genet. 12:746666. doi: 10.3389/fgene.2021.746666

Abbreviations: NSCLC, non-small cell lung cancer; LUAD, lung adenocarcinoma; m^6A, N^6-methyladenosine; TCGA, The Cancer Genome Atlas; GEO, Gene Expression Omnibus; GSVA, Gene set variation analysis; ESTIMATE, Estimation of Stromal and Immune Cells in Malignant Tumors Using Expression Data; HLA, human leukocyte antigen; ssGSEA, single-sample gene set enrichment analysis; DEGs, differentially expressed genes; FC, fold change; FDR, false discovery rate; LASSO, least absolute shrinkage and selection operator; OS, overall survival; ROC, receiver operating characteristic; AUC, area under curv; HR, hazard ratio.

Conclusion: Collectively, our study may provide a clinically useful tool for precise prognostic management and optimization of immunotherapeutic strategies for LUAD patients.

Keywords: lung adenocarcinoma, N6-methyladenosine, molecular subtypes, risk score, prognosis, tumor immunity

INTRODUCTION

Lung cancer has the high incidence and mortality globally, occupying almost 20% of cancer-related deaths in 2018 (Bray et al., 2018). It was estimated that there were 2.1 million new lung cancer cases and 1.8 million deaths in 2018 (Bray et al., 2018). This disease mainly includes two histological subtypes: non-small cell lung cancer (NSCLC; 85%) and small cell lung cancer (SCLC). NSCLC contains lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (Zhu et al., 2019). LUAD is the main histology, and its incidence is constantly on the rise. Conventional therapeutic options against NSCLC include surgery resection, chemotherapy, and radiotherapy. Despite the progress in combined and personalized therapies such as tyrosine kinase inhibitors and immunotherapies (PD1/PD-L1 inhibitors), the 5-year survival rate is only 16% (Zhang C. et al., 2020). Diagnosis of LUAD usually occurs at an advanced stage, and most patients experience badly toxic treatment and poor clinical benefit (Schmidt et al., 2019). Hence, it is of importance to explore specific prognostic models for predicting patients' survival, which can assist design appropriate therapeutic strategies and management choice for distinct LUAD subgroups.

N^6-methyladenosine (m^6A) is the most abundant type of RNA post-transcriptional modification in eukaryotes, which plays a key role in a variety of biological processes by regulating the translation, splicing, stabilization, and degradation of mRNAs (Zhang H. et al., 2020). Typically, m^6A regulators contain three types: writers (including VIRMA, METTL14, METTL3, RBM15, RBM15B, RBMX, WTAP, and ZC3H13), erasers (including ALKBH5 and FTO) and readers (including HNRNPA2B1, HNRNPC, IGF2BP1, IGF2BP2, IGF2BP3, YTHDC1, YTHDC2, YTHDF1, YTHDF2, and YTHDF3) (Fu et al., 2014). Emerging evidence suggests that the 20 m^6A regulators display tight relationships with LUAD (Li F. et al., 2020; Chao et al., 2020; Li Y. et al., 2020). For instance, YTHDC2 suppresses LUAD carcinogenesis through inhibiting SLC7A11-dependent antioxidant function (Ma et al., 2021). Moreover, FTO accelerates LUAD progression through mRNA demethylation (Ding et al., 2020). ALKBH5 facilitates proliferation and invasion of LUAD cells following intermittent hypoxia (Chao et al., 2020). YTHDF1 is linked to hypoxia adaptation and LUAD progression (Shi et al., 2019). FTO triggers LUAD progression through activating cell migration via mRNA demethylation (Ding et al., 2020). These experimental evidences suggest that an in-depth understanding of m^6A regulators may deepen our understanding on the role of m^6A modification in the progression of LUAD. Here, we comprehensively analyzed the expression and clinical implication of m^6A regulators in LUAD.

MATERIALS AND METHODS

Dataset Preparation

The Cancer Genome Atlas database (TCGA) RNA-seq data (FPKM values) and matched clinical features of 502 LUAD patients were retrieved from the Genomic Data Commons website (https://portal.gdc.cancer.gov/). The FPKM values were normalized with transcripts per million (TPM) method, followed by log2 conversion. Microarray expression profiling and clinical information of 274 LUAD samples and 398 LUAD samples were separately obtained from the GSE30219 dataset (Rousseaux et al., 2013) and the GSE72094 dataset (Schabath et al., 2016) in the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). All data were obtained from the publicly available databases. Therefore, it was not applicable for the ethical approval. A total of 20 m^6A regulators including 8 writers (VIRMA, METTL14, METTL3, RBM15, RBM15B, RBMX, WTAP, and ZC3H13), 2 erasers (ALKBH5 and FTO), and 10 readers (HNRNPA2B1, HNRNPC, IGF2BP1, IGF2BP2, IGF2BP3, YTHDC1, YTHDC2, YTHDF1, YTHDF2, and YTHDF3) were collected from the published literature. The location of these m^6A regulators on the human chromosomes was plotted through Rcircos package (version 1.2.1) (Zhang et al., 2013). Protein-protein interaction analysis of the m^6A regulators was performed by the STRING online database (version: 11.0; https://string-db.org/) (Szklarczyk et al., 2017).

Consensus Clustering Analysis

Consensus clustering analysis was carried out utilizing ConsensusClusterPlus package (version 1.48.0) to assign LUAD patients in the TCGA dataset into different m^6A modification patterns with 50 iterations and resample rate of 80% based on the expression matrix of the 20 m^6A regulators (Wilkerson and Hayes, 2010). Kaplan-Meier curves of overall survival (OS) were conducted between two m^6A modification patterns. The survival difference was compared with log-rank test. The t-distributed stochastic neighbor embedding (t-SNE) was presented to validate the accuracy of this classification.

Gene Set Variation Analysis

The activation of pathways was quantified in each LUAD sample from the TCGA dataset by single-sample gene set enrichment analysis (ssGSEA) method derived from GSVA package (version 1.32.0) in an unsupervised manner (Hänzelmann et al., 2013). The gene set of "c2. cp.kegg.v7.2. symbols" was obtained from the Molecular Signatures Database, which was used as the reference set (Liberzon et al., 2015).

Estimation of Tumor Immunity

According to the normalized expression matrix, stromal and immune scores across LUAD samples in the TCGA dataset were estimated via the Estimation of Stromal and Immune Cells in Malignant Tumors Using Expression Data (ESTIMATE) method (https://sourceforge.net/projects/estimateproject/). (Yoshihara et al., 2013) that was applied for inferring the overall infiltrations of stromal and immune cells in LUAD tissues based on gene symbols. The tumor purity was calculated via ESTIMATE and consensus measurement of purity estimations methods. Tumor immune signatures were assessed in LUAD samples, including the mRNA expression of human leukocyte antigen (HLA) family genes and immune checkpoints. The infiltration levels of immune cells were quantified across LUAD samples based on the published gene signatures utilizing the ssGSEA algorithm (Charoentong et al., 2017; Jia et al., 2018).

Identification of m^{6}A-Related Differentially Expressed Genes

The DEGs were screened between two m^{6}A modification patterns in the TCGA dataset through limma package (version 3.40.6) (Ritchie et al., 2015). The cut-off was |log2 fold change (FC)|>1 and false discovery rate (FDR) < 0.001. FDR was calculated with Benjamin–Hochberg method. The m^{6}A-related DEGs were visualized into volcano and heat maps via pheatmap package (version 1.0.12).

Construction of a Least Absolute Shrinkage and Selection Operator-Cox Regression Model

LASSO represents a regularization and descending dimension method that has been applied for prognostic Cox models. Univariate-cox regression analysis was utilized to assess the correlation between overall survival (OS) of LUAD patients in the TCGA dataset and the m^{6}A-related DEGs. The genes with $p < 0.05$ were input into the LASSO-cox regression model through glmnet package (version 2.0.16) (Engebretsen and Bohlin, 2019). Variable selection was presented for penalizing the data fitting criteria, which reduced the complexity and made the model more interpretable. The coefficient of each variable was the average estimate of the coefficient obtained from 10-fold cross-verification. The m^{6}A risk score was developed following the formula: risk score = $\sum_{i=1}^{n} Coef\ (i)X\ (i)$, where n indicated the number of variables in this LASSO model, Coef 1) represented the regression coefficient, and X 1) meant the mRNA expression levels of variables in LUAD samples. To evaluate the prediction utility of the LASSO model, time-dependent receiver operating characteristic (ROC) curves were conducted by survivalROC package (version 1.0.3) in the TCGA, GSE30219 and GSE72094 datasets, followed by calculation of one, three and 5-year area under curve (AUC). In the two datasets, LUAD patients were separately split into two groups according to the median m^{6}A risk score through survminer (version 0.4.9) and survival (version 3.2–13) packages. Kaplan-Meier curves of OS were depicted for two groups via survival package and OS difference was compared with log-rank test. The distribution of survival status in two groups was then visualized. By pheatmap package, heatmap was established to visualize the mRNA expression pattern of each variable in the LASSO model.

Estimation of the Prediction Independency of the m^{6}A Risk Score

To estimate whether the m^{6}A risk score independently predicted LUAD patients' OS, univariate- and multivariate-cox regression analysis was carried out following adjusting clinical features (gender, stage, T, N and M) in the TCGA, GSE30219 and GSE72094 datasets. Hazard ratio (HR) and p value were calculated for each variable.

Construction of a Nomogram Model

To better apply the m^{6}A risk score in clinical practice, a nomogram that included independent prognostic indicators was conducted to predict LUAD patients' one, three and 5-year OS in the TCGA, GSE30219 and GSE72094 datasets via rms package (version 6.2–0). Calibration plot was presented to evaluate predictive performance of the m^{6}A risk score. Furthermore, decision curve analysis was carried out for calculating the clinical net benefit of every model in comparison to all or none strategies. The none plots indicated the assumption that no subjects had one, three or 5-year OS. Meanwhile, all plots indicated the assumption that each subject had one, three or 5-year OS at specific threshold probabilities. The best model was the one with the highest net benefit.

Statistical Analysis

All statistical analysis was implemented through the R software (version 3.6.3). Wilcoxon test was used for comparison between two groups. $p < 0.05$ was statistically significant.

RESULTS

Landscape of Expression and Prognostic Implications of m^{6}A Regulators in Lung Adenocarcinoma

Totally, 20 m^{6}A regulators including 8 writers (VIRMA, METTL14, METTL3, RBM15, RBM15B, RBMX, WTAP, and ZC3H13), 2 erasers (ALKBH5 and FTO), and 13 readers (HNRNPA2B1, HNRNPC, IGF2BP1, IGF2BP2, IGF2BP3, YTHDC1, YTHDC2, YTHDF1, YTHDF2, and YTHDF3) were collected in this study. **Figure 1A** and **Supplementary Table S1** depicted their location on the chromosomes. Also, there were close direct (physical) or indirect (functional) interactions between the m^{6}A regulators (**Figure 1A**). Pan-cancer analysis revealed the prognostic implication of the m^{6}A regulators in the TCGA cohort (**Figure 1B**). For LUAD, IGF2BP1, IGF2BP2, IGF2BP3, HNRNPA2B1, HNRNPC, VIRMA, RBM15 and ALKBH5 were significant risk factors. The mRNA expression of the 20 m^{6}A regulators was compared between LUAD and normal tissues.

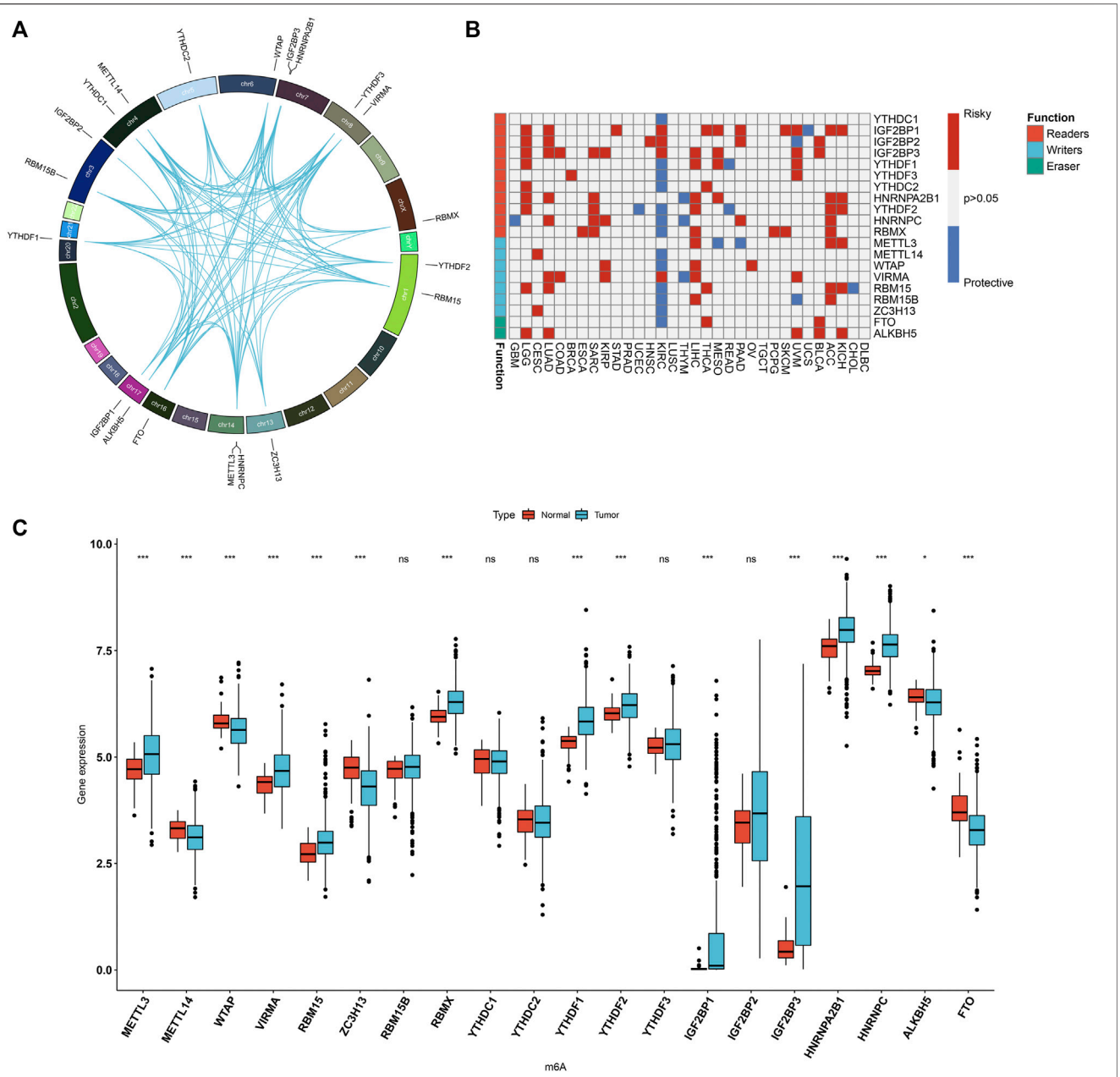

FIGURE 1 | Landscape of expression patterns and prognostic value of 20 m^{6}A regulators in LUAD. **(A)** The location of 20 m^{6}A regulators on the human chromosomes. Lines in circle represented their interactions according to the STRING database. **(B)** Heatmap showing the correlations between m^{6}A regulators and prognosis of pan-cancer in TCGA cohorts. The left side of the heatmap showed the functions (writer, eraser and reader) of m^{6}A regulators. Red indicated that a specific regulator was a risk factor for a type of cancer and blue indicated that a specific regulator was a protective factor for a type of cancer. **(C)** Comparison of the expression of m^{6}A regulators between LUAD and normal tissues in the TCGA dataset with Wilcoxon test. Ns: not significant; *p < 0.05; ***p < 0.001.

We found that METTL3, VIRMA, RBM15, RBMX, YTHDF1, YTHDF2, IGF2BP1, IGF2BP3, HNRNPA2B1 and HNRNPC displayed higher mRNA expression in LUAD compared to normal specimens (**Figure 1C**). Meanwhile, METTL14, WTAP, ZC3H13, ALKBH5 and FTO were significantly down-regulated in LUAD than normal tissues. These data were indicative of the important implication of m^{6}A regulators in the progression of LUAD.

Construction of m^{6}A Regulators-Mediated m^{6}A Modification Patterns in Lung Adenocarcinoma

A total of 502 LUAD samples were clustered based on the expression similarity of the m^{6}A regulators via consensus clustering analysis. Our data found that when the number of groups (k) = 2, there was an excellent clustering among LUAD samples in the consensus matrix (**Figure 2A**). Consensus

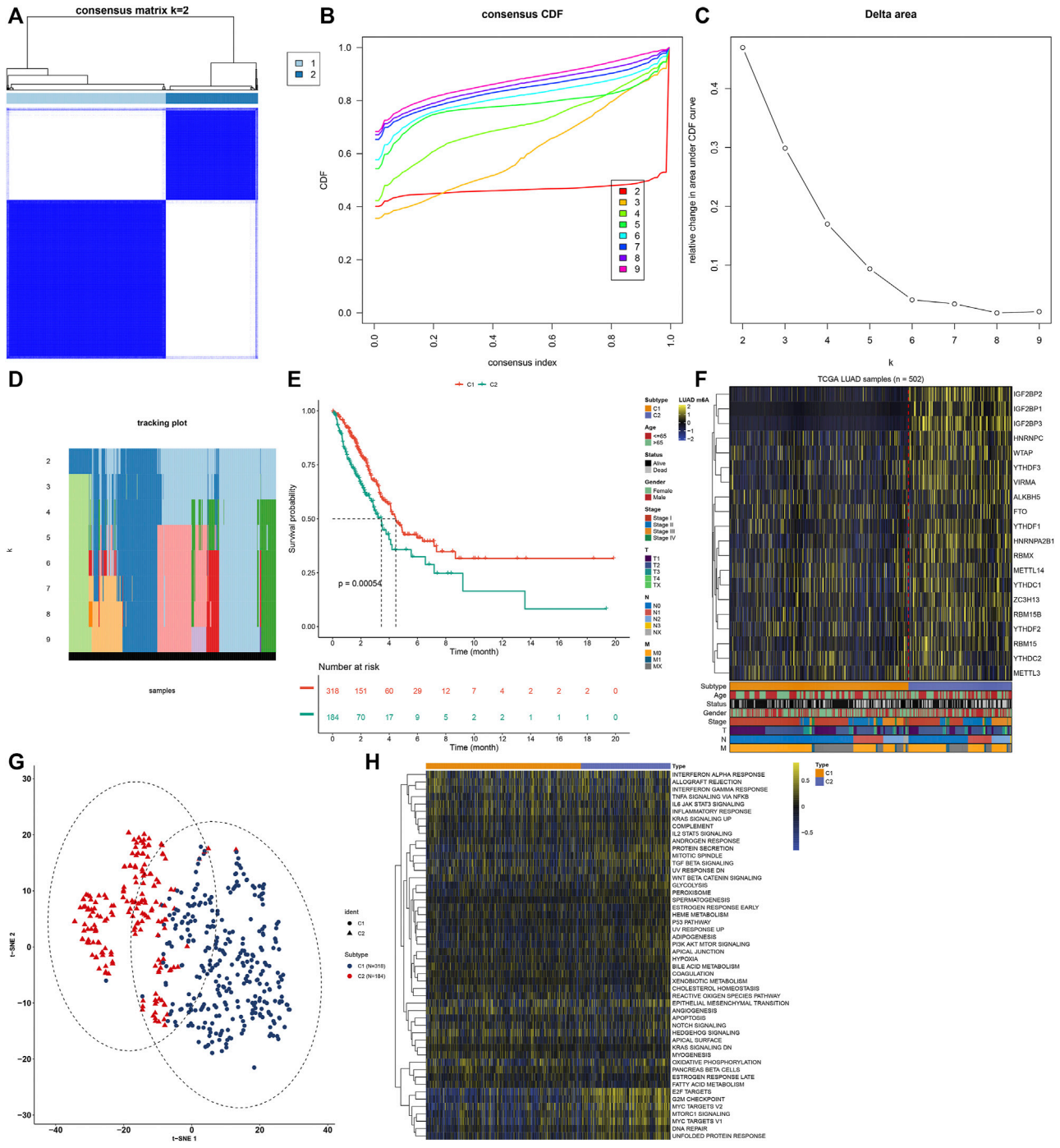

FIGURE 2 | Construction of m^{6}A regulator-mediated m^{6}A modification patterns in LUAD from the TCGA cohort. **(A)** Consensus matrix when number of groups (k) = 2. In the consensus matrix, white meant that samples were impossibly clustered together, and dark blue meant that samples were always clustered together. Both rows and columns of the matrix represented samples. **(B)** Consensus cumulative distribution function (CDF) diagram when different k values. **(C)** Delta area plot for relative change in the area under CDF curve for k compared to k-1. **(D)** Tracking plot for sample clustering when different k values. **(E)** Kaplan-Meier curves of OS between two m^{6}A modification patterns. The survival probabilities were compared with log-rank test. **(F)** Heatmap for the expression patterns of 20 m^{6}A regulators in different m^{6}A modification patterns, age, survival status, gender, stage, T, N and M. **(G)** The t-SNE plots for verifying the differences between two m^{6}A modification patterns. **(H)** GSVA for the activation of signaling pathways in LUAD samples between two m^{6}A modification patterns.

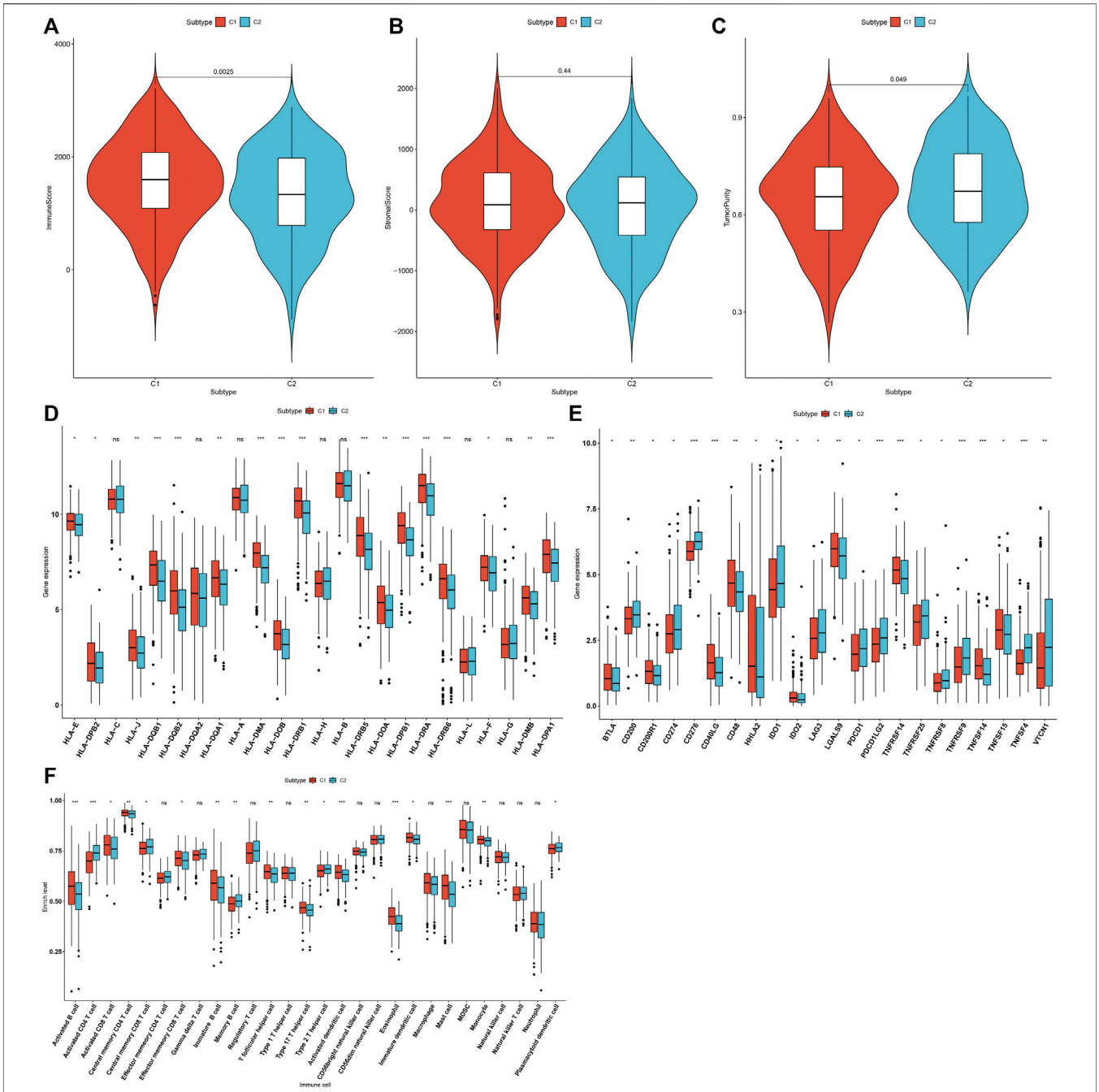

FIGURE 3 | Two m^{6}A modification patterns characterized by distinct tumor immunity across 502 LUAD specimens from the TCGA cohort. **(A–C)** Quantification of immune score, stromal score, and tumor purity in LUAD samples from two m^{6}A modification patterns through ESTIMATE algorithm. **(D)** Comparison of the mRNA expression of HLA genes between two m^{6}A modification patterns. **(E)** Comparison of the mRNA expression of immune checkpoints between two m^{6}A modification patterns. **(F)** Quantification of the infiltration levels of immune cells in LUAD samples from two m^{6}A modification patterns via ssGSEA algorithm. Ns: not significant; $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

cumulative distribution function (CDF) diagram showed that when k value = 2, CDF reached an approximate maximum (**Figure 2B**). Delta area plot depicted the relative change in the area under CDF curve for k compared to k-1 (**Figure 2C**). As shown in tracking plot, when k value = 2, sample classification was stable (**Figure 2D**). Hence, we clustered LUAD patients into two m^{6}A modification patterns, named as C1 (N = 318) and C2 (N = 184). To further understand the characteristics of m^{6}A modification patterns clustered by consensus clustering analysis in LUAD, we firstly analyzed the difference in OS. The data showed that C2 exhibited a more unfavorable OS in comparison to C1 (p = 0.00054; **Figure 2E**). Furthermore, we visualized the

expression patterns of the m^{6}A regulators in two m^{6}A modification patterns. As shown in **Figure 2F**, IGF2BP2, IGF2BP1, IGF2BP3 had distinctly higher expression in C2 compared to C1. The t-SNE was carried out for verifying whether the categories were appropriate. Our results showed that most of samples from C1 and C2 were separately gathered (**Figure 2G**), indicating that the clustering of two m^{6}A modification patterns was a relatively good choice. By applying GSVA algorithm, activation of several signaling pathways was quantified in each LUAD sample. We found that E2F targets, G2M checkpoint, MYC targets, mTORC1 signaling, DNA repair and unfolded protein response had higher activations in C2 than C1 (**Figure 2H**).

Two m^{6}A Modification Patterns Characterized by Distinct Tumor Immunity

The overall infiltration levels of immune and stromal cells were estimated in 502 LUAD samples from the TCGA dataset via ESTIMATE algorithm. Compared to C1, C2 pattern had a significantly decreased immune score (p = 0.0025; **Figure 3A**). But no significant difference in stromal score was detected between m^{6}A modification patterns (**Figure 3B**). There was significantly increased tumor purity in C2 than C1 (p = 0.049; **Figure 3C**). The mRNA expression of HLA genes and immune checkpoints was compared between m^{6}A modification patterns. Most of HLAs were highly expressed in C1 compared to C2, including HLA-E, HLA-DPB2, HLA-J, HLA-DQB1, HLA-DQB2, HLA-DQA1, HLA-DMA, HLA-DOB, HLA-DRB1, HLA-DRB5, HLA-DOA, HLA-DPB1, HLA-DRA, HLA-DRB6, HLA-F, HLA-DMB and HLA-DPA1 (**Figure 3D**). We also evaluated the differences in mRNA expression of common immune checkpoints between two m^{6}A modification patterns. Our results showed that BTLA, CD200R1, CD40LG, CD48, HHLA2, IDO2, LGALS9, TNFRSF14, TNFSF14 and TNFSF15 displayed higher mRNA expression in C1 compared to C2 (**Figure 3E**). Meanwhile, C2 pattern had increased mRNA expression of CD200, CD274, CD276, IDO1, LAG3, PDCD1, PDCD1LG2, TNFRSF25, TNFRSF8, TNFRSF9, TNFSF4 and VTCN1 in comparison to C1. The infiltration levels of immune cells were quantified in each LUAD specimen via ssGSEA algorithm. Compared to C2, there were increased infiltration levels of activated B cell, activated CD8 T cell, central memory CD4 T cell, effector memory CD8 T cell, immature B cell, T follicular helper cell, type 17 T helper cell, activated dendritic cell, eosinophil, immature dendritic cell, mast cell and monocyte in C1 (**Figure 3F**). The higher infiltration levels of activated CD4 T cell, central memory CD8 T cell, memory B cell, type 2 T helper cell and plasmacytoid dendritic cell were found in C2 than C1.

Identification of DEGs Between Two m^{6}A Modification Patterns

To explore the molecular mechanisms underlying two m^{6}A modification patterns, we presented differential expression analysis. With the cutoff of |log2FC|>1 and adjusted p < 0.001, a total of 297 genes exhibited abnormal expression between two m^{6}A modification patterns (**Figures 4A,B**). Among them, 111 genes were down-regulated and 186 were up-regulated in C1 compared to C2 (**Supplementary Table S2**). These genes could be affected by m^{6}A methylation modification in LUAD.

Development of an m^{6}A Risk Score for Lung Adenocarcinoma

Prognostic implications of the 297 m^{6}A-related DEGs were assessed via univariate-cox regression analysis. As a result, 213 genes had significant correlations to LUAD prognosis in the TCGA dataset (**Supplementary Table S3**). Candidate prognostic m^{6}A-related DEGs were further screened with LASSO-Cox regression analysis (**Figures 5A,B**). As a result, 12 candidate m^{6}A-related DEGs were identified for constructing a LASSO-cox regression model. Non-zero coefficients and the expression of 12 m^{6}A-related DEGs in this LASSO-cox regression model were calculated in the TCGA dataset. The m^{6}A risk score formula was as follows: m^{6}A risk score = 0.0301860610758127 * mRNA expression of ANLN + 0.0263443393604996 * mRNA expression of PLK1 + 0.0576834829109261 * mRNA expression of IGF2BP1 + 0.0236368688862302 * mRNA expression of HMMR + 0.0356239951044486 * mRNA expression of NEIL3 + (-0.00157287764972551) * mRNA expression of SFTA3 + (-0.0250630515726943) * mRNA expression of CXCL17 + (-0.0244060686771379) * mRNA expression of IRX5 + 0.0277060677471147 * mRNA expression of PKP2 + 0.0281636442957098 * mRNA expression of LYPD3 + 0.00917559407956536 * mRNA expression of ABCC2 + 0.157371504648263 * mRNA expression of DKK1. ROC curves were conducted to evaluate whether the m^{6}A risk score accurately and sensitively estimated the survival likelihood of LUAD patients in the TCGA dataset. The AUC values of one, three and 5-year OS were separately 0.751, 0.690 and 0.611 (**Figure 5C**). These indicated the good predictive performance of the m^{6}A risk score. **Figure 5D** depicted the distribution of the m^{6}A risk score across LUAD patients. According to the median m^{6}A risk score, patients were split into high- and low-m^{6}A risk score groups. The OS difference was compared between groups. As shown in **Figure 5E**, low m^{6}A risk score group displayed a potential survival advantage in comparison to high m^{6}A risk score group (p < 0.0001). The distribution of survival status was visualized in **Figure 5F**. We found that high m^{6}A risk score group had the relatively increased number of dead status than low m^{6}A risk score group. **Figure 5G** showed the mRNA expression of 12 variables in the model between high- and low-m^{6}A risk score groups. DKK1, PKP2, LYPD3, NEIL3, HMMR, ANLN, PLK1, IGF2BP1 and ABCC2 displayed higher mRNA expression in high-m^{6}A risk score group compared to low-m^{6}A risk score group.

External Verification of the m^{6}A Risk Score in Lung Adenocarcinoma Prognosis

To externally verify the prognostic implication of the m^{6}A risk score, we acquired the transcriptome data and follow-up information of 274 LUAD patients from the GSE30219 cohort.

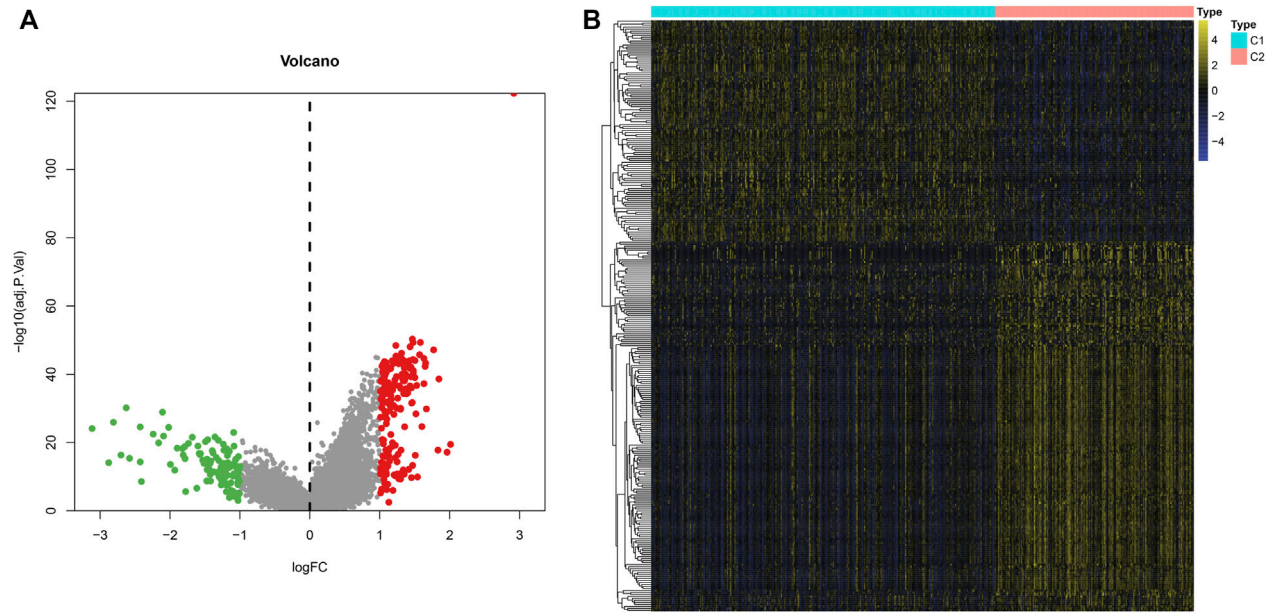

FIGURE 4 | Identification of DEGs between two m^{6}A modification patterns across LUAD samples from the TCGA dataset. **(A)** The volcano map of 297 genes that were abnormally expressed between two m^{6}A modification patterns. Red dots indicated genes with high expression in C1 and green dots represented genes with low expression in C1. **(B)** Hierarchical clustering analysis of the 297 m^{6}A-related DEGs in LUAD samples from two m^{6}A modification patterns.

The AUC values under one, three and 5-year OS were separately 0.663, 0.677 and 0.694 (**Figure 6A**). According to the median m^{6}A risk score, we clustered these LUAD patients into two groups (**Figure 6B**). High m^{6}A risk score group had more patients with dead status than low m^{6}A risk score group (**Figure 6C**). Consistently, high m^{6}A risk score was markedly associated with worse prognosis of LUAD in comparison to low m^{6}A risk score ($p < 0.0001$; **Figure 6D**). Furthermore, DKK1, PKP2, LYPD3, NEIL3, HMMR, ANLN, PLK1, IGF2BP1, and ABCC2 were highly expressed in high m^{6}A risk score group than low m^{6}A risk score group (**Figure 6E**). We also verified the prognostic significance of the m^{6}A risk score in the GSE72094 dataset. The AUC values at one, three and 5-year OS were 0.699, 0.635, and 0.663 (**Figure 6F**). As expected, high m^{6}A risk score indicated poorer OS than low m^{6}A risk score (**Figure 6G**). Therefore, the m^{6}A risk score possessed the potential in accurately predicting survival outcomes of LUAD patients.

The m^{6}A Risk Score Acts as an Independent Prognostic Indicator of Lung Adenocarcinoma

In the TCGA dataset, univariate-cox regression analysis showed that the m^{6}A risk score ($p < 0.001$; HR: 5.227 (3.347–8.163)), stage ($p < 0.001$; HR: 1.674 (1.458–1.923)), T ($p < 0.001$; HR: 1.530 (1.271–1.843)), N ($p < 0.001$; HR: 1.705 (1.437–2.023)) and M ($p = 0.007$; HR: 2.106 (1.229–3.609)) were significantly associated with LUAD prognosis (**Figure 7A**). These prognostic factors were input into multivariate-cox regression analysis. In **Figure 7B**, m^{6}A risk score [$p < 0.001$; HR: 4.373 (2.618–7.306)] and stage [$p = 0.029$; HR: 1.533 (1.046–2.246)] were independent prognostic factors of LUAD. The prognostic implication of the m^{6}A risk score was externally verified in the GSE30219 and GSE72094 cohorts. Our results confirmed that the m^{6}A risk score could independently predict LUAD prognosis both in the GSE30219 cohort (**Figures 7C,D**) and GSE72094 cohort (**Figures 7E,F**).

Establishment and Verification of a Prognostic Nomogram for Lung Adenocarcinoma Patients

A nomogram was built for predicting one, three and 5-year OS likelihood of LUAD patients in the TCGA dataset, including the m^{6}A risk score and stage that were obtained from multivariate-cox regression analysis (**Figure 8A**). Calibration diagram demonstrated that there was a high consistency in this nomogram-predicted and actual one, three and 5-year OS probabilities (**Figures 8B–D**). Moreover, decision curves suggested that the nomogram showed the best clinical net benefit for one, three and 5-year OS (**Figures 8E–G**). The nomogram was externally verified in the GSE30219 cohort (**Supplementary Figure S1A–G**) and GSE72094 cohort (**Supplementary Figure S2A–G**).

DISCUSSION

LUAD is usually diagnosed at an advanced stage, characterized by a high mortality (Zhu et al., 2020). The development of more effective therapeutic strategies requires an in-depth understanding of factors that impact the initiation and

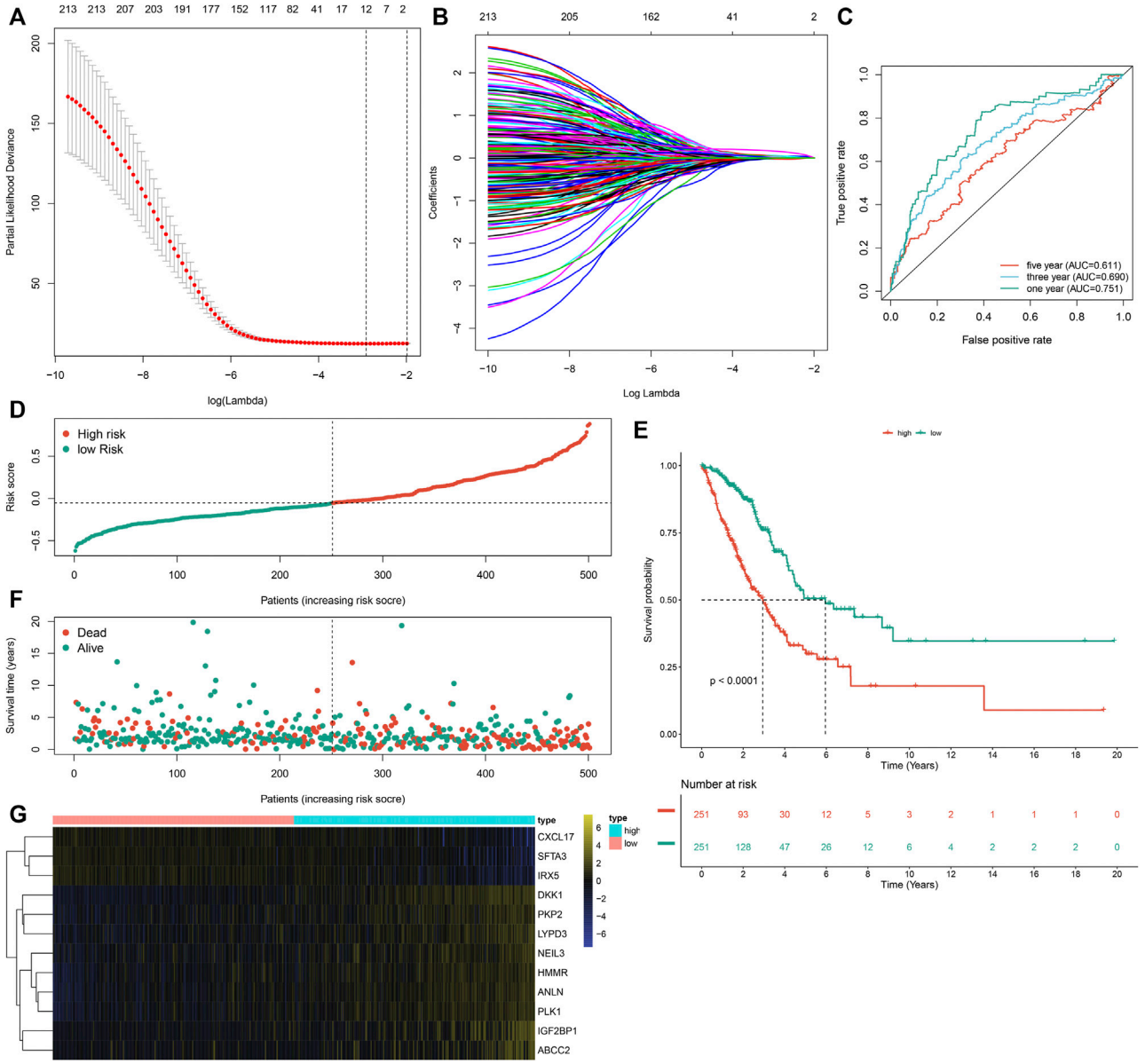

FIGURE 5 | Development of an m^{6}A risk score for LUAD by LASSO-cox regression analysis in LUAD samples from the TCGA cohort. **(A)** Partial likelihood deviance in each lambda value for the LASSO-cox regression analysis. **(B)** Process of variable selection in the LASSO-cox regression model by 10-fold cross-verification. **(C)** ROC curves under one, three and 5-year OS based on the m^{6}A risk score. **(D)** The distribution of the m^{6}A risk score and identification of the median m^{6}A risk score indicated by vertical dotted line. **(E)** Kapan-Meier curves of OS for high- and low-m^{6}A risk score groups, followed by log-rank test. **(F)** The distribution of survival status in high- and low-m^{6}A risk score groups. Red dots indicated dead status and blue dots indicated alive status. **(G)** Hierarchical clustering analysis for the mRNA expression patterns in LUAD samples from high- and low-m^{6}A risk score groups.

progression of LUAD. Recently, several studies have reported the key implications of m^{6}A regulators during LUAD development (Li F. et al., 2020; Chao et al., 2020; Li Y. et al., 2020). For example, Zhuang et al. constructed a diagnostic score model and a prognostic model for LUAD based on m^{6}A regulators (Zhuang et al., 2020). Zhou et al. characterized two molecular subtypes with diverse prognosis and tumor microenvironment in LUAD based on m^{6}A RNA methylation modification (Zhou et al., 2021). Wu et al. developed a five-m^{6}A regulatory gene signature as a prognostic biomarker in LUAD patients (Wu X. et al., 2021). Xu et al. proposed m^{6}A-related lncRNAs as potential biomarkers for prediction of prognosis and immune response in patients with LUAD (Xu et al., 2021). However, more studies should be presented for investigating the biological significance of m^{6}A regulators in LUAD progression and prognosis. Hence, this study systematically analyzed the abnormal expression and clinical implications of m^{6}A regulators.

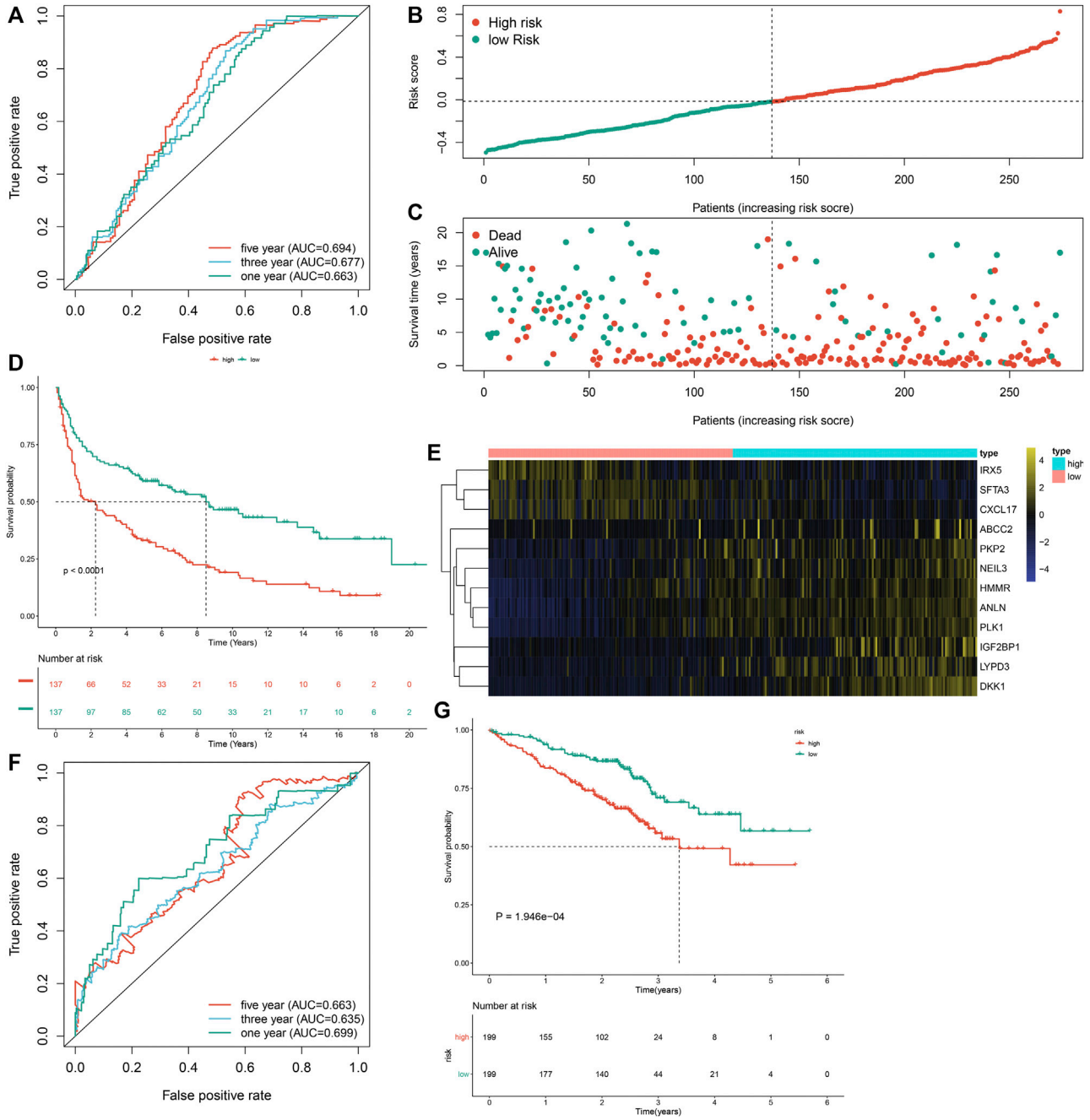

FIGURE 6 | External verification of the m^{6}A risk score for LUAD samples in the GSE30219 and GSE72094 datasets. **(A)** ROC curves for estimating the efficacy of the m^{6}A risk score in prediction of one, three and 5-year OS in the GSE30219 dataset. **(B)** The distribution of m^{6}A risk score and identification of the median m^{6}A risk score indicated by vertical dotted line in the GSE30219 dataset. **(C)** The distribution of survival status in high- and low-m^{6}A risk score groups. Red dots indicated dead status and blue dots indicated alive status in the GSE30219 dataset. **(D)** Kapan-Meier curves of OS for high- and low-m^{6}A risk score groups, followed by log-rank test in the GSE30219 dataset. **(E)** Hierarchical clustering analysis for the mRNA expression patterns in LUAD samples from high- and low-m^{6}A risk score groups in the GSE30219 dataset. **(F)** ROC curves for assessing the efficacy of the m^{6}A risk score in prediction of one, three and 5-year OS in the GSE72094 dataset. **(G)** Kapan-Meier curves of OS for high- and low-m^{6}A risk score groups, followed by log-rank test in the GSE72094 dataset.

By consensus clustering analysis, we established two m^{6}A modification patterns with distinct survival outcomes based on the expression matrix of 20 m^{6}A regulators for LUAD (**Figure 2**). Immune evasion represents a "hallmark of cancer," which reflects that immune effector constitutes a key determinant in the tumor microenvironment (Liu J. et al., 2020). Exploring the interactions between tumors and corresponding immune cells may reveal powerful and novel treatment options against LUAD.

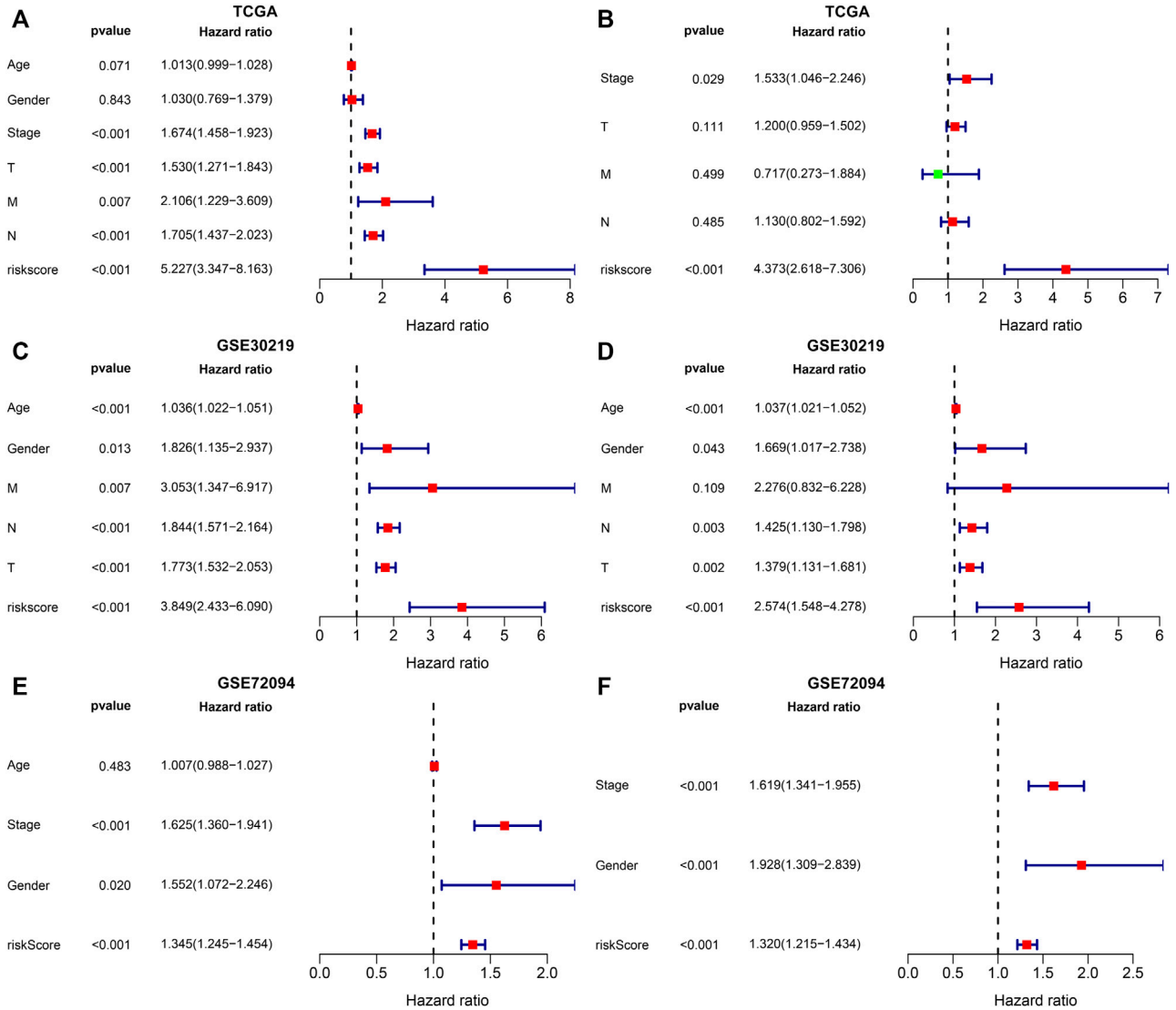

FIGURE 7 | Evaluation and external verification of the predictive independency of the m^{6}A risk score in LUAD patients. **(A)** Univariate-cox regression analysis for the correlations between the m^{6}A risk score and LUAD patients' OS in the TCGA dataset. **(B)** Multivariate-cox regression analysis for assessing the independent prognostic factors in the TCGA dataset. **(C)** External validation of the correlations between the m^{6}A risk score and LUAD patients' OS in the GSE30219 dataset by univariate-cox regression analysis. **(D)** External validation of the independent prognostic factors in the GSE30219 dataset through multivariate-cox regression analysis. **(E)** External verification of the relationships of the m^{6}A risk score with LUAD patients' OS in the GSE72094 dataset through univariate-cox regression analysis. **(F)** External verification of the independent prognostic factors in the GSE72094 dataset through multivariate-cox regression analysis.

Immunotherapies like PD1/PD-L1 inhibitors have become the standard-of-care therapeutic strategy against NSCLC (Cui et al., 2019). But, only 20–30% patients respond to such therapy (Rittmeyer et al., 2017). Limited data are available concerning the interactions between markers and immunotherapy response. Hence, it is necessary to explore and identify effective tumor immunity-related markers for LUAD, thereby reducing the mortality and developing innovative targeted therapeutic options. Compared to C2, C1 had an increased immune score and most of HLAs were highly expressed in C1, indicating that patients in C1 pattern exhibited higher tumor immunity (**Figure 3**).

Much progress in genome-wide methods like RNA-seq and microarrays has accelerated the evolution of cancer biomarker-related research. Numerous genetic markers of LUAD have been discovered, which are significantly correlated to diagnosis, survival outcomes, and drug resistance. But most of studies are limited to a single marker or a small sample population, leading to the limited accuracy and availability of markers. Furthermore, due to tumor heterogeneity, conventional clinical parameters like TNM staging are difficult to meet the requirement of accuracy and individuation in predicting prognosis. Thus, combined multiple markers or large sample analysis are necessary. Here, we established an m^{6}A risk score

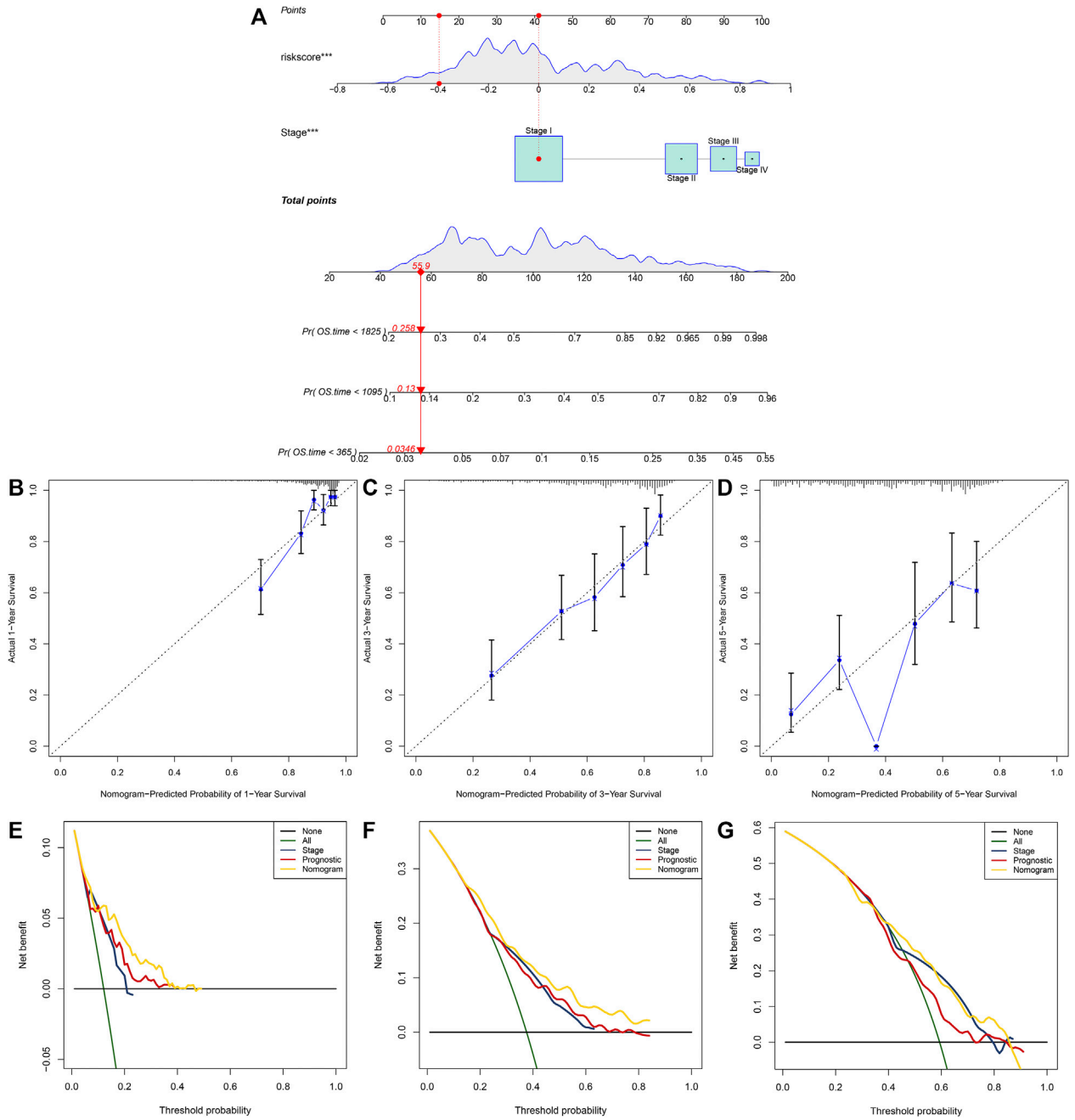

FIGURE 8 | Establishment of the nomogram for predicting survival likelihood of LUAD patients in the TCGA dataset. **(A)** The nomogram that included independent prognostic indicators obtained from multivariate-cox regression analysis for predicting one, three and 5-year OS likelihood of LUAD patients. Each variable (m^6A risk score and stage) was projected to the point scale for obtaining a value. **(B–D)** Calibration plots of the nomogram-predicted one, three and 5-year OS probabilities against the observed rates. The *x*-axis represented nomogram-predicted survival and the *y*-axis represented actual survival. **(E–G)** Decision curves for the clinical net benefit of each model in comparison to all or none strategies. The *x*-axis indicated the threshold probability, and the *y*-axis indicated the net clinical benefit.

containing ANLN, PLK1, IGF2BP1, HMMR, NEIL3, SFTA3, CXCL17, IRX5, PKP2, LYPD3, ABCC2, and DKK1 for LUAD prognosis (**Figure 5**). High m^6A risk score indicated reduced OS duration of LUAD patients after external validation in two cohorts (**Figure 6**). Following multivariate-cox regression analysis, m^6A risk score was an independent prognostic indicator of LUAD (**Figure 7**). ROC curves confirmed the excellent performance in predicting LUAD patients' outcomes.

Previously, ANLN up-regulation is in relation to LUAD metastasis (Xu et al., 2019). ANLN promotes LUAD progression via activating RHOA and involving the PI3K/AKT signaling (Suzuki et al., 2005). PLK1/vimentin pathway promotes immune escape

through recruitment of Smad2/3 to PD-L1 promoter in LUAD metastasis (Jang et al., 2021). PLK1 induces migration of LUAD epithelial cells through STAT3 (Yan et al., 2018). IGF2BP1 induces LUAD progression via interaction with circXPO1 (Huang et al., 2020). Up-regulation of IGF2BP1 contributes to an unfavorable prognosis of LUAD (Huang et al., 2019). HMMR acts as an oncogene of LUAD and induces tumor progression (Li W. et al., 2020). NEIL3 that is correlated to immune infiltrations serves as an independent indicator for prediction of LUAD survival (Zhao et al., 2021). CXCL17 is an important determinator for LUAD spine metastasis (Liu W. et al., 2020). IRX5 as an oncogene is in relation to LUAD outcomes (Zhang et al., 2018). PKP2 accelerates the development of LUAD through increasing focal adhesion and EMT (Wu Y. et al., 2021). Elevated expression of LYPD3 contributes to LUAD carcinogenesis and unfavorable survival outcomes (Hu et al., 2020). ABCC2, a multidrug resistance-associated protein, displays an increased expression in LUAD (Maruhashi et al., 2018). DKK1 is an immune-associated prognostic marker in LUAD (Zhang et al., 2019). Above findings revealed the critical biological implications of the variables in the m^6A risk score in the progression of LUAD.

Furthermore, our data indicated that in comparison to the nomogram established by a single prognostic indicator, the nomogram established by the m^6A risk score and clinical features might become the best model in prediction of short- and long-term survival of LUAD patients, thereby possibly assisting clinical management and therapy (**Figure 8**). However, there are some limitations in our study, as follows: firstly, more information should be provided for internal mechanisms of m^6A modification; secondly, the prognostic value of the m^6A risk score should be verified in prospective research.

CONCLUSION

Collectively, this study comprehensively characterized the expression and clinical implication of m^6A regulators in LUAD. Two m^6A modification patterns were conducted, with different OS and tumor immunity. Furthermore, we developed the m^6A risk score, which had high accuracy in predicting LUAD prognosis. Thus, our data may provide a reliable tool for prediction of prognosis and optimization of immunotherapy for LUAD patients.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**supplementary material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

HZ conceived and designed the study. YM conducted most of the experiments and data analysis, and wrote the manuscript. All authors reviewed and approved the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.746666/full#supplementary-material

REFERENCES

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J. Clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Chao, Y., Shang, J., and Ji, W. (2020). ALKBH5-m6A-FOXM1 Signaling axis Promotes Proliferation and Invasion of Lung Adenocarcinoma Cells under Intermittent Hypoxia. *Biochem. Biophysical Res. Commun.* 521 (2), 499–506. doi:10.1016/j.bbrc.2019.10.145

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Cui, Y., Fang, W., Li, C., Tang, K., Zhang, J., Lei, Y., et al. (2019). Development and Validation of a Novel Signature to Predict Overall Survival in "Driver Gene-Negative" Lung Adenocarcinoma (LUAD): Results of a Multicenter Study. *Clin. Cancer Res.* 25 (5), 1546–1556. doi:10.1158/1078-0432.Ccr-18-2545

Ding, Y., Qi, N., Wang, K., Huang, Y., Liao, J., Wang, H., et al. (2020). FTO Facilitates Lung Adenocarcinoma Cell Progression by Activating Cell Migration through mRNA Demethylation. *Ott* 13, 1461–1470. doi:10.2147/ott.S231914

Engebretsen, S., and Bohlin, J. (2019). Statistical Predictions with Glmnet. *Clin. Epigenet* 11 (1), 123. doi:10.1186/s13148-019-0730-1

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene Expression Regulation Mediated through Reversible m6A RNA Methylation. *Nat. Rev. Genet.* 15 (5), 293–306. doi:10.1038/nrg3724

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Hu, P., Huang, Y., Gao, Y., Yan, H., Li, X., Zhang, J., et al. (2020). Elevated Expression of LYPD3 Is Associated with Lung Adenocarcinoma Carcinogenesis and Poor Prognosis. *DNA Cel Biol.* 39 (4), 522–532. doi:10.1089/dna.2019.5116

Huang, H., Wang, D., Guo, W., Zhuang, X., and He, Y. (2019). Correlated Low IGF2BP1 and FOXM1 Expression Predicts a Good Prognosis in Lung Adenocarcinoma. *Pathol. - Res. Pract.* 215 (7), 152433. doi:10.1016/j.prp.2019.152433

Huang, Q., Guo, H., Wang, S., Ma, Y., Chen, H., Li, H., et al. (2020). A Novel Circular RNA, circXPO1, Promotes Lung Adenocarcinoma Progression by Interacting with IGF2BP1. *Cell Death Dis* 11 (12), 1031. doi:10.1038/s41419-020-03237-8

Jang, H.-R., Shin, S.-B., Kim, C.-H., Won, J.-Y., Xu, R., Kim, D.-E., et al. (2021). PLK1/vimentin Signaling Facilitates Immune Escape by Recruiting Smad2/3 to PD-L1 Promoter in Metastatic Lung Adenocarcinoma. *Cell Death Differ* 28, 2745–2764. doi:10.1038/s41418-021-00781-4

Jia, Q., Wu, W., Wang, Y., Alexander, P. B., Sun, C., Gong, Z., et al. (2018). Local Mutational Diversity Drives Intratumoral Immune Heterogeneity in Non-small Cell Lung Cancer. *Nat. Commun.* 9 (1), 5361. doi:10.1038/s41467-018-07767-w

Li, F., Wang, H., Huang, H., Zhang, L., Wang, D., and Wan, Y. (2020). m6A RNA Methylation Regulators Participate in the Malignant Progression and Have Clinical Prognostic Value in Lung Adenocarcinoma. *Front. Genet.* 11, 994. doi:10.3389/fgene.2020.00994

Li, W., Pan, T., Jiang, W., and Zhao, H. (2020). HCG18/miR-34a-5p/HMMR axis Accelerates the Progression of Lung Adenocarcinoma. *Biomed. Pharmacother.* 129, 110217. doi:10.1016/j.biopha.2020.110217

Li, Y., Gu, J., Xu, F., Zhu, Q., Chen, Y., Ge, D., et al. (2020). Molecular Characterization, Biological Function, Tumor Microenvironment

Association and Clinical Significance of m6A Regulators in Lung Adenocarcinoma. *Brief Bioinform* 22, bbaa225. doi:10.1093/bib/bbaa225

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cel Syst.* 1 (6), 417–425. doi:10.1016/j.cels.2015.12.004

Liu, J., Han, X., Chen, L., Han, D., Mu, X., Hu, X., et al. (2020). TRIM28 Is a Distinct Prognostic Biomarker that Worsens the Tumor Immune Microenvironment in Lung Adenocarcinoma. *aging* 12 (20), 20308–20331. doi:10.18632/aging.103804

Liu, W., Xie, X., and Wu, J. (2020). Mechanism of Lung Adenocarcinoma Spine Metastasis Induced by CXCL17. *Cell Oncol.* 43 (2), 311–320. doi:10.1007/s13402-019-00491-7

Ma, L., Chen, T., Zhang, X., Miao, Y., Tian, X., Yu, K., et al. (2021). The m6A Reader YTHDC2 Inhibits Lung Adenocarcinoma Tumorigenesis by Suppressing SLC7A11-dependent Antioxidant Function. *Redox Biol.* 38, 101801. doi:10.1016/j.redox.2020.101801

Maruhashi, R., Akizuki, R., Sato, T., Matsunaga, T., Endo, S., Yamaguchi, M., et al. (2018). Elevation of Sensitivity to Anticancer Agents of Human Lung Adenocarcinoma A549 Cells by Knockdown of Claudin-2 Expression in Monolayer and Spheroid Culture Models. *Biochim. Biophys. Acta (Bba) - Mol. Cel Res.* 1865 (3), 470–479. doi:10.1016/j.bbamcr.2017.12.005

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007

Rittmeyer, A., Barlesi, F., Waterkamp, D., Park, K., Ciardiello, F., von Pawel, J., et al. (2017). Atezolizumab versus Docetaxel in Patients with Previously Treated Non-small-cell Lung Cancer (OAK): a Phase 3, Open-Label, Multicentre Randomised Controlled Trial. *The Lancet* 389 (10066), 255–265. doi:10.1016/s0140-6736(16)32517-x

Rousseaux, S., Debernardi, A., Jacquiau, B., Vitte, A.-L., Vesin, A., Nagy-Mignotte, H., et al. (2013). Ectopic Activation of Germline and Placental Genes Identifies Aggressive Metastasis-Prone Lung Cancers. *Sci. Translational Med.* 5 (186), 186ra166. doi:10.1126/scitranslmed.3005723

Schabath, M. B., Welsh, E. A., Fulp, W. J., Chen, L., Teer, J. K., Thompson, Z. J., et al. (2016). Differential Association of STK11 and TP53 with KRAS Mutation-Associated Gene Expression, Proliferation and Immune Surveillance in Lung Adenocarcinoma. *Oncogene* 35 (24), 3209–3216. doi:10.1038/onc.2015.375

Schmidt, L., Eskiocak, B., Kohn, R., Dang, C., Joshi, N. S., DuPage, M., et al. (2019). Enhanced Adaptive Immune Responses in Lung Adenocarcinoma through Natural Killer Cell Stimulation. *Proc. Natl. Acad. Sci. USA* 116 (35), 17460–17469. doi:10.1073/pnas.1904253116

Shi, Y., Fan, S., Wu, M., Zuo, Z., Li, X., Jiang, L., et al. (2019). YTHDF1 Links Hypoxia Adaptation and Non-small Cell Lung Cancer Progression. *Nat. Commun.* 10 (1), 4892. doi:10.1038/s41467-019-12801-6

Suzuki, C., Daigo, Y., Ishikawa, N., Kato, T., Hayama, S., Ito, T., et al. (2005). ANLN Plays a Critical Role in Human Lung Carcinogenesis through the Activation of RHOA and by Involvement in the Phosphoinositide 3-kinase/AKT Pathway. *Cancer Res.* 65 (24), 11314–11325. doi:10.1158/0008-5472.Can-05-1507

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible. *Nucleic Acids Res.* 45 (D1), D362–d368. doi:10.1093/nar/gkw937

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a Class Discovery Tool with Confidence Assessments and Item Tracking. *Bioinformatics* 26 (12), 1572–1573. doi:10.1093/bioinformatics/btq170

Wu, X., Sheng, H., Wang, L., Xia, P., Wang, Y., Yu, L., et al. (2021). A Five-m6A Regulatory Gene Signature Is a Prognostic Biomarker in Lung Adenocarcinoma Patients. *Aging* 13 (7), 10034–10057. doi:10.18632/aging.202761

Wu, Y., Liu, L., Shen, X., Liu, W., and Ma, R. (2021). Plakophilin-2 Promotes Lung Adenocarcinoma Development via Enhancing Focal Adhesion and Epithelial-Mesenchymal Transition. *Cmar* 13, 559–570. doi:10.2147/cmar.S281663

Xu, F., Huang, X., Li, Y., Chen, Y., and Lin, L. (2021). m6A-related lncRNAs Are Potential Biomarkers for Predicting Prognoses and Immune Responses in Patients with LUAD. *Mol. Ther. - Nucleic Acids* 24, 780–791. doi:10.1016/j.omtn.2021.04.003

Xu, J., Zheng, H., Yuan, S., Zhou, B., Zhao, W., Pan, Y., et al. (2019). Overexpression of ANLN in Lung Adenocarcinoma Is Associated with Metastasis. *Thorac. Cancer* 10 (8), 1702–1709. doi:10.1111/1759-7714.13135

Yan, W., Yu, H., Li, W., Li, F., Wang, S., Yu, N., et al. (2018). Plk1 Promotes the Migration of Human Lung Adenocarcinoma Epithelial Cells via STAT3 Signaling. *Oncol. Lett.* 16 (5), 6801–6807. doi:10.3892/ol.2018.9437

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, D.-L., Qu, L.-W., Ma, L., Zhou, Y.-C., Wang, G.-Z., Zhao, X.-C., et al. (2018). Genome-wide Identification of Transcription Factors that Are Critical to Non-small Cell Lung Cancer. *Cancer Lett.* 434, 132–143. doi:10.1016/j.canlet.2018.07.020

Zhang, H., Meltzer, P., and Davis, S. (2013). RCircos: an R Package for Circos 2D Track Plots. *BMC Bioinformatics* 14, 244. doi:10.1186/1471-2105-14-244

Zhang, M., Zhu, K., Pu, H., Wang, Z., Zhao, H., Zhang, J., et al. (2019). An Immune-Related Signature Predicts Survival in Patients with Lung Adenocarcinoma. *Front. Oncol.* 9, 1314. doi:10.3389/fonc.2019.01314

Zhang, C., Zhang, G., Sun, N., Zhang, Z., Zhang, Z., Luo, Y., et al. (2020). Comprehensive Molecular Analyses of a TNF Family-Based Signature with Regard to Prognosis, Immune Features, and Biomarkers for Immunotherapy in Lung Adenocarcinoma. *EBioMedicine* 59, 102959. doi:10.1016/j.ebiom.2020.102959

Zhang, H., Shi, X., Huang, T., Zhao, X., Chen, W., Gu, N., et al. (2020). Dynamic Landscape and Evolution of m6A Methylation in Human. *Nucleic Acids Res.* 48 (11), 6251–6264. doi:10.1093/nar/gkaa347

Zhao, C., Liu, J., Zhou, H., Qian, X., Sun, H., Chen, X., et al. (2021). NEIL3 May Act as a Potential Prognostic Biomarker for Lung Adenocarcinoma. *Cancer Cel Int* 21 (1), 228. doi:10.1186/s12935-021-01938-4

Zhou, H., Zheng, M., Shi, M., Wang, J., Huang, Z., Zhang, H., et al. (2021). Characteristic of Molecular Subtypes in Lung Adenocarcinoma Based on m6A RNA Methylation Modification and Immune Microenvironment. *BMC Cancer* 21 (1), 938. doi:10.1186/s12885-021-08655-1

Zhu, J., Wang, M., and Hu, D. (2020). Deciphering N6-Methyladenosine-Related Genes Signature to Predict Survival in Lung Adenocarcinoma. *Biomed. Res. Int.* 2020, 1–13. doi:10.1155/2020/2514230

Zhu, X., Chen, L., Liu, L., and Niu, X. (2019). EMT-mediated Acquired EGFR-TKI Resistance in NSCLC: Mechanisms and Strategies. *Front. Oncol.* 9, 1044. doi:10.3389/fonc.2019.01044

Zhuang, Z., Chen, L., Mao, Y., Zheng, Q., Li, H., Huang, Y., et al. (2020). Diagnostic, Progressive and Prognostic Performance of m6A Methylation RNA Regulators in Lung Adenocarcinoma. *Int. J. Biol. Sci.* 16 (11), 1785–1797. doi:10.7150/ijbs.39046

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 17 December 2021
doi: 10.3389/fgene.2021.790888

N6-methyladenosine Regulator-Mediated Immune Genes Identify Breast Cancer Immune Subtypes and Predict Immunotherapy Efficacy

Meng-Meng Zhang[1,2†], Yi-Lin Lin[2†], Wen-Feng Zeng[3], Yang Li[2], Yang Yang[1], Miao Liu[1], Ying-Jiang Ye[2], Ke-Wei Jiang[2], Shu Wang[1,2] and Shan Wang[2*]*

[1]Department of Breast Surgery, Peking University People's Hospital, Beijing, China, [2]Laboratory of Surgical Oncology, Beijing Key Laboratory of Colorectal Cancer Diagnosis and Treatment Research, Peking University People's Hospital, Beijing, China, [3]Breast Tumor Center, Sun Yat-Sen Memorial Hospital, Sun Yat-Sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Dongyu Zhao,
Peking University, China

Reviewed by:
Jingfa Xiao,
Chinese Academy of Sciences, China
Xinlei Gao,
Boston Children's Hospital, Harvard Medical School, United States

***Correspondence:**
Shu Wang
shuwang@pkuph.edu.cn
Shan Wang
shanwang60@sina.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: 07 October 2021
Accepted: 12 November 2021
Published: 17 December 2021

Citation:
Zhang M-M, Lin Y-L, Zeng W-F, Li Y, Yang Y, Liu M, Ye Y-J, Jiang K-W, Wang S and Wang S (2021) N6-methyladenosine Regulator-Mediated Immune Genes Identify Breast Cancer Immune Subtypes and Predict Immunotherapy Efficacy. Front. Genet. 12:790888. doi: 10.3389/fgene.2021.790888

Breast cancer (BRCA) is a heterogeneous malignancy closely related to the tumor microenvironment (TME) cell infiltration. *N*6-methyladenosine (m6A) modification of mRNA plays a crucial regulator in regulating the immune microenvironment of BRCA. Immunotherapy represents a paradigm shift in BRCA treatment; however, lack of an appropriate approach for treatment evaluation is a significant issue in this field. In this study, we attempted to establish a prognostic signature of BRCA based on m6A-related immune genes and to investigate the potential association between prognosis and immunotherapy. We comprehensively evaluated the m6A modification patterns of BRCA tissues and non-tumor tissues from The Cancer Genome Atlas and the modification patterns with TME cell-infiltrating characteristics. Overall, 1,977 TME-related genes were identified in the literature. Based on LASSO and Cox regression analyses, the m6A-related immune score (m6A-IS) was established to characterize the TME of BRCA and predict prognosis and efficacy associated with immunotherapy. We developed an m6A-IS to effectively predict immune infiltration and the prognosis of patients with BRCA. The prognostic score model represented robust predictive performance in both the training and validation cohorts. The low-m6A-IS group was characterized by enhanced antigen presentation and improved immune checkpoint expression, further indicating sensitivity to immunotherapy. Compared with the patients in the high-score group, the overall survival rate after treatment in the low-score group was significantly higher in the testing and validation cohorts. We constructed an m6A-IS system to examine the ability of the m6A signature to predict the infiltration of immune cells of the TME in BRCA, and the m6A-IS system acted as an independent prognostic biomarker that predicts the response of patients with BRCA in immunotherapy.

Keywords: N6-methyladenosine, breast cancer, tumor microenvironment, immunotherapy, prediction

INTRODUCTION

Breast cancer is the biggest threat to female health worldwide (Siegel et al., 2020). Although comprehensive efforts have been made in breast cancer treatment, including chemotherapy, radiotherapy, and molecular therapeutics, breast cancer continues to be associated with significant mortality in women and an equally substantial socioeconomic burden (DeSantis et al., 2015). Immunotherapy is revolutionizing the therapeutic approach for solid malignancies, and accumulating data indicate that immune checkpoint antagonists such as programmed cell death-1 (PD-1)/programmed death ligand-1 (PD-L1) inhibitors can induce efficacious and durable clinical responses in a proportion of patients with breast cancer, especially metastatic breast cancer (Emens, 2018; Franzoi et al., 2021). A community of epithelial-derived tumor cells mixed with a community of stromal components, referred to as "tumor microenvironment" (TME), is increasingly recognized as indispensable for mammary tumorigenesis. Immune cells, including regulatory T cells (Tregs), myeloid-derived suppressor cells, and B cells, extracellular matrix components, cancer-associated fibroblasts, blood vessels, and cancer-associated adipocytes are critical components of the TME (Deepak et al., 2020). Because of their complexity and heterogeneity, cancer cells can escape immune surveillance by the TME and induce antitumor immune system suppression, drug resistance, and recurrence of breast cancer (Hanahan and Weinberg, 2011). Therefore, focusing on recognizing the composition and the alterations in the molecular signatures of cells in the TME comprehensively may help in identifying the different immune phenotypes of breast cancer and predicting immunotherapeutic responsiveness.

Posttranscriptional regulation of the transcriptome is an important biological process, and over 170 chemical modifications in RNA have been identified to date (Frye et al., 2018). *N*6-methyladenosine (m6A) modification is a dynamic process of RNA posttranscriptional modification and exerts multiple functions in many biology processes, such as transportation, RNA processing, splicing, stability, and degradation of the target RNA (Alarcon et al., 2015; Patil et al., 2016). The m6A machinery is a dynamic and reversible process, including methylation by the methyltransferase complex ("writer"), removal by demethylases ("eraser"), and recognition by m6A-binding proteins ("reader") (Yang et al., 2018). m6A modification regulates oncogenesis and tumor development, which has been testified (Li et al., 2019; Vu et al., 2019). Overexpression of the m6A reader *YTHDF3* or the m6A demethylase *ALKBH5* may enhance the transcription of m6A-enriched genes in breast cancer, facilitating breast cancer brain metastasis (Zhang et al., 2016; Chang et al., 2020).

Since numerous studies have concentrated on tumor intrinsic cells, the potential role of m6A modification in TME and antitumor immune response has been rarely reported. However, little is known about whether messenger RNA (mRNA) m6A methylation in immune cells is responsible for regulating the TME, which leads to inhibition of immune function and tumor migration. Recent studies have provided some clues. YTHDF1, a well-known m6A-binding protein, undermines the cross-presentation of engulfed neoantigen-specific immunity by interacting with transcripts encoding lysosomal proteases in dendritic cells, suggesting that altered m6A modification may facilitate immune evasion in tumors (Han et al., 2019).

In our study, we initially identified the characteristics of immune cell types and 24 m6A regulators in the literature. Then, we identified the m6A-related immune genes with TME cell-infiltrating characteristics using Pearson's correlation analysis. A new m6A-related immune score (m6A-IS) prediction system was constructed based on m6A-related genes to assess the prognosis of breast cancer. Moreover, we showed the relationship between m6A-IS and the response to immunotherapy. Ultimately, a predictive nomogram for survival prediction of individual patients with breast cancer verified that m6A modification is non-negligible in drawing different TME characterizations. A flowchart of our research is shown in **Figure 1**.

MATERIALS AND METHODS

Data Acquisition

The mRNA expression profile data and DNA mutation data (VarScan2) of breast cancer samples were obtained from The Cancer Genome Atlas (TCGA) database; 1,068 samples with complete prognostic information were included. A total of 1,094 breast cancer samples with complete prognostic information were obtained in the METABRIC cohort from cBioportal (www.cbioportal.org). Transcripts per million data were used for subsequent analysis. The data for a cohort of patients with metastatic urothelial cancer treated with anti-PD-L1 therapy were obtained from the R software package IMvigor210CoreBiologies (IMvigor210, version 1.0.0) (Mariathasan et al., 2018).

Immune Score and Gene Set Enrichment Analysis

The gene set file "c2.cp.kegg.v6.2" was downloaded from the Molecular Signatures database (MSigDB; https://www.gsea-msigdb.org/gsea/index.jsp). The CIBERSORT (Newman et al., 2015) algorithm was used to evaluate the infiltration levels of immune cells in the sample, and 23 immune cell signatures were also used to evaluate the infiltration state of the TME (Charoentong et al., 2017; Jia et al., 2018). ESTIMATE was used to evaluate the immune and stromal scores of each sample (Yoshihara et al., 2013). Single-sample gene set enrichment analysis (ssGSEA) was used to calculate the enrichment scores of the samples using the GSVA package (Hänzelmann et al., 2013).

Identification of *N*^{6}-Methyladenosine-Related Immune Genes

We obtained a total of 24 m6A regulators from the study by Zhang et al. (2020). A total of 1,997 immune genes were collected

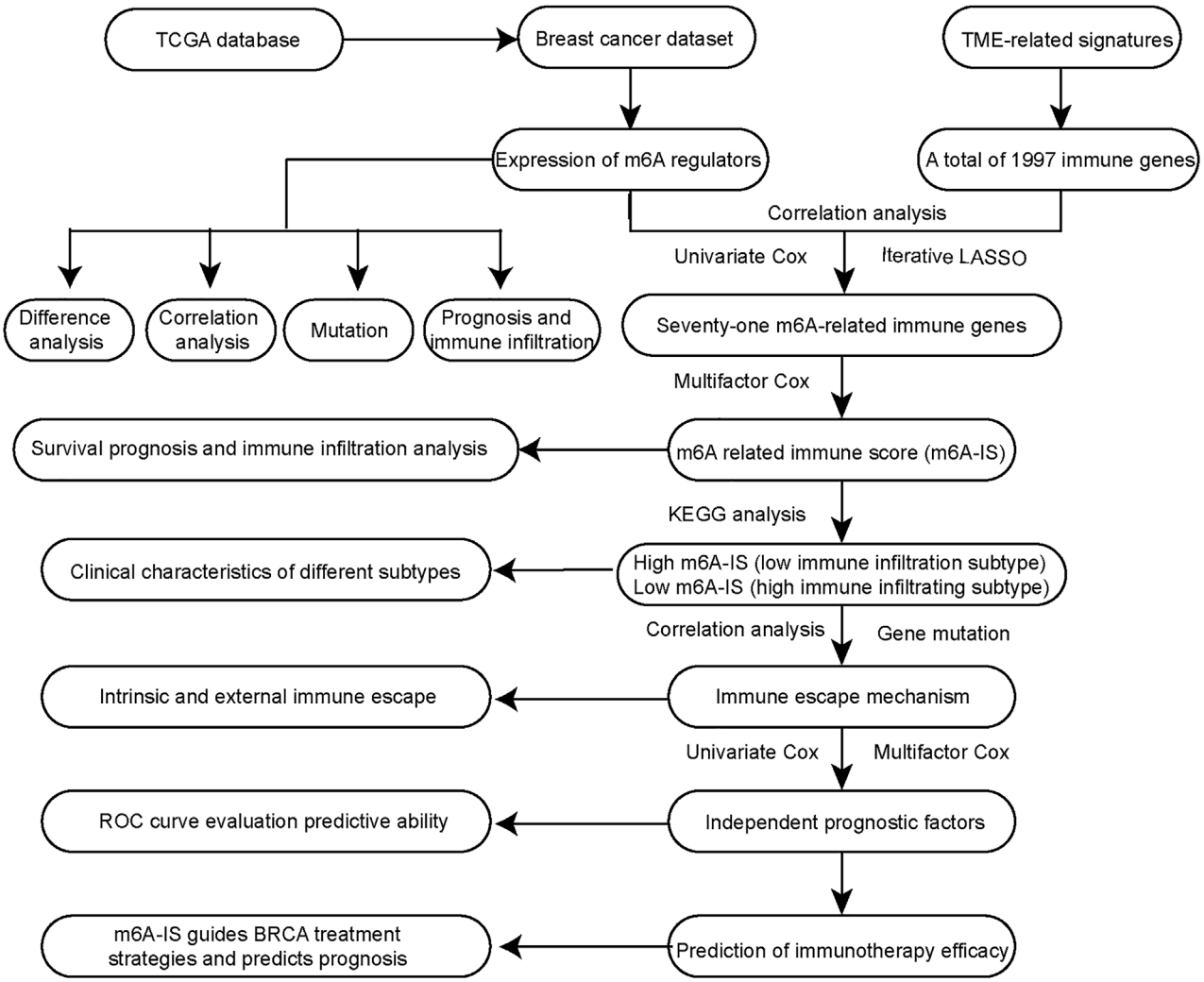

FIGURE 1 | Flowchart of the research.

from 184 TME-related signatures (Wang et al., 2021). We identified m6A-related immune genes based on Pearson's correlation analysis between the expression levels of m6A regulators and immune genes in breast cancer. |Pearson's correlation coefficient| ≥ 0.5 and $p < 0.001$ were set as cutoff values. A Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to identify the m6A-related immune genes associated with prognosis.

Construction of *N*^{6}-Methyladenosine-Related Immune Score

Univariate Cox regression analysis was used to screen the m6A-related immune genes ($p < 0.05$ was set as the cutoff value). Iterative LASSO (least absolute shrinkage and selection operator) was used to screen the m6A-related immune genes for subsequent analysis (Wang et al., 2021). The number of iterations was 500, and genes with a frequency greater than 50 were consensus genes for the iteration LASSO. The order of frequency represents the degree of influence of these features, and these features were then incorporated into the multivariate Cox regression model; the inclusion was stopped when the area under the curve (AUC) value of the receiver operating characteristic curve (ROC) reached its peak. The m6A-related immune genes obtained were used to construct the m6A-IS. The m6A-IS was expressed as follows: m6A-IS = (coefficient $mRNA_1$ × $mRNA_1$ expression) + (coefficient $mRNA_2$ ×expression of $mRNA_2$) + . . .+ (coefficient $mRNA_n$ × expression $mRNA_n$). Visualization of the prognostic value of m6A-IS was obtained using a nomogram.

Survival Analysis

We arranged the m6A-IS from low to high, starting from one low-expressing patient by setting the loop, calculating the corresponding individual *p*-value and hazard ratio (HR), and saving the calculation results. The minimum *p*-value obtained

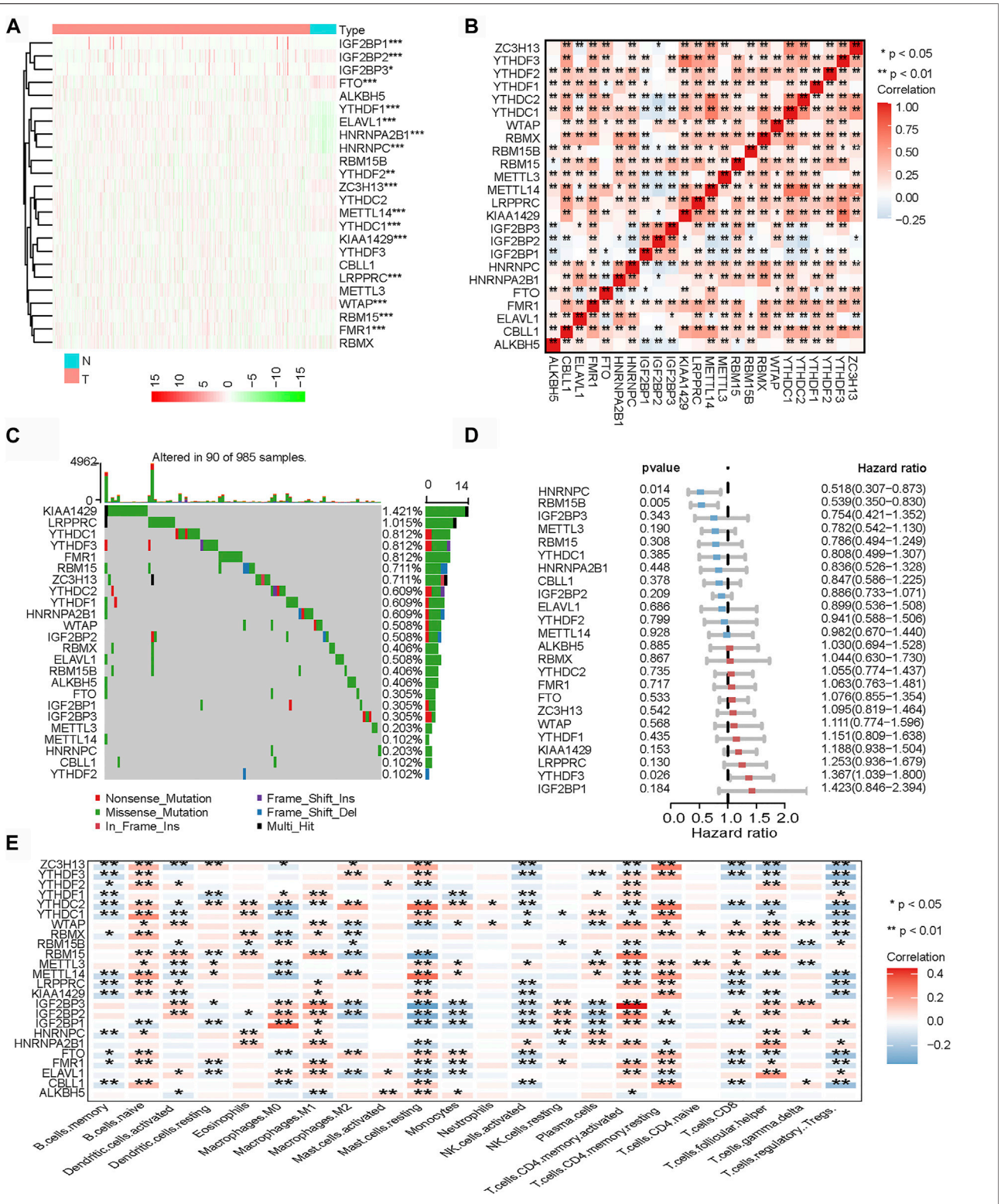

FIGURE 2 | Landscape of N^6-methyladenosine (m6A) regulators in breast cancer. **(A)** Heatmap of all m6A regulators in breast cancer tissues and normal tissues from The Genome Cancer Atlas (TCGA) database. Each *column* represents individual patients (*blue* indicates normal tissue, *pink* indicates breast cancer tissue; the *darker the red*, the more obvious the upregulation of gene expression, and the *darker the green*, the more obvious the downregulation of gene expression). **(B)** Co-expression of m6A regulators (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(C)** Mutation frequency of 24 m6A regulators in 985 breast cancer samples from TCGA-BRCA cohort. **(D)** Univariate Cox regression was performed to analyze the hazard ratio of each m6A-related gene in predicting overall survival in breast cancer. **(E)** CIBERSORT was used to analyze the component correlation between the 22 immune cells [tumor microenvironment (TME) infiltrating cells] and 24 m6A regulators (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$).

was used to determine the grouping information, and the samples were then divided into two groups. The Kaplan–Meier method estimated the overall survival (OS) curve, and the difference between survival distributions was evaluated using the two-sided log-rank test implemented in the R package survival. The Kaplan–Meier survival curve constructed *via* the R package survmin.

Statistical Analysis

The limma package in R was used to determine differentially expressed genes in the breast cancer cohorts. The ggplot2 package and ComplexHeatmap package were used to draw heatmaps and other maps. The R package forest plot was used to plot the forest plots. Wilcoxon's rank-sum test was used to analyze the differences between the two groups. The Kruskal–Wallis test was used to compare differences between three or more groups. The maftools package was used to map the gene mutations. The AUC was quantified using the pROC R package. All statistical *p*-values were two-sided, and statistical significance was set at $p < 0.05$. All analyses were performed using the R software (version 4.0.2).

RESULTS

Landscape of N^6-Methyladenosine Regulators in Breast Cancer

In the aggregate, 24 m6A regulators were identified in our study. In order to enrich the dysregulated genes and their correlations, the RNA transcriptomic datasets containing the next-generation sequencing (RNA sequencing, RNA-seq) data of 1,109 breast cancer tissues and 113 non-tumor tissues from TCGA project (TCGA-BRCA) was downloaded. Then, we explored the differential expressions in breast cancer tissues and normal mammary tissues in TCGA. Compared with normal tissues, there were 17 m6A regulators that were significantly differentially expressed in breast cancer tissues (**Figure 2A**). A correlation analysis was performed for the 24 m6A regulators (**Figure 2B**). The results showed that the expressions of the 24 m6A regulators were significantly correlated. Subsequently, we summarized the frequency of somatic mutations in the 24 m6A regulators in breast cancer. Of 985 samples, only 90 had mutations in the m6A regulators; indeed, there was a very low frequency of mutations (9.14%). Interestingly, we observed few mutations in breast cancer populations (**Figure 2C**). By performing univariate Cox regression analysis, we identified the affected prognosis of the 24 m6A-associated genes in patients with breast cancer (**Figure 2D**). Among the 24 m6A regulators, *HNRNPC* and *RBM15B* were identified as significant protective factors for survival ($p < 0$ 0.05), and *YTHDF3* was a risk factor for survival ($p < 0$ 0.05).

Related research works have shown that the expressions of m6A regulators are associated with the heterogeneity of the TME (Han et al., 2019; Zhang et al., 2020). Therefore, we further analyzed the relationship between the m6A regulators and cell infiltration in the TME. We analyzed the association between 22 types of immune cells (CIBERSORT algorithm) and m6A regulators (**Figure 2E**). The heatmap showed that most of the significant enrichments were found in immune cells. The infiltration levels of Tregs, activated natural killer (NK) cells, and memory B cells were significantly related to the expressions of most m6A regulators. The infiltration levels of naive B cells, M1 macrophages, and memory CD4 T cells were significantly positively correlated with most m6A regulators. The results showed that the expressions of m6A regulators were significantly correlated with the levels of immune cell infiltration ($p < 0.05$); indeed, they play a non-negligible role in the regulation of the TME in breast cancer.

Construction of the N^6-Methyladenosine-Related Immune Score and Analysis of Its Characteristics

Correlation analysis was used for preliminary screening to further analyze which TME-related immune genes were related to m6A-regulated expression. Through correlation analysis, a total of 534 TME-related genes were found to be significantly positively or negatively correlated with m6A regulators (| Pearson's correlation coefficient| ≥ 0.5 and $p < 0.001$) (**Figure 3A**). Univariate Cox regression analysis further revealed that 71 TME-related immune genes were significantly associated with breast cancer prognosis ($p < 0.05$; **Figure 3A**). Finally, 71 TME-related immune genes were included in the iterative LASSO algorithm for analysis. The results showed that, when these 28 TME-related immune genes were included (**Supplementary Table S1**), the model had the highest accuracy in predicting the prognosis (**Figure 3B**).

Multivariate Cox regression analysis was performed to construct the m6A-IS. The m6A-IS was expressed as follows: m6A-IS = (coefficient $mRNA_1$ × $mRNA_1$ expression) + (coefficient $mRNA_2$ × expression of $mRNA_2$) + ...+ (coefficient $mRNA_n$ × expression $mRNA_n$). The results of the multivariate Cox regression analysis and m6A-IS are shown in **Supplementary Tables S2** and **S3**. The results of the survival analysis showed that m6A-IS = 1.38 was the best cutoff value, and the low-m6A-IS group had a longer survival time in breast cancer than the high-m6A-IS group (**Figure 3C**). We also observed the same result from the METABRIC cohort (**Supplementary Figure S1A**).

To identify the signatures of the low- and high-m6A-IS subgroups, we analyzed the differences in the expressions of m6A modulators in the two subgroups and discovered that a large proportion of m6A regulators were differentially expressed in the high- and low-m6A-IS groups dramatically (**Figure 3D**). The clinicopathological information of the breast cancer patients has been shown in **Table 1**. The results of the Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis showed that the low-m6A-IS group was mainly enriched in the activation of some immune pathways, indicating that patients in this group may be with immune subtypes (**Figure 3E**). The high-m6A-IS group was mainly enriched in glucose metabolism and lipid metabolism pathways, indicating that patients in this group may be with metabolic subtypes (**Figure 3E**).

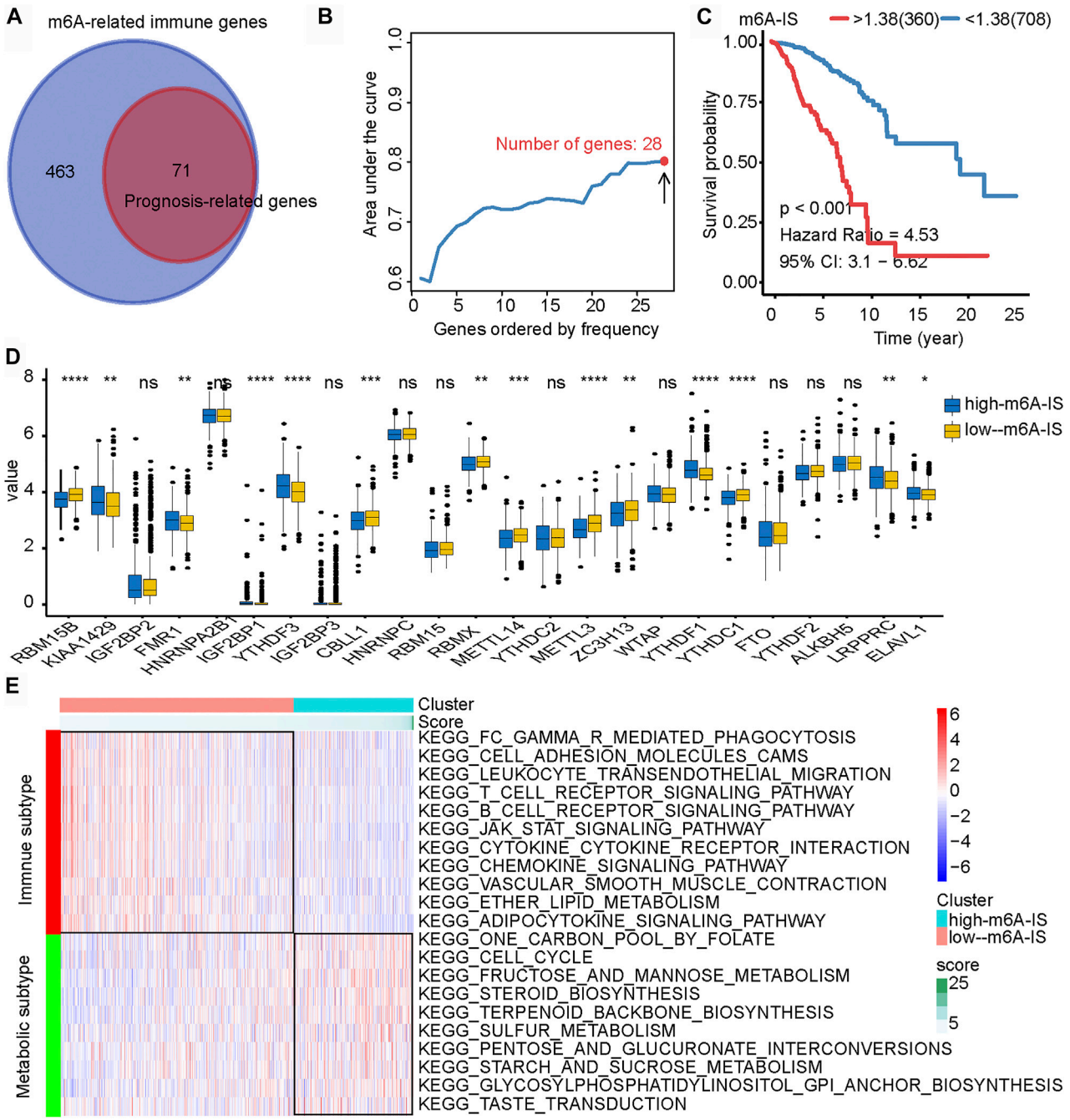

FIGURE 3 | Construction of the N^6-methyladenosine-related immune score (m6A-IS) and analysis of its characteristics. **(A)** Through correlation analysis, a total of 534 tumor microenvironment (TME)-related genes were found to be significantly positively or negatively correlated with m6A regulators (|Pearson's correlation coefficient| ≥ 0.5 and $p < 0.001$; **Figure2A**). Then, univariate Cox regression analysis further revealed that 71 TME-related immune genes were significantly related to the prognosis of breast cancer ($p < 0.05$; **Figure2A**). **(B)** The iterative LASSO algorithm was used to analyze a total of 71 TME-related genes. **(C)** The low m6A-IS subgroup showed longer survival than the high-m6A-IS subgroup (HR = 4.53, 95% CI = 3.04–6.66, $p < 0.001$). **(D)** Expressions of the 24 m6A regulators between the high- and low-m6A-IS groups of patients with breast cancer. High score, *blue*; low score, *yellow*. The *box bounds* represent an interquartile range of values, *center lines* represent the median value, and *black dots* show outliers. *Asterisks* represent the statistical *p*-value (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(E)** Heatmap of the significantly enriched terms in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways.

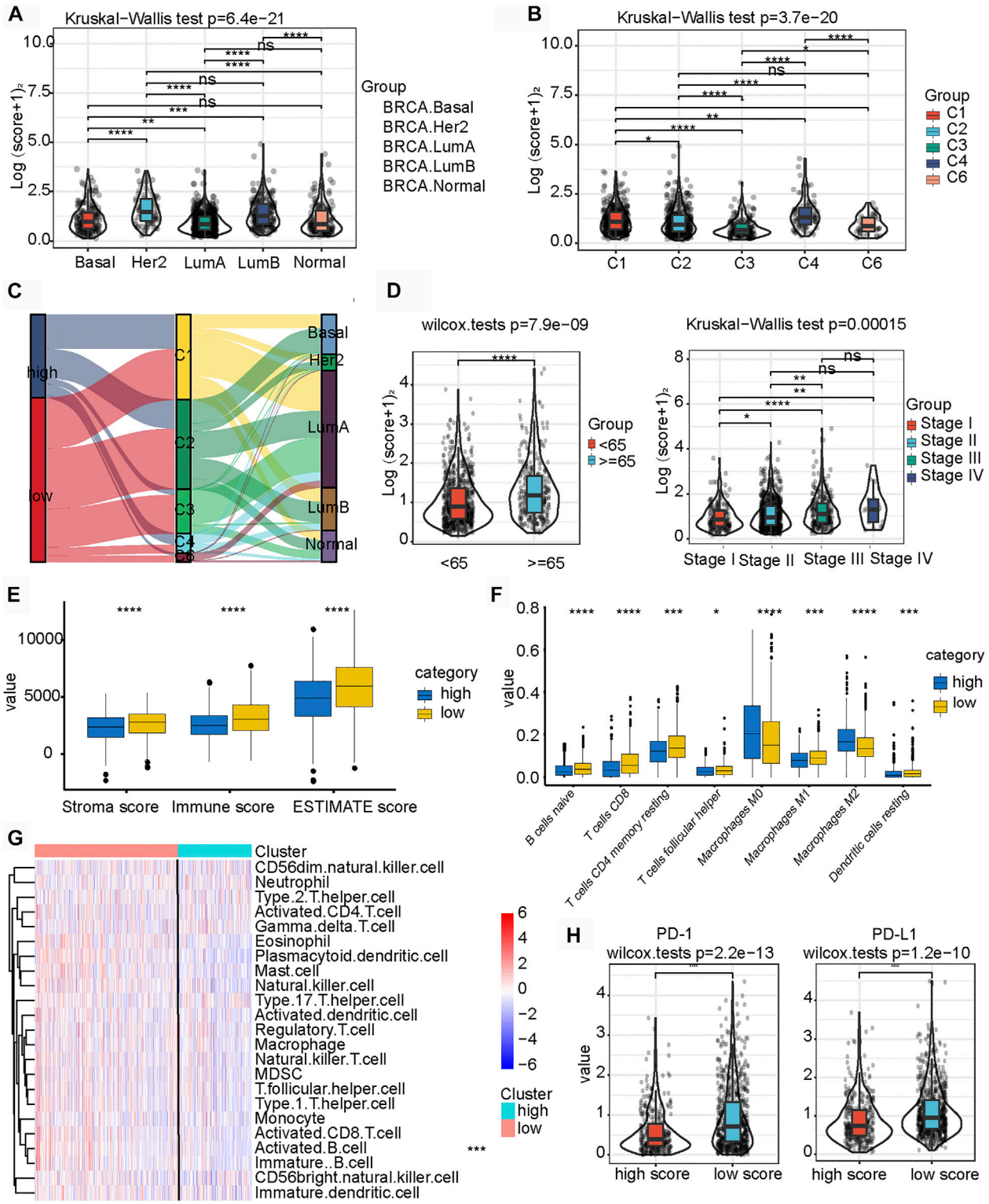

FIGURE 4 | Relationship between clinicopathological characteristics and the N^6-methyladenosine-related immune score (m6A-IS). **(A)** Difference in m6A-IS in breast cancer with different molecular subtypes. Statistical difference in different breast cancer molecular subtypes was compared by the Kruskal–Wallis test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(B)** Difference in m6A-IS between different immunogenomic subtypes. The Kruskal–Wallis test was performed to compare the statistical difference in each immunogenomic subtype (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(C)** Alluvial diagram representing the association of m6A-IS, six immune subtypes, and clinical molecular subtypes in breast cancer. **(D)** Breast cancer patients with different clinicopathological features had different expression levels of m6A-IS

(*Continued*)

FIGURE 4 | (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(E)** Difference between the stroma score, the immune score, and the ESTIMATE score in the high- and low-m6A-IS groups (*blue* represents the high-m6A-IS group and *yellow* the low-m6A-IS group) (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(F)** Difference in the abundance of 8 types of infiltrating immune cells between the high- and low-m6A-IS groups (*blue* represents the high-m6A-IS group and *yellow* the low-m6A-IS group) (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(G)** Heatmap of the tumor-infiltrating cell proportions between the high- and low-m6A-IS groups. **(H)** Expression levels of PD-1/PD-L1 between the low- and high-m6A-IS groups (Wilcox test: $p < 0.0001$).

Relationship Between Clinicopathological Characteristics and N^6-Methyladenosine-Related Immune Score

We investigated the expression levels of m6A-IS from the five breast cancer subtypes previously reported to further analyze its characteristics (Parker et al., 2009). We found that the Her-2 and luminal B subtypes had the highest m6A-IS, followed by the basal and normal subtypes. The luminal A subtype had the lowest m6A-IS level (**Figure 4A** and **Supplementary Figure S1B**). This is consistent with the results of previous studies: luminal A subtype has the best prognosis, Her-2 and luminal B subtypes have the worst prognosis, and the prognosis of the basal and normal subtypes can be characterized at a level between the best and worst. We classified the breast cancer samples based on the six immune subtypes previously identified using immunogenomic features (Thorsson et al., 2018). The results showed that m6A-IS expression was the lowest in the C3 subtype; meanwhile, it was the highest in the C4 subtype. This result suggests that the C3 subtype had the best prognosis, while the C4 subtype had the worst prognosis (**Figure 4B**). Thorsson et al. discovered that the C4 subtype displayed high M2 macrophage domination and low lymphocytic infiltrate, which induced poor outcomes. In contrast, the C3 subtype, which showed a type I immune response and remarkable Th17 signature, had favorable prognosis. Our results are in line with the conclusions of a previous report (Thorsson et al., 2018), indicating that the m6A-IS has good robustness in different breast cancer molecular subtypes. A Sankey diagram was used to visualize the relationship between m6A-IS and the two breast cancer subtypes (**Figure 4C**). In addition, between-group comparisons of age were performed using the Wilcoxon test, and there was a significant difference in the distributions of age. Patients older than 65 years were more likely to have higher scores than patients aged 65 years or younger (**Figure 4D** and **Supplementary Figure S1C**). Interestingly, we found that, with increased malignancy, the m6A-IS showed a significant stepwise increase in non-metastatic breast cancer; meanwhile, no significant increase was observed in patients with metastatic breast cancer (stage IV) (**Figure 4D**).

These results indicate that m6A-IS can be used to characterize the existing subtypes and clinical features of patients with breast cancer. However, the relationship between m6A-IS and TME remains unknown. For this reason, we found through ssGSEA that the low-m6A-IS group had a higher expression of stromal and immune scores than the high-m6A-IS group (**Figure 4E**), which is consistent with previous results (**Figure 3E**). In addition, the CIBERSORT algorithm (**Figure 4F**) and the enrichment analysis of 23 immune cells (**Figure 4G** and **Supplementary Figure S1D**) showed that the infiltration levels of B cells, CD8 T cells, and M1 tumor-associated macrophages in the low-m6A-IS group were significantly higher than those in the high-m6A-IS group. In contrast, the infiltration level of M2 tumor-associated macrophages increased significantly in the high-m6A-IS group. Finally, we found that the expression levels of the immune checkpoint molecules (PD1 and PD-L1) in the low-m6A-IS group were significantly higher than those in the high-m6A-IS group (**Figure 4H**). These results suggest that the low-m6A-IS group had increased immune cell infiltration and, simultaneously, had increased immunosuppression.

Potential Immune Escape Mechanisms in High and Low N^6-Methyladenosine-Related Immune Score

We further compared the two groups and found that the low-m6A-IS group had a higher expression of major histocompatibility complex (MHC)-related antigen-presenting molecules than the high-m6A-IS group (all $p < 0.05$; **Figure 5A**). We analyzed some co-inhibitory molecules to compare their expressions in the high and low-m6A-IS groups, and the results showed that the expression levels of these molecules were significantly correlated with m6A-IS. The results showed that the low-m6A-IS group had higher expressions of immunosuppressive molecules than the high-m6A-IS group (all $p < 0.05$; **Figure 5A**). Further analysis revealed that the low-m6A-IS group showed higher chemokine expressions (**Figure 5B**). These results indicate that the low-m6A-IS group had higher immunogenicity than the high-m6A-IS group. However, at the same time, there was a significant immunosuppressive state in the low-m6A-IS group. This suggests that there is a potential immune escape mechanism in low-m6A-IS.

Finally, the top 20 genes were further analyzed with the highest mutation frequency in the high and low-m6A-IS groups (**Figure 5C**). The results showed that *PIK3CA* and *CDH1* had increased mutation frequencies in the low-m6A-IS group, while *TP53* had an increased mutation frequency in the high-m6A-IS group. These results suggest that *PIK3CA* and *CDH1* mutations may be related to a high immune infiltration, while *TP53* mutations may be related to immunosuppression, which may require further analysis.

N^6-Methyladenosine-Related Immune Score Was an Independent Prognostic Factor for Patients With Breast Cancer

Univariate and multivariate Cox analyses represented that age, tumor stage, and the m6A-IS were independent prognostic

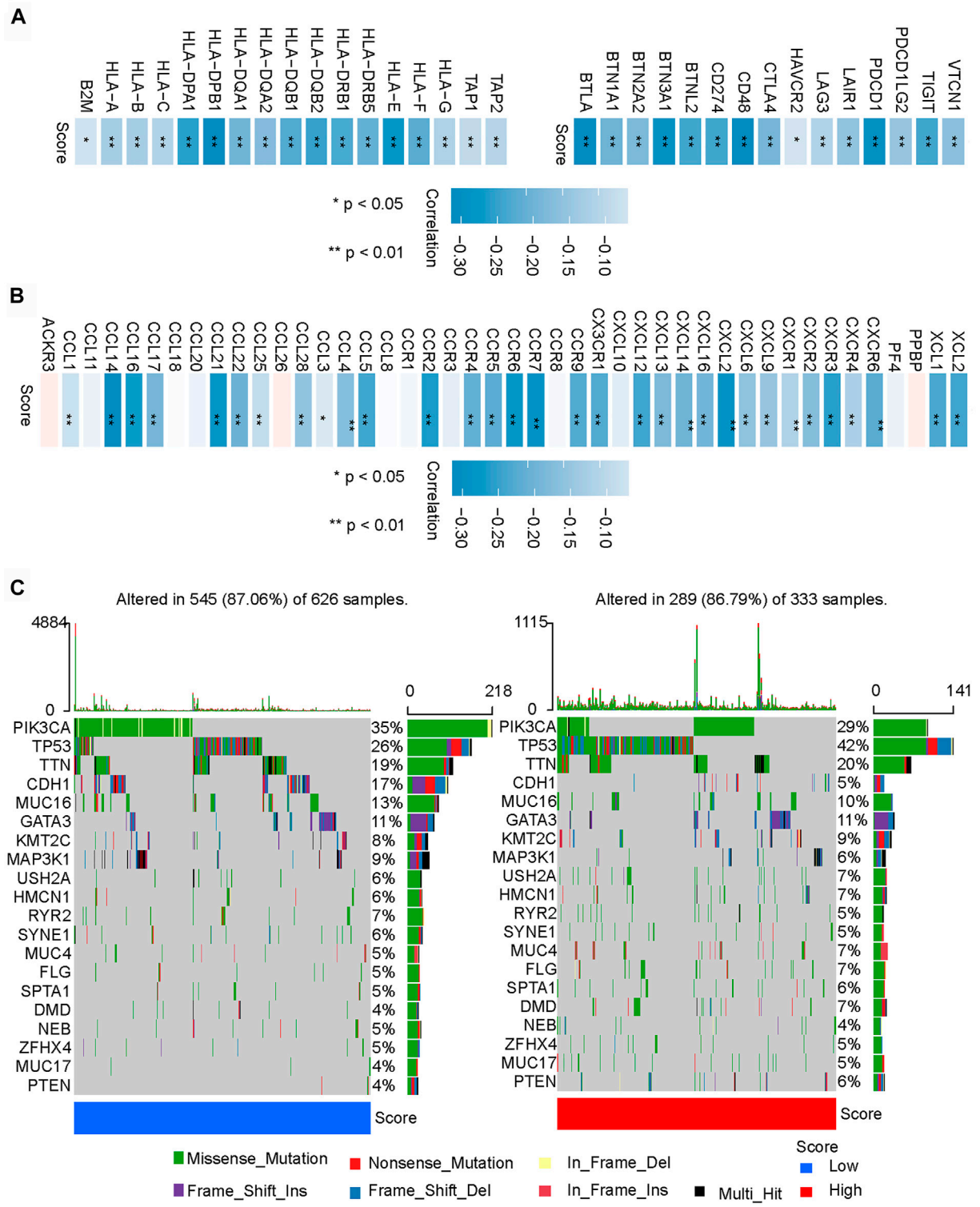

FIGURE 5 | Potential immune escape mechanisms in high and low N^6-methyladenosine-related immune scores (m6A-IS). **(A)** Correlation analysis between the expressions of MHC molecules and co-inhibitory molecules and m6A-IS (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(B)** Correlation analysis of the expressions of chemokines and m6A-IS (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). **(C)** Waterfall plot of tumor somatic mutations in the low- ***(left)*** and high-m6A-IS ***(right)*** groups. Individual patients are represented in each column. Missense mutation, *green*; nonstop mutation, *grey*; nonsense mutation, *red*; multi-hit, *black*. The *right bar plot* shows the mutation frequency of each gene in separate subgroups.

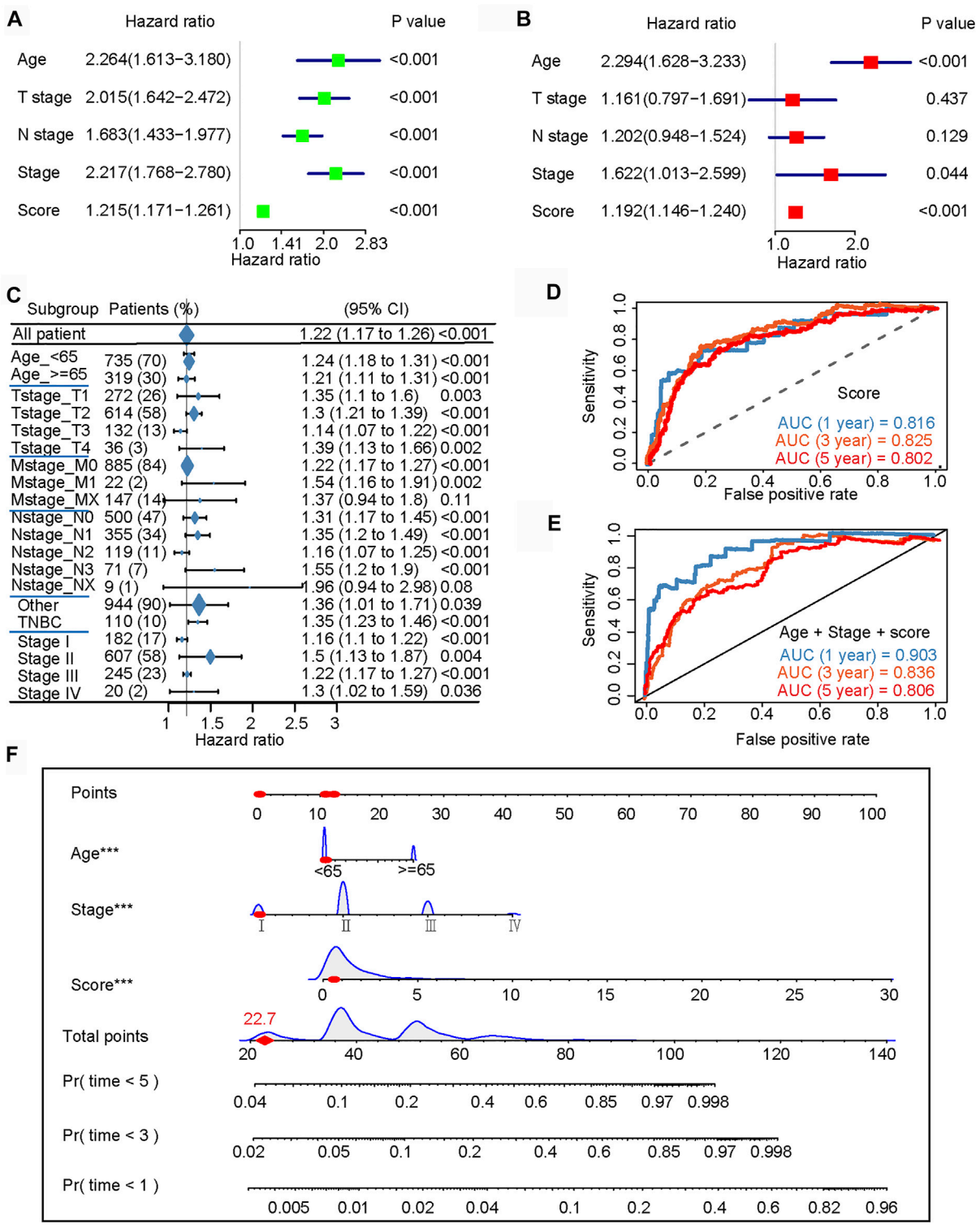

FIGURE 6 | N^6-methyladenosine-related immune score (m6A-IS) is an independent prognostic factor for patients with breast cancer (BRCA). **(A,B)** Univariate and multivariate Cox analyses further showed that age, tumor stage, and m6A-IS are independent prognostic predictors for patients with BRCA. **(C)** Univariate Cox regression analysis of the overall prognostic value of m6A-IS in each clinical feature subgroup. **(D)** ROC curve constructed based on m6A-IS. The AUCs of the ROC curve at 1, 3, and 5 years were 0.816, 0.825, and 0.802, respectively. **(E)** The m6A-IS combines age and tumor stage to construct a final predictive prognostic model. The results showed that the multi-predictor ROC has excellent accuracy for 1-, 3-, and 5-year overall survival (OS) (AUC = 0.903, 0.836, and 0.806, respectively). **(F)** Nomogram based on m6A-IS, age, and stage in The Cancer Genome Atlas (TCGA) cohort.

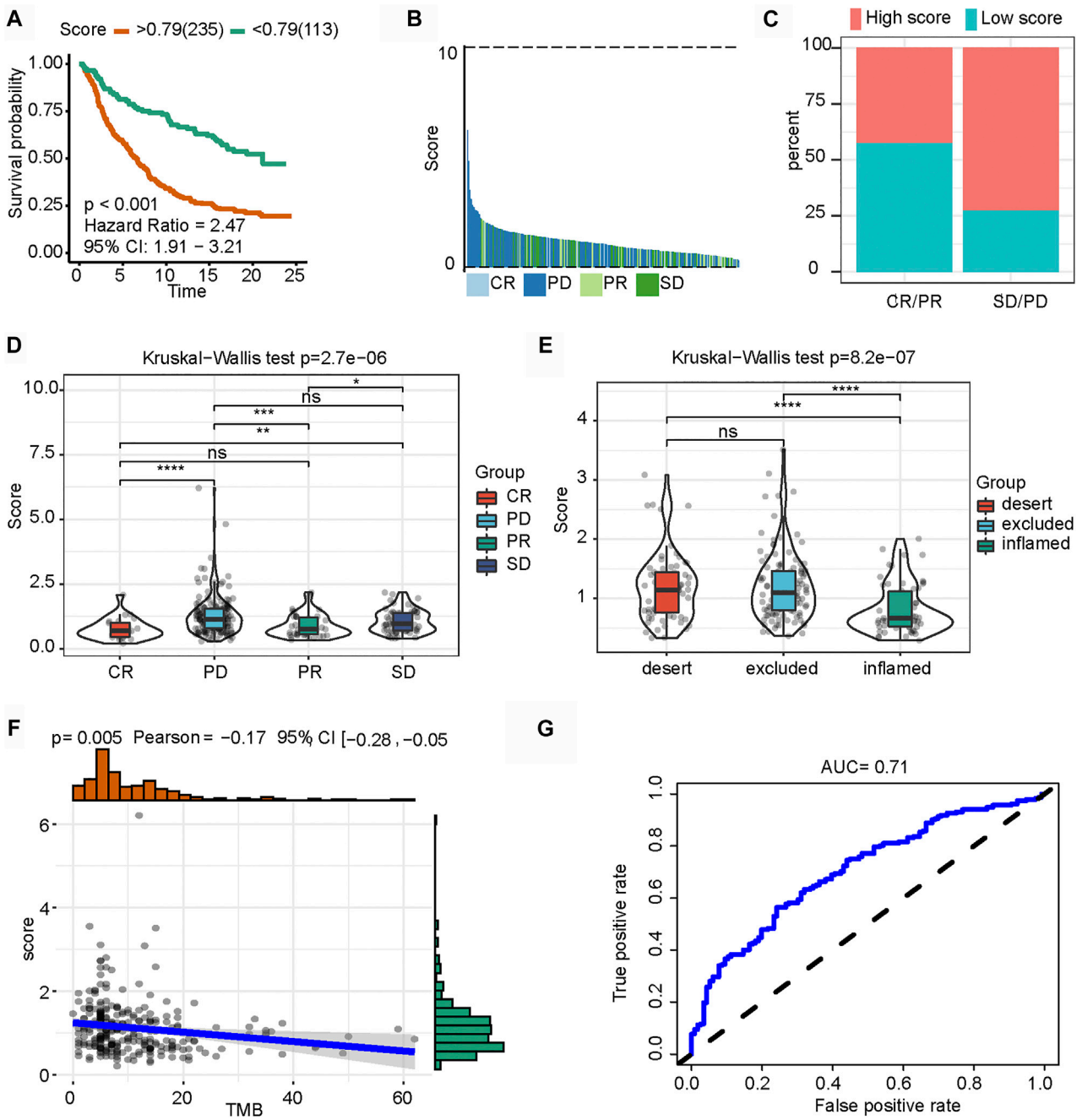

FIGURE 7 | Role of N^6-methyladenosine-related immune score (m6A-IS) in immunotherapy. **(A)** Patient survival analysis defined by the high- (n = 235) and low-m6A-IS (n = 113) groups based on the IMvigor210 cohort (log-rank test: $p < 0.001$). **(B)** Waterfall plot of m6A-IS for distinct clinical response groups based on the IMvigor210 cohort. **(C)** Stacked bar chart representing the percentage of clinical response patients assigned to two m6A-IS subgroups in the IMvigor210 cohort. High-score, *red*; low score, *blue*. **(D)** Boxplot of m6A-IS for distinct clinical response groups *via* the Kruskal–Wallis test ($p < 0.001$). **(E)** Differences in m6A-IS among three distinct tumor immune phenotypes in the IMvigor210 cohort by the Kruskal–Wallis test ($p < 0.001$). **(F)** Spearman's correlation analysis between m6A-IS and tumor mutation burden ($p = 0.005$). **(G)** Predictive value of the quantification of m6A-IS in patients treated with anti-PD-L1 immunotherapy (AUC = 0.71).

predictors for patients with breast cancer (**Figures 6A**, **B**). Subsequently, we conducted a more detailed stratification of breast cancer patients based on clinical characteristics. Univariate Cox regression analysis revealed that m6A-IS has a good prognostic value in each subgroup (**Figure 6C**). Furthermore, the triple-negative (TNBC) subtype has the strongest tumor immunogenicity of all BC subtypes (Liu et al., 2018); meanwhile, m6A-IS can also be used to assess prognosis in

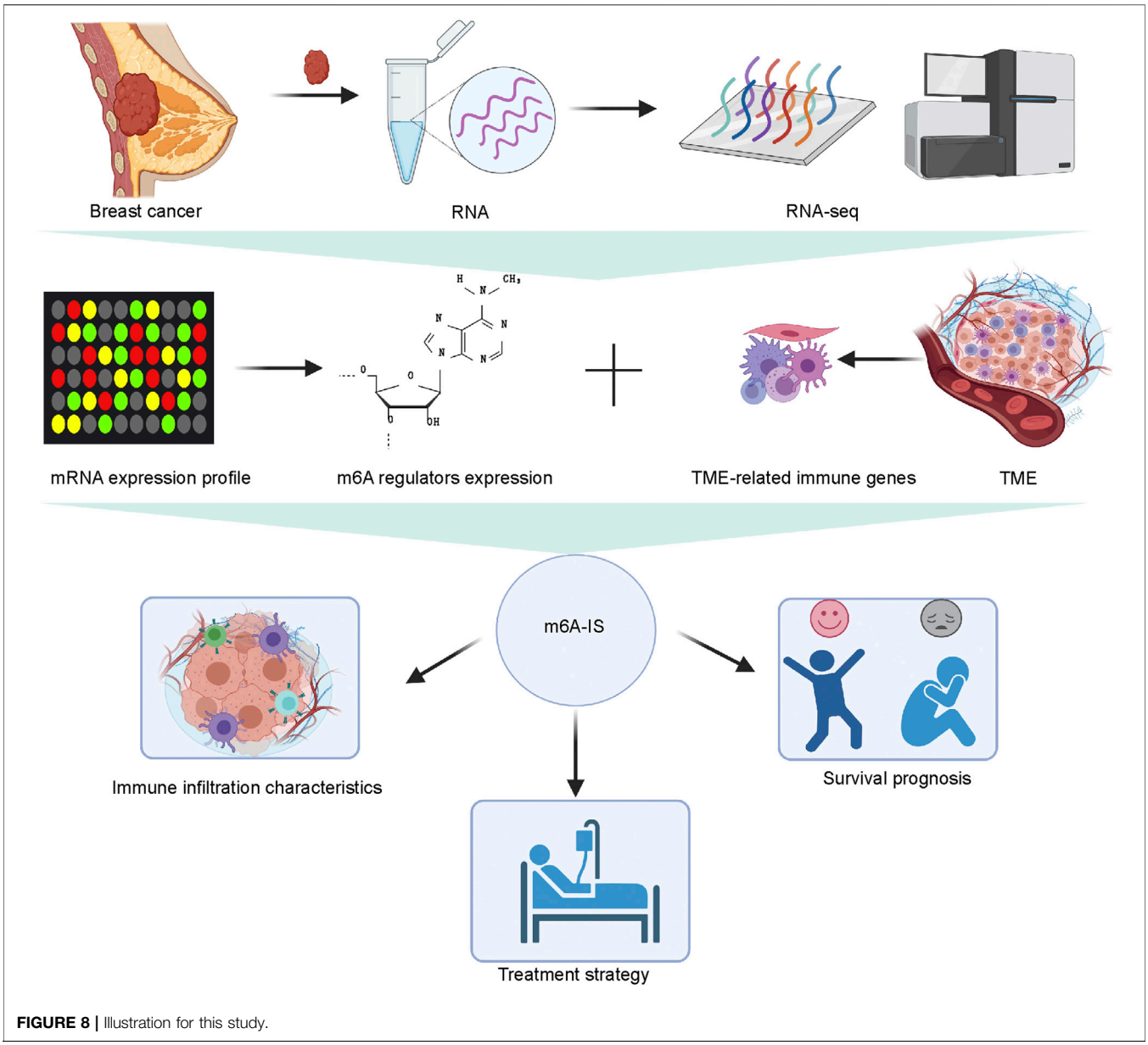

FIGURE 8 | Illustration for this study.

this subtype (**Figure 6C**). These results indicate that m6A-IS, as an independent prognostic indicator, may be useful for clinical prognosis evaluation.

An ROC curve was constructed based on the m6A-IS. We found that the AUCs of the ROC curve at 1, 3, and 5 years were 0.816, 0.825, and 0.802, respectively (**Figure 6D**). Next, we verified whether the m6A-IS can predict the prognosis of breast cancer patients in the METABRIC cohort (**Supplementary Figure S1E**). It was shown that m6A-IS has good ability to predict prognosis. The previous results found that age and tumor stage were also independent factors; therefore, we combined them with the m6A-IS to build the final predictive prognosis model (**Figure 6E**). The results showed that the multi-predictor ROC had excellent accuracy for 1-, 3-, and 5-year OS (AUC = 0.903, 0.836, and 0.806, respectively). Based on the results of the logistic regression analysis and the ROC curves, a nomogram was graphically depicted (**Figure 6F**). By calculating the total scores of each selected variable, the survival of individual breast cancer patients at 1, 3, and 5 years can be easily estimated by plotting a vertical line between the total points and each prognosis axis of the nomogram.

Role of the N^6-Methyladenosine-Related Immune Score in Anti-PD-L1 Immunotherapy

Recently, the connection of m6A modification with PD-L1 expression has been reported. In an anti-PD-L1 cohort (IMvigor210 cohort), the low-m6A-IS group exhibited significant clinical benefits and a markedly longer survival

TABLE 1 | Clinical characteristics of the high- and low-m6A-IS groups

	High m6A-IS (360)	Low m6A-IS (708)	*p*-value
Age (years)			
<65	236	510	0.0292
≥65	124	198	
T stage			
T1	60	219	<0.0001
T2	236	380	
T3	48	84	
T4	15	23	
Unknown	1	2	
N stage			
N0	173	329	0.1094
N1	107	249	
N2	44	76	
N3	32	41	
Unknown	4	13	
M stage			
M0	295	594	0.0363
M1	12	10	
Unknown	53	104	
Tumor stage			
Stage I	42	139	0.0010
Stage II	216	391	
Stage III	85	158	
Stage IV	12	8	
Unknown	5	12	
Subtype			
TNBC	16	94	<0.0001
Other	344	614	

time (**Figure 7A**). Compared with those with high m6A-IS, patients with low m6A-IS showed significant survival advantages and clinical response to anti-PD-L1 therapy (**Figures 7B–D**). Next, we examined the differences in m6A-IS among the different immune phenotypes in the IMvigor210 cohort. Interestingly, patients in the higher m6A-IS group were remarkably relevant to the exclusion and desert immune phenotypes, and it was difficult to achieve antitumor effects of the checkpoint inhibitors in these phenotypes (**Figure 7E**). Based on accumulating evidence, high levels of tumor mutation burden (TMB) and PD-L1 expression have long-lasting clinical responses to immunotherapy. A high TMB is generally considered to be a preexisting adaptive immune response to the tumor, and patients with a high TMB who received PD-1 blocking immunotherapy showed an improved response and enhanced clinical efficacy compared to patients with a moderate or low TMB (Hellmann et al., 2018). Further analysis revealed that m6A-IS was significantly negatively correlated with TMB (**Figure 7F**). These results suggest that the quantification of the modification patterns of m6A is a potential and robust biomarker for the assessment of prognosis and clinical response to immunotherapy (**Figure 7G**). In conclusion, this study emphasizes that m6A methylation modification is significantly correlated with tumor immunophenotype and anti-PD-L1 immunotherapy response, and established m6A modification characteristics may help predict the response to anti-PD-L1 immunotherapy. An illustration of this study is shown in **Figure 8**.

DISCUSSION

Breast cancer is characterized by high morbidity and ranks first among female malignant tumors globally (Harbeck et al., 2019). Within the genomic heterogeneity and diverse histological features, patients with breast cancer present individual responses to traditional treatments, including surgery, chemotherapy, hormonal therapy, and target therapy (Colozza et al., 2007; Foukakis et al., 2011). Immunotherapy combined with chemotherapy is emerging as a novel treatment regimen for breast cancer (Adams et al., 2019b). Although breast cancer is considered less immunogenic, lower mutational load than other solid carcinomas, the synergism between anti-PD1/PD-L1 agents and chemotherapy has been supported *via* multiple preclinical pieces of evidence, particularly in the TNBC subtype (Adams et al., 2019a; Adams et al., 2019c; Loibl et al., 2019; Planes-Laine et al., 2019). Similarly, more studies are needed to identify novel immunotherapy result prediction.

Sample classification methods based on a predefined multi-gene signature is a proven approach for predicting the treatment benefits of immune checkpoint inhibition in a variety of malignancies (Cristescu et al., 2015). Dysregulation of m6A methylation has been shown to be closely associated with the antitumor immune dysregulation through interactions with various m6A regulators (Wu et al., 2019; Yang et al., 2019; Wang et al., 2020b). In this study, we screened and identified 24 m6A regulators and drew a gene signature based on the dataset from TCGA. A prognostic model (m6A-IS) was constructed to correlate with the clinical outcomes of breast cancer patients by combining the roles of immune infiltrating cells in the TME with these m6a-related genes.

In our study, we constructed the m6A-IS to quantify the prognostic outcomes based on two groups (high and low), providing strong evidence for individualized immunotherapy in breast cancer. Indeed, the m6A-IS reflects the heterogeneity of patients. Secondly, the model links m6A methylation to the prognosis of breast cancer immunotherapy. The m6A-IS score includes, but is not limited to, regulatory factors related to m6A, such as *HNRNPC*, *RBM15B*, and *YTHDF3*. *HNRNPC* appeared to be related to good outcomes in glioma patients (Wang et al., 2020a) and was associated with an increased proportion of patients at low risk of lung squamous cell carcinoma based on an immune-related prediction model (Xu et al., 2020) that is similar to our model. However, Wu et al. found that aberrant upregulation of *HNRNPC* resulted in the accumulation of endogenous double-stranded RNA and tumorigenesis in breast cancer cell lines (Wu et al., 2018). Indeed, more samples are needed to draw more precise conclusions. The methyltransferase *RBM15B* is a paralog of *RBM15* and can be involved in regulating immunological phenotypes. High *RBM15B* levels are correlated with multiple immune signatures and cancer-related pathways (Fang et al., 2020). It has been reported that the m6A reader *YTHDF3* promotes ribosome loading with *YTHDF1*, and a high *YTHDF3* expression in breast cancer clinically correlates with brain metastases (Chang et al., 2020). However, further studies are needed to determine whether these genes can be new targets for improving the response to immunotherapy. Additionally, the

m6A-IS represents patients with different clinical characteristics and is related to immunotherapy. High grouping showed an m6A modification pattern characterized by an immune-excluded phenotype, suggesting worse clinical features and a lower predicted survival time. The pattern characterized by the immune-inflamed phenotype showed lower m6A-IS. Moreover, the infiltration of TME cells indicates that m6A-IS is important for immunotherapy. In low-group patients, the upregulation of immune cell infiltration associated with immune activation correlated with improved prognosis for immunotherapy. For example, a study showed that naive B cells and $CD8^+$ T cells appeared to be anticancer immune cells (Zhang et al., 2019; Bu et al., 2020); meanwhile, M2 macrophages, immune cells that promote tumor proliferation and metastasis, were increased in this study (Tariq et al., 2017). Interestingly, clinical trials have shown promising prospects for immunotherapy in patients with the TNBC subtype. In our predicted model, the TNBC subtype in patients was significantly associated with lower m6A-IS. Moreover, TMB is a determinant of the immune-mediated survival of breast cancer patients and can be an independent predictor of immunotherapeutic response in various cancers (Goodman et al., 2017; Thomas et al., 2018; Lee et al., 2019). Our data revealed a significant negative correlation between the m6A score and TMB. Additionally, m6A-IS may also be used for other tumors and immune-related tumors. Researchers have studied prognostic models of breast cancer immunotherapy under modified conditions in hypoxia, ferroptosis, or autophagy (Lin et al., 2020; Zheng et al., 2020; Tang et al., 2021). In our research, we focused on exploring the extensive regulatory mechanism of m6A methylation modification in the breast cancer microenvironment. Thus, our model is valuable in facilitating breast cancer treatment. In addition, these results were validated in the IMvigor210 cohort with a defined immunophenotype. This indicates that m6A-IS has a predictive advantage for selecting the appropriate immunotherapy for breast cancer. Moreover, our model can be used to further determine the TME infiltration pattern, which is a tumor immunophenotype.

Despite conducting multi-pronged and multi-database verification, our study has several limitations. Firstly, our prognostic model needs to be further validated using forward-looking, multicenter, real-world data. In addition, the underlying mechanism between m6A regulators and TME needs to be further tested *via* clinical molecular experiments of the potential molecular mechanism of the breast cancer immunotherapy response. The results of single-cell sequencing should contribute to an increased understanding of the specific changes in the TME, which is also an aspect of our future concern. These not only increase the challenges but also motivate us to conduct future research.

CONCLUSION

We constructed the m6A-IS to assess the prognosis of patients suffering from breast cancer. Patients with low m6A-IS had a longer survival time. The results of our study provide insights into the mechanism of immune infiltration and immune evasion in breast cancer based on m6A-IS stratification. The m6A-IS was used to stratify patients and determine those who will gain a survival benefit from immunotherapy, thereby contributing to improved diagnosis and treatment of breast cancer.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This study was supported by the Natural Science Foundation of Beijing (no. 7202212).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.790888/full#supplementary-material.

REFERENCES

Adams, S., Diamond, J. R., Hamilton, E., Pohlmann, P. R., Tolaney, S. M., Chang, C.-W., et al. (2019a). Atezolizumab Plus Nab-Paclitaxel in the Treatment of Metastatic Triple-Negative Breast Cancer with 2-Year Survival Follow-Up. *JAMA Oncol.* 5 (3), 334–342. doi:10.1001/jamaoncol.2018.5152

Adams, S., Gatti-Mays, M. E., Kalinsky, K., Korde, L. A., Sharon, E., Amiri-Kordestani, L., et al. (2019b). Current Landscape of Immunotherapy in Breast Cancer. *JAMA Oncol.* 5 (8), 1205–1214. doi:10.1001/jamaoncol.2018.7147

Adams, S., Loi, S., Toppmeyer, D., Cescon, D. W., De Laurentiis, M., Nanda, R., et al. (2019c). Pembrolizumab Monotherapy for Previously Untreated, PD-L1-Positive, Metastatic Triple-Negative Breast Cancer: Cohort B of the Phase II KEYNOTE-086 Study. *Ann. Oncol.* 30 (3), 405–411. doi:10.1093/annonc/mdy518

Alarcón, C. R., Lee, H., Goodarzi, H., Halberg, N., and Tavazoie, S. F. (2015). N6-methyladenosine marks Primary microRNAs for Processing. *Nature* 519 (7544), 482–485. doi:10.1038/nature14281

Bu, F., Nie, H., Zhu, X., Wu, T., Lin, K., Zhao, J., et al. (2020). A Signature of 18 Immune-related Gene Pairs to Predict the Prognosis of Pancreatic Cancer Patients. *Immun. Inflamm. Dis.* 8 (4), 713–726. doi:10.1002/iid3.363

Chang, G., Shi, L., Ye, Y., Shi, H., Zeng, L., Tiwary, S., et al. (2020). YTHDF3 Induces the Translation of m6A-Enriched Gene Transcripts to Promote Breast Cancer Brain Metastasis. *Cancer Cell* 38 (6), 857–871. doi:10.1016/j.ccell.2020.10.004

Charoentong, P., Finotello, F., Angelova, M., Mayer, C., Efremova, M., Rieder, D., et al. (2017). Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cel Rep.* 18 (1), 248–262. doi:10.1016/j.celrep.2016.12.019

Colozza, M., de Azambuja, E., Personeni, N., Lebrun, F., Piccart, M. J., and Cardoso, F. (2007). Achievements in Systemic Therapies in the Pregenomic Era in Metastatic Breast Cancer. *Oncol.* 12 (3), 253–270. doi:10.1634/theoncologist.12-3-253

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.-M., Ting, J. C., Wong, S. S., et al. (2015). Molecular Analysis of Gastric Cancer Identifies Subtypes Associated with Distinct Clinical Outcomes. *Nat. Med.* 21 (5), 449–456. doi:10.1038/nm.3850

Deepak, K. G. K., Vempati, R., Nagaraju, G. P., Dasari, V. R., S., N., Rao, D. N., et al. (2020). Tumor Microenvironment: Challenges and Opportunities in Targeting Metastasis of Triple Negative Breast Cancer. *Pharmacol. Res.* 153, 104683. doi:10.1016/j.phrs.2020.104683

DeSantis, C. E., Bray, F., Ferlay, J., Lortet-Tieulent, J., Anderson, B. O., and Jemal, A. (2015). International Variation in Female Breast Cancer Incidence and Mortality Rates. *Cancer Epidemiol. Biomarkers Prev.* 24 (10), 1495–1506. doi:10.1158/1055-9965.EPI-15-0535

Emens, L. A. (2018). Breast Cancer Immunotherapy: Facts and Hopes. *Clin. Cancer Res.* 24 (3), 511–520. doi:10.1158/1078-0432.CCR-16-3001

Fang, J., Hu, M., Sun, Y., Zhou, S., and Li, H. (2020). Expression Profile Analysis of m6A RNA Methylation Regulators Indicates They Are Immune Signature Associated and Can Predict Survival in Kidney Renal Cell Carcinoma. *DNA Cel Biol.* 39, 2194–2211. doi:10.1089/dna.2020.5767

Foukakis, T., Fornander, T., Lekberg, T., Hellborg, H., Adolfsson, J., and Bergh, J. (2011). Age-specific Trends of Survival in Metastatic Breast Cancer: 26 Years Longitudinal Data from a Population-Based Cancer Registry in Stockholm, Sweden. *Breast Cancer Res. Treat.* 130 (2), 553–560. doi:10.1007/s10549-011-1594-z

Franzoi, M. A., Romano, E., and Piccart, M. (2021). Immunotherapy for Early Breast Cancer: Too Soon, Too Superficial, or Just Right? *Ann. Oncol.* 32 (3), 323–336. doi:10.1016/j.annonc.2020.11.022

Frye, M., Harada, B. T., Behm, M., and He, C. (2018). RNA Modifications Modulate Gene Expression during Development. *Science* 361 (6409), 1346–1349. doi:10.1126/science.aau1646

Goodman, A. M., Kato, S., Bazhenova, L., Patel, S. P., Frampton, G. M., Miller, V., et al. (2017). Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol. Cancer Ther.* 16 (11), 2598–2608. doi:10.1158/1535-7163.MCT-17-0386

Han, D., Liu, J., Chen, C., Dong, L., Liu, Y., Chang, R., et al. (2019). Anti-tumour Immunity Controlled through mRNA m6A Methylation and YTHDF1 in Dendritic Cells. *Nature* 566 (7743), 270–274. doi:10.1038/s41586-019-0916-x

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of Cancer: the Next Generation. *Cell* 144 (5), 646–674. doi:10.1016/j.cell.2011.02.013

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: Gene Set Variation Analysis for Microarray and RNA-Seq Data. *BMC Bioinformatics* 14, 7. doi:10.1186/1471-2105-14-7

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., et al. (2019). Breast Cancer. *Nat. Rev. Dis. Primers* 5 (1), 66. doi:10.1038/s41572-019-0111-2

Hellmann, M. D., Callahan, M. K., Awad, M. M., Calvo, E., Ascierto, P. A., Atmaca, A., et al. (2018). Tumor Mutational Burden and Efficacy of Nivolumab Monotherapy and in Combination with Ipilimumab in Small-Cell Lung Cancer. *Cancer Cell* 33 (5), 853–861. doi:10.1016/j.ccell.2018.04.001

Jia, Q., Wu, W., Wang, Y., Alexander, P. B., Sun, C., Gong, Z., et al. (2018). Local Mutational Diversity Drives Intratumoral Immune Heterogeneity in Non-small Cell Lung Cancer. *Nat. Commun.* 9 (1), 5361. doi:10.1038/s41467-018-07767-w

Lee, D.-W., Han, S.-W., Bae, J. M., Jang, H., Han, H., Kim, H., et al. (2019). Tumor Mutation Burden and Prognosis in Patients with Colorectal Cancer Treated with Adjuvant Fluoropyrimidine and Oxaliplatin. *Clin. Cancer Res.* 25 (20), 6141–6147. doi:10.1158/1078-0432.CCR-19-1105

Li, Y., Xiao, J., Bai, J., Tian, Y., Qu, Y., Chen, X., et al. (2019). Molecular Characterization and Clinical Relevance of m6A Regulators across 33 Cancer Types. *Mol. Cancer* 18 (1), 137. doi:10.1186/s12943-019-1066-3

Lin, Q.-G., Liu, W., Mo, Y.-z., Han, J., Guo, Z.-X., Zheng, W., et al. (2020). Development of Prognostic index Based on Autophagy-Related Genes Analysis in Breast Cancer. *Aging* 12 (2), 1366–1376. doi:10.18632/aging.102687

Liu, Z., Li, M., Jiang, Z., and Wang, X. (2018). A Comprehensive Immunologic Portrait of Triple-Negative Breast Cancer. *Translational Oncol.* 11 (2), 311–329. doi:10.1016/j.tranon.2018.01.011

Loibl, S., Untch, M., Burchardi, N., Huober, J., Sinn, B. V., Blohmer, J.-U., et al. (2019). A Randomised Phase II Study Investigating Durvalumab in Addition to an Anthracycline Taxane-Based Neoadjuvant Therapy in Early Triple-Negative Breast Cancer: Clinical Results and Biomarker Analysis of GeparNuevo Study. *Ann. Oncol.* 30 (8), 1279–1288. doi:10.1093/annonc/mdz158

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., et al. (2018). TGFβ Attenuates Tumour Response to PD-L1 Blockade by Contributing to Exclusion of T Cells. *Nature* 554 (7693), 544–548. doi:10.1038/nature25501

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust Enumeration of Cell Subsets from Tissue Expression Profiles. *Nat. Methods* 12 (5), 453–457. doi:10.1038/nmeth.3337

Parker, J. S., Mullins, M., Cheang, M. C. U., Leung, S., Voduc, D., Vickery, T., et al. (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. *Jco* 27 (8), 1160–1167. doi:10.1200/jco.2008.18.1370

Patil, D. P., Chen, C.-K., Pickering, B. F., Chow, A., Jackson, C., Guttman, M., et al. (2016). m6A RNA Methylation Promotes XIST-Mediated Transcriptional Repression. *Nature* 537 (7620), 369–373. doi:10.1038/nature19342

Planes-Laine, G., Rochigneux, P., Bertucci, F., Chrétien, A. S., Viens, P., Sabatier, R., et al. (2019). PD-1/PD-L1 Targeting in Breast Cancer: The First Clinical Evidences Are Emerging. A Literature Review. *Cancers* 11 (7), 1033. doi:10.3390/cancers11071033

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer Statistics, 2020. *CA A. Cancer J. Clin.* 70 (1), 7–30. doi:10.3322/caac.21590

Tang, Y., Li, C., Zhang, Y.-J., and Wu, Z.-H. (2021). Ferroptosis-Related Long Non-coding RNA Signature Predicts the Prognosis of Head and Neck Squamous Cell Carcinoma. *Int. J. Biol. Sci.* 17 (3), 702–711. doi:10.7150/ijbs.55552

Tariq, M., Zhang, J., Liang, G., Ding, L., He, Q., and Yang, B. (2017). Macrophage Polarization: Anti-cancer Strategies to Target Tumor-Associated Macrophage in Breast Cancer. *J. Cel. Biochem.* 118 (9), 2484–2501. doi:10.1002/jcb.25895

Thomas, A., Routh, E. D., Pullikuth, A., Jin, G., Su, J., Chou, J. W., et al. (2018). Tumor Mutational burden Is a Determinant of Immune-Mediated Survival in Breast Cancer. *Oncoimmunology* 7 (10), e1490854. doi:10.1080/2162402X.2018.1490854

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T. H., et al. (2018). The Immune Landscape of Cancer. *Immunity* 48 (4), 812. doi:10.1016/j.immuni.2018.03.023

Vu, L. P., Cheng, Y., and Kharas, M. G. (2019). The Biology of m6A RNA Methylation in Normal and Malignant Hematopoiesis. *Cancer Discov.* 9 (1), 25–33. doi:10.1158/2159-8290.CD-18-0959

Wang, L.-c., Chen, S.-h., Shen, X.-l., Li, D.-c., Liu, H.-y., Ji, Y.-l., et al. (2020a). M6A RNA Methylation Regulator HNRNPC Contributes to Tumorigenesis and Predicts Prognosis in Glioblastoma Multiforme. *Front. Oncol.* 10, 536875. doi:10.3389/fonc.2020.536875

Wang, S., Xiong, Y., Zhang, Q., Su, D., Yu, C., Cao, Y., et al. (2021). Clinical Significance and Immunogenomic Landscape Analyses of the Immune Cell Signature Based Prognostic Model for Patients with Breast Cancer. *Brief Bioinform* 22 (4), bbaa311. doi:10.1093/bib/bbaa311

Wang, Y., Zhang, Y., Du, Y., Zhou, M., Hu, Y., and Zhang, S. (2020b). Emerging Roles of N6-Methyladenosine (m6A) Modification in Breast Cancer. *Cell Biosci* 10 (1), 136. doi:10.1186/s13578-020-00502-3

Wu, H.-X., Chen, Y.-X., Wang, Z.-X., Zhao, Q., He, M.-M., Wang, Y.-N., et al. (2019). Alteration in TET1 as Potential Biomarker for Immune Checkpoint Blockade in Multiple Cancers. *J. Immunotherapy Cancer* 7 (1), 264. doi:10.1186/s40425-019-0737-3

Wu, Y., Zhao, W., Liu, Y., Tan, X., Li, X., Zou, Q., et al. (2018). Function of HNRNPC in Breast Cancer Cells by Controlling the dsRNA-induced Interferon Response. *EMBO J.* 37 (23), e99017. doi:10.15252/embj.201899017

Xu, F., Zhang, H., Chen, J., Lin, L., and Chen, Y. (2020). Immune Signature of T Follicular Helper Cells Predicts Clinical Prognostic and Therapeutic Impact in Lung Squamous Cell Carcinoma. *Int. Immunopharmacology* 81, 105932. doi:10.1016/j.intimp.2019.105932

Yang, S., Wei, J., Cui, Y.-H., Park, G., Shah, P., Deng, Y., et al. (2019). m6A mRNA Demethylase FTO Regulates Melanoma Tumorigenicity and Response to Anti-PD-1 Blockade. *Nat. Commun.* 10 (1), 2782. doi:10.1038/s41467-019-10669-0

Yang, Y., Hsu, P. J., Chen, Y.-S., and Yang, Y.-G. (2018). Dynamic Transcriptomic m6A Decoration: Writers, Erasers, Readers and Functions in RNA Metabolism. *Cell Res.* 28 (6), 616–624. doi:10.1038/s41422-018-0040-8

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., et al. (2013). Inferring Tumour Purity and Stromal and Immune Cell Admixture from Expression Data. *Nat. Commun.* 4, 2612. doi:10.1038/ncomms3612

Zhang, B., Wu, Q., Li, B., Wang, D., Wang, L., and Zhou, Y. L. (2020). m6A Regulator-Mediated Methylation Modification Patterns and Tumor Microenvironment Infiltration Characterization in Gastric Cancer. *Mol. Cancer* 19 (1), 53. doi:10.1186/s12943-020-01170-0

Zhang, C., Samanta, D., Lu, H., Bullen, J. W., Zhang, H., Chen, I., et al. (2016). Hypoxia Induces the Breast Cancer Stem Cell Phenotype by HIF-dependent and ALKBH5-Mediated m6A-Demethylation of NANOG mRNA. *Proc. Natl. Acad. Sci. USA* 113 (14), E2047–E2056. doi:10.1073/pnas.1602883113

Zhang, Z., Ma, L., Goswami, S., Ma, J., Zheng, B., Duan, M., et al. (2019). Landscape of Infiltrating B Cells and Their Clinical Significance in Human Hepatocellular Carcinoma. *Oncoimmunology* 8 (4), e1571388. doi:10.1080/2162402X.2019.1571388

Zheng, S., Zou, Y., Liang, J. y., Xiao, W., Yang, A., Meng, T., et al. (2020). Identification and Validation of a Combined Hypoxia and Immune index for Triple-negative Breast Cancer. *Mol. Oncol.* 14 (11), 2814–2833. doi:10.1002/1878-0261.12747

Conflict of Interest: The Handling Editor DZ declared a shared parent affiliation with the authors MMZ, YLL, YL, YY, ML, YJY, KJ, SW, SW at the time of the review.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 28 March 2022
doi: 10.3389/fgene.2022.839841

Mouse Paternal RNAs Initiate a Pattern of Metabolic Disorders in a Line-Dependent Manner

Guzide Satir-Basaran[1,2†], Leila Kianmehr[3†], Ecmel Mehmetbeyoglu[1,4†], Kezban Korkmaz Bayram[1,5], Mehmet Memis[1], Zeynep Yilmaz[1], Esra Tufan[1], Serpil Taheri[1,6], Fahrettin Kelestimur[7] and Minoo Rassoulzadegan[1,8]*

[1]Betul Ziya Eren Genome and Stem Cell Center, Erciyes University, Kayseri, Turkey, [2]Department of Biochemistry, Faculty of Pharmacy, Erciyes University, Kayseri, Turkey, [3]Departement of Animal Sciences and Marine Biology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran, [4]Department of Cancer and Genetics, Cardiff University, Cardiff, United Kingdom, [5]Department of Medical Genetics, Medical Faculty, Yıldırım Beyazıt University, 06800 Ankara, Turkey, [6]Department of Medical Biology, Erciyes University Medical School, Kayseri, Turkey, [7]Department of Endocrinology, Yeditepe University Medical School, Istanbul, Turkey, [8]INSERM-CNRS, Université de Nice, Nice, France

OPEN ACCESS

Edited by:
Dongyu Zhao,
Peking University, China

Reviewed by:
Ionel Sandovici,
University of Cambridge, United Kingdom
Raffaele Teperino,
Helmholtz Association of German Research Centres (HZ), Germany

***Correspondence:**
Minoo Rassoulzadegan
minoo@unice.fr

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to RNA, a section of the journal Frontiers in Genetics

Received: *20 December 2021*
Accepted: *08 February 2022*
Published: *28 March 2022*

Citation:
Satir-Basaran G, Kianmehr L, Mehmetbeyoglu E, Korkmaz Bayram K, Memis M, Yilmaz Z, Tufan E, Taheri S, Kelestimur F and Rassoulzadegan M (2022) Mouse Paternal RNAs Initiate a Pattern of Metabolic Disorders in a Line-Dependent Manner. Front. Genet. 13:839841. doi: 10.3389/fgene.2022.839841

A wide range of diseases result from environmental effects, and the levels of many native transcripts are altered. The alteration of non-coding RNAs (ncRNAs) and transmission of the variation to the next generation is increasingly recognized as a marker of disease. However, the determining signals and mechanisms of RNA-induced heritability remain unclear. We performed functional tests with four different genotypes of mice maintained on a high-fat diet to trace the transfer of the obesity/diabetes phenotype to the next generation in order to detect common signals. Two founders of four mouse lines (*B6/D2* hybrid and *Dnmt2*$^{-/-C57BL/6}$) resist and do not change their phenotype while their sperm RNAs after microinjection into fertilized mouse eggs transfer the newly acquired phenotypes in a susceptible inbred line (*C57BL/6* or *Balb/c*). Unexpectedly, in the same line of experiments, sperm RNA from animals raised on a normal diet when mixed with the sperm RNA from animals raised on a diet high in fat or synthetic miR-19b (inducer of obesity) affects or prevents the development of obesity and diabetes. However, it remains unclear what happens to ncRNA signaling under diet. With a comprehensive new analysis of the transcripts maintained as an RNA/DNA hybrid in sperm, we suggest that a fraction of the RNAs are stably attached to the genome. Thus, we propose that changes in the dynamics of ncRNA retention on DNA by factors such as transcriptional variations or lack of adequate methylation could serve as molecular markers to trace these epigenetics events.

Keywords: sperm, diet, mouse, epigenetic, Dnmt2, non-coding RNAs, miRNAs, DNA/RNA hybrids

INTRODUCTION

Parents' diet can be a concern for future generations, ultimately contributing to the development of diseases (Buettner et al., 2007; Hur et al., 2017). In mice raised on a high-fat diet (HFD), the offspring may develop obesity/type 2 diabetes (OB/T2D) like phenotypes (West and York, 1998; Pomp et al., 2008; Massiera et al., 2010; Sims et al., 2013; EASD, 2015). Inbred laboratory mouse models with known genetics raised under defined conditions allow investigation of this problem. However, unlike other lines (Winzell and Ahre´n, 2004), some laboratory mice are resistant to the development of OB/T2D like phenotypes when fed on HFD (Montgomery et al., 2013). To address a common

mechanism that determines susceptibility to HFD, we analyzed the phenotypes of four lineages of laboratory mice and their two successive generations F1 and F2. We studied the influence of genotypes on the response to HFD and its transgenerational maintenance. We followed four lines of mice, two inbred, *C57BL/6* (Rossmeisl et al., 2003) (often used) and *Balb/c*, one hybrid *B6/D2* (F1 cross of *C57BL/6* and *DBA2*), and a mutant $Dnmt2^{-/-C57BL/6}$ (missing of functional RNA methyltransferase see for review (Lyko, 2018)) maintaining all mice under the same housing and diet conditions. Individuals with apparently similar dietary intake varied in weight and expression of phenotypes. Our results indicated that despite varying levels of obesity, the lineages differ markedly in the transmission of regulatory factors of obesity, glucose homeostasis, and insulin sensitivity.

Several studies now show changes in diet induce RNA variations also in the sperm cells (see recent review (Zhang et al., 2019)). The RNA content of sperm is complex (Kawano et al., 2012; Burl et al., 2018) and varies considerably with diet (Chen et al., 2016a; Chen et al., 2016b; Klastrup et al., 2019; Nätt et al., 2019). A comparison based on RNA-sequencing of sperm from founder males raised on a normal diet (ND) versus HFD confirmed variations in several transcripts, including small noncoding RNAs (sncRNAs) (Klastrup et al., 2019; Zhang et al., 2019). Previously, it had been observed that the total RNAs of spermatozoa obtained from males (founders) raised on HFD after microinjection into naive fertilized eggs induced the same pathological variation as that observed in the founders and was transmitted to offspring (Grandjean et al., 2015; Chen et al., 2016a).

In addition, Chen et al. (Chen et al., 2016a) identified purified RNAs from aqueous phase of total sperm enriched in transfer RNA fragments (tRNA) derived from 5′ tRNA as inducers of phenotypic changes. HFD-mediated phenotypes (HFD-MPs) are transferable simply by microinjection of RNAs (total or fractionated) from spermatozoa into naive fertilized mouse eggs. On the other hand, our laboratory studied the variations induced by food on the miRNAs of spermatozoa, one in particular (miR-19b) with increasing levels under HFD (Grandjean et al., 2015). In fact, microinjection of synthetic miR-19b into naive fertilized mouse eggs can transfer disease to offspring, thus showing a role of miRNA in the establishment of the HFD-MPs (Grandjean et al., 2015). In contrast, microinjection of synthetic tsRNAs is not sufficient to induce disease (Chen et al., 2016a; Zhang et al., 2018), indicating that the tsRNA fraction of sperm contains more information such as RNA modification (see below).

In addition to changes in the level of ncRNAs (miRNAs, piRNAs, cRNAs, and lncRNAs etc.), modification of RNAs such as C methylation is also involved in the induction and transfer of disease heritability. Initially, the Bestor's group (Goll et al., 2006) demonstrated that methylation of tRNAs at position 38 in $tRNA^{Asp}GUC$ occurs by the Dnmt2 protein. Subsequently, a group of RNA modifying enzymes has also been reported [see recent review for more information (Jeltsch et al., 2017)]. In addition, we have previously reported a role of the Dnmt2 protein in the transmission of epigenetic traits from one generation to another (Kiani et al., 2013). In fact, homozygous $Dnmt2^{-/-}$ mice fail to transmit the epigenetic variation in coat color initiated by alteration of *c-kit* transcripts (Kiani et al., 2013). In preparing this report, colleagues reported that the absence of the methyltransferase, Dnmt2, also abolished the transmission of obesity and diabetes to offspring (Zhang et al., 2018). According to this latest report, in the absence of the Dnmt2 protein, the 30–40 nt RNAs in the spermatozoa of obese animals do not show the elevation of the modifications (m^5C, m^2G) normally induced by HFD (Zhang et al., 2018). Thus, in addition to the changes in the proportions of transcripts in sperm, RNA modifications are also important for conveying hereditary information. However, it is not yet clear why RNA levels (tsRNAs fragments, miR-19b or other noncoding RNAs) vary depending on HFD and the underlying mechanism. Alterations in tRNA fragments (tRFs) have already been reported in several circumstances (stress, drugs, and food uptake), yet it is difficult to predict the specificity of tRFs in HFD-MPs, while miR-19b is a unique sequence in the mouse genome with multiple predicted mRNA targets in several normal tissues and the development of the disease (Li and YaoMicroRNAs, 2012).

On the other hand, we have shown that the sperm ncRNAs in the $Dnmt2^{-/-C57BL/6}$ inbred background are found at varying levels from male to male regardless of diet. These varying levels of ncRNAs in $Dnmt2^{-/-C57BL/6}$ mice are suspected to be produced by uncertain changes in RNA stability in the absence of Dnmt2. In addition to $Dnmt2^{-/-C57BL/6}$, we reveal here another line *B6/D2* hybrid mice (wild type) resistant to HFD. Organisms with hybrid genetics are known for their variation in the levels of transcripts (Botet and Keurentjes, 2020) and hybrid vigor commonly used in transgenesis (Clark et al., 2020). Thus, data from these two HFD resistant lines suggest a tempting explanation that predisposition to ncRNA variability, either by transcripts proportions or by stability, may allow unlimited possibilities of unsuccessful heritability.

Finally, we recently developed a direct protocol to detect DNA-associated RNA molecules in sperm, which reveals stable DNA/RNA hybrid regions in the genome as an R-loop structure. We performed sperm RNA preparations followed by RNA-sequencing combined with a bioinformatic analysis comparison between two fractions of RNA (attached to DNA versus free RNA molecules) in the same male (Kianmehr et al., 2019; Rassoulzadegan et al., 2021). Our results suggest changes in sperm RNA ratios between free and DNA-associated RNA fractions in the $Dnmt2^{-/-C57BL/6}$ genome compared to wild type. Hybrid DNA/RNA annotation in spermatozoa may represent a signal to be retained from life experiences.

Together, these results highlight the combinatorial effects of ncRNAs in the field of RNA-mediated epigenetic hereditary variation and its consequence on cross-generations.

RESULTS

Health and genealogical monitoring of four founder lines of mice (males) raised in a diet rich in fat

To answer the question of susceptibility to HFD (21% butter) among different genotypes and its transgenerational

maintenance, we maintained two inbred lines, *C57BL/6* and *Balb/c with B6/D2* hybrids (F1 cross between the *C57BL/6* and *DBA2* lines) and the $Dnmt2^{-/-C57BL/6}$ mutants. To minimize the effect from the environment, all experiments were conducted using animals born during the same week under the same housing and feeding conditions. From founder males (F0) raised in HFD, we derived two generations by sexual mating (F1 and F2). All mice other than F0 males were kept on normal diet (ND) and were followed by monitoring their body weight and health for up to 20 weeks. During this time, lineage-dependent differences of mice in test results for weight gain, glucose and insulin tolerance were observed. A summary of the experiments with four lines (*C57BL/6*, *Balb/c, B6/D2* and including $Dnmt2^{-/-C57BL/6}$) is presented in the Supplementary Figures: founder (F0) **Supplementary Figures S1–S4**, F1 generation **Supplementary Figures S5–S8**, F2 generation **Supplementary Figures S9–S12**.

Additionally, all glucose tolerance test (GTT) and insuline tolerance test (ITT) plots with area under the curve (AUC) were analyzed before and after normalization to baseline glucose levels (see the guidelines provided in PMID: 34117483). Analyses of AUC for GTTs and ITTs involve first normalizing the data (i.e. subtracting the initial glucose reading at time 0' from all subsequent readings). This is thought to be important, because a GTT and an ITT the response to a glucose and insulin boost, respectively. The calculated AUC, after normalizing data to baseline glucose levels, is proposed to reveal a much more subtle effect than usually reported without normalization. However, there are conflicting opinions on the application of AUC in chronically treated animals [see references (KC et al., 2018)]. Basically, both methods reveal the differences between animals maintained on HFD and ND. The GTT results are all in agreement but the difference between ND/HFD are more supported with the normalization. The ITT tests while keeping the differences between the two conditions (HFD vs. ND) had however a less pronounced difference with normalization.

Variations in test results (*C57BL/6*) for body weight gain, glucose, and insulin tolerance were noticeable in the F0 generation (**Supplementary Figure S4**) and are clearly perpetuated in the F2 generation, see **Supplementary Figures S8** (F1) and **Supplementary Figures S12** (F2) in *C57BL/6* according as previously reported (Grandjean et al., 2015). Glucose and insulin tolerance tests demonstrated prominent pathology in the founder animals and showed larger differences in a lineage-dependent manner in the F1 and F2 generations (**Supplementary Figures S5–12**). Two founder inbred lines *Balb/c* (**Supplementary Figure S1**) and *C57BL/6* (**Supplementary Figure S4**) are sensitive to HFD, and the mice show high fat diet mediated phenotypes (HFD-MPs) and transmit to the F1 and F2 generations. On the other hand, the founders of the *B6/D2* line (**Supplementary Figure S2**) and particularly $Dnmt2^{-/-\ C57BL/6}$ (**Supplementary Figure S3** and **Figure 1**) are resistant to HFD (no change in weight). The *B6/D2* F2 generation (**Supplementary Figure S10**) differs from that of the founder due to genetic segregation (visible on coat color). The founder's *B6/D2* hybrid animals (**Supplementary Figure S2**) are not affected in their weight and are unresponsive to GTT and ITT challenging, but the next generation develops HFD-MPs (**Supplementary Figure S6**, F1 and **Supplementary Figure S10**, F2). These results suggest that the initial signals that program the health of the offspring are perceived differently in the somatic and germ cells. The results on the $Dnmt2^{-/-C57BL/6}$ line are also shown in **Figure 1** and in **Supplementary Figure S3** (F0), **Supplementary Figure S7** (F1), and **Supplementary Figure S9** (F2). $Dnmt2^{-/-C57BL/6}$ mutants exhibit a defect in methylation detectable with highly expressed tRNA molecules, which disrupts their metabolism (Lyko, 2018). We have previously reported a role of methylation of RNA other than tRNA in the transmission of hereditary epigenetic signals (Kiani et al., 2013). Our results here confirm a role of the Dnmt2 protein in the hereditary modification of the HFD-MPs.

A summary of the number of mice used in **Supplementary Figures S1–3** can be found in the Supplementary Figures.

The $Dnmt2^{-/-C57BL/6}$ line did not develop complete metabolic diseases even when fed a diet of up to 33% butter fat

In addition to the above experiments with the $Dnmt2^{-/-C57BL/6}$ line, two independent experiments with $Dnmt2^{-/-C57BL/6}$ males are shown in **Figure 1**, one again as above with 21% and one at 33% fat diet. Raw data for each group of mice in **Figure 1** can be found in **Supplementary Table S1** and in individual raw data of **Figure 1**.

We examined the ability of $Dnmt2^{-/-C57BL/6}$ mutants to gain weight under the same conditions on a 21% butter (HFD) diet as previously described using a larger number of animals, 10 males and eight females; neither sex gained weight (see **Figure 1** for males and **Supplementary Figure S3** for females). Using a higher concentration of sugar and 33% butter (see Materials and Methods), we tested 12 males in each condition as shown in **Figure 1**. $Dnmt2^{-/-C57BL/6}$ mutant males did not gain weight. The physiology of the $Dnmt2^{-/-C57BL/6}$ mutant was different from that of wild-type mice, and even under 33% HFD, the mice did not get weight but are sensitive to glucose and insulin without symptoms. A higher concentration of up to 60% HFD has been reported previously (Zhang et al., 2018), and $Dnmt2^{-/-C57BL/6}$ mice at this higher concentration of HFD could eventually become obese but are unable to transmit this phenotype to the next generation.

Microinjection of total RNAs from spermatozoa of different mouse lines transfer obesity and diabetes in the susceptible Balb/C line

Another generation (G1) was obtained in **Figure 2** by microinjecting total sperm RNAs from different founders into fertilized mouse eggs (naïve *Balb/c*). We observed that lineage-dependent phenotypic variations in the disease spectrum (weight, GTT and ITT) were transferable by microinjection of total RNA from spermatozoa of *C57BL/6*, *B6/D2*, *Balb/c* and

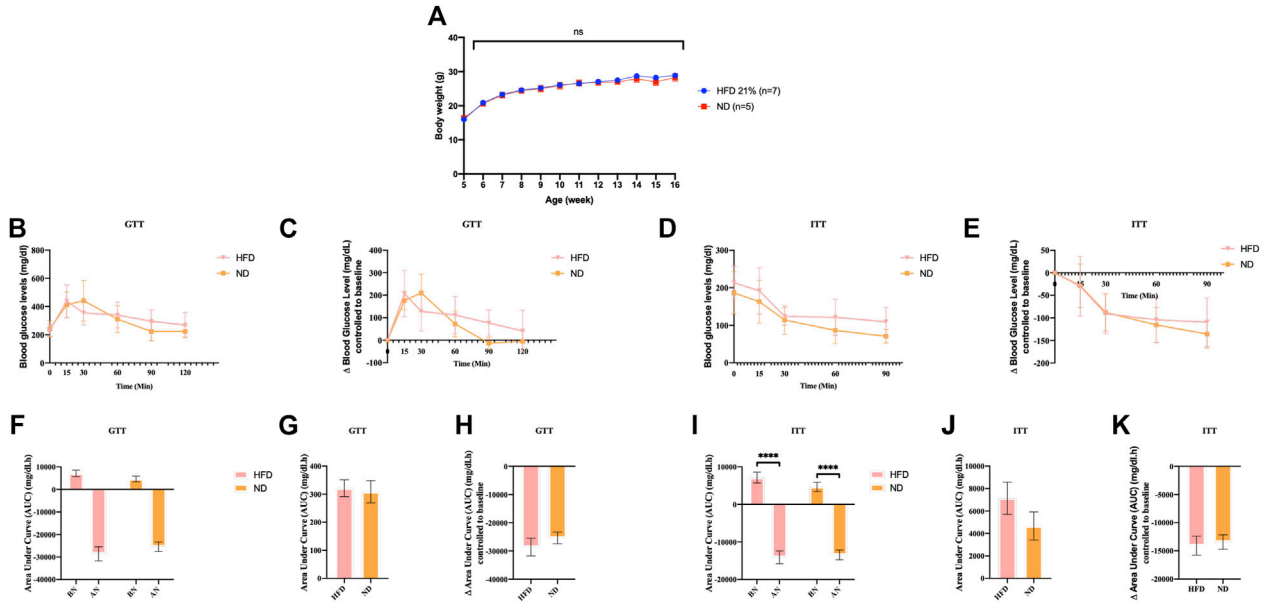

FIGURE 1 | Founder animals *Dnmt2*$^{-/-\ C57BL/6}$ are resistant to high-fat diet. Thirty-six *Dnmt2*$^{-/-C57BL/6}$ males (*C57Bl/6* background) from 20 litters weaned at 4 weeks old were divided into two groups. One group was maintained on an ND, and the other group was on HFD containing 21% **(Figure 1-1)** or 33% **(Figures 1-2)**, butter (see Methods) and monitored for body weight **(A)**, fasting blood glucose **(B,C)**, and insulin tolerance test **(D,E)** at 15 weeks. The area under curve (AUC) (Virtue and Vidal-Puig, 2021) **(F–K)** was also calculated after normalization based on the glucose levels of each time point to 0 point (HFD and control ND), before normalization (BN) and after normalization (AN). Statistical analysis was performed by two-tailed, two-way ANOVA, uncorrected Fisher's LSD. **p < 0.01, ****p < 0.0001 (ND vs. HFD), ns, not significant. All data are plotted as mean ± S.E.M. The raw data are presented in **Supplementary Table S1** and individual raw data in Supplementary Table for **Figure 1**.

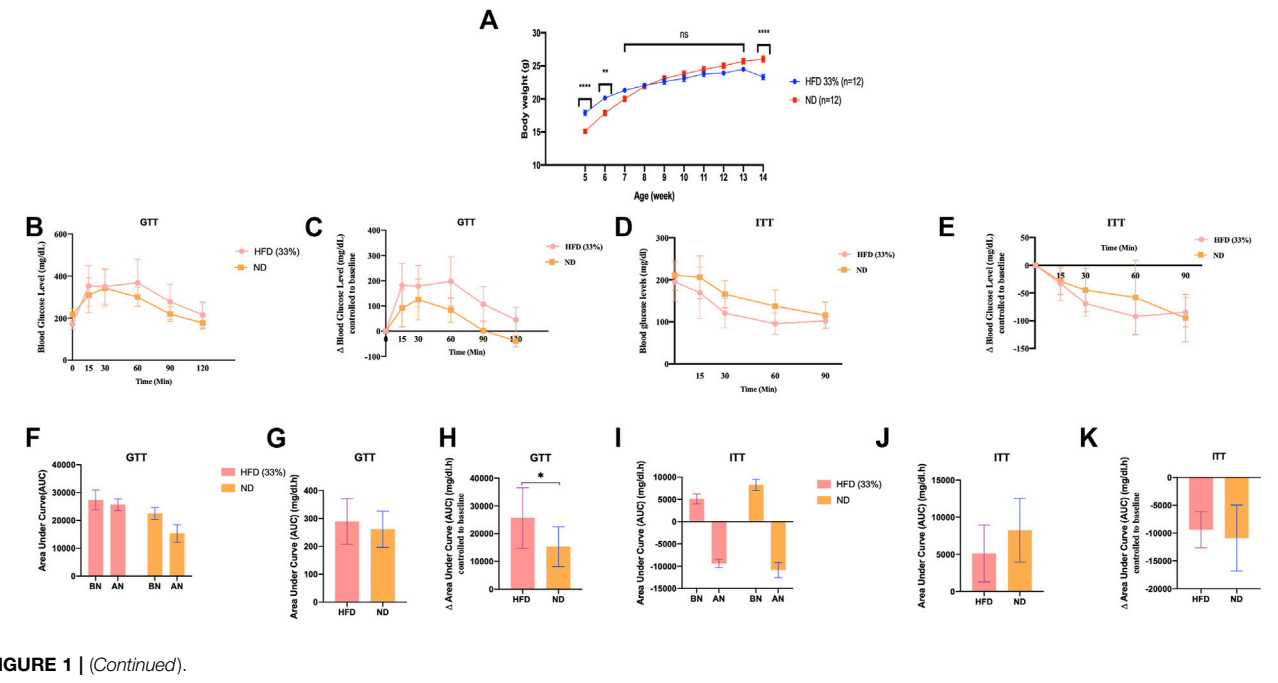

FIGURE 1 | (*Continued*).

Dnmt2$^{-/-C57BL/6}$ mice in fertilized eggs from naïve *Balb/c* mice. A higher effect is observed on female mice born after microinjection with sRNAs.

Thus, the total spermatic F0 RNAs of HFD-resistant founders, in particular the *Dnmt2*$^{-/-C57BL/6}$ and *B6/D2* hybrids, also contained RNAs that could recapitulate the metabolic

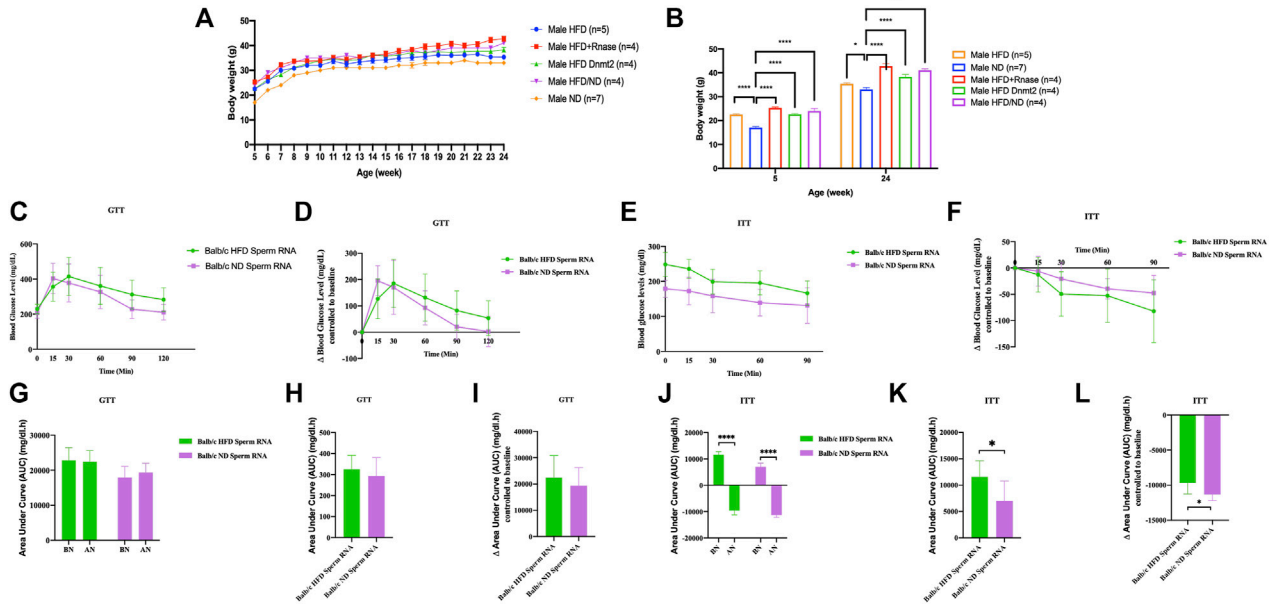

FIGURE 2 | Microinjection of total RNA of sperm from HFD-resistant males perturbs the health of susceptible offspring (Balb/c). Total sperm RNAs prepared from males raised on ND or HFD were microinjected into Balb/c fertilized eggs, as described in the Methods section. Microinjection performed with total sperm RNAs for the three lines, Balb/c, ND, HFD, Mixed ND/HFD and HFD RNA added RNase before microinjection and finally HFD Dnmt2-/-C57BL/6 RNA. n = number of mice in each group. All mice were raised on an ND. Body weights **(A, B)**, fasting blood glucose (GTT) **(C)** and insulin tolerance (ITT) **(D)** at 15 weeks are presented for males. The area under curve (AUC) 30 after normalization were also calculated based on the glucose levels of each time point **(E, F)** (Each group vs control ND). Females are shown in **Supplementary Figure S13**. Comparison with or without normalization of GTT **(G–I)** and ITT **(J–L)** challenges analysis. Statistical analysis was performed by two-tailed, one-way analysis of variance (ANOVA), uncorrected Fisher's least significant difference (LSD). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ ****p < 0.0001 (ND versus each group). Ns, not significant. All data are plotted as mean ± s.e.m. All raw data are presented in **Supplementary Table S2** and individual raw data in Supplementary Table for **Figure 2**.

phenotypes after microinjection into fertilized eggs of a susceptible line (*Balb/c*). We assume that unmethylated sRNAs from $Dnmt2^{-/-C57BL/6}$ are methylated by wild-type Dnmt2 protein present in fertilized *Balb/c* eggs. The dynamism of methylation remains to be assessed. These results again confirm the role of RNA methylation mediated by Dnmt2 protein in the transfer of hereditary acquired epigenetic traits. In addition, these results reveal signals from as yet unexplained sRNAs from the *B6/D2* hybrid line which required further investigation.

In addition, we microinjected RNase enzymes with sRNAs to initially abolish RNA inducing signals. The results of the RNase enzyme microinjection are however contradictory with the expected results. Under these conditions we believe that the RNase treatment removes more than the sRNAs at the time of microinjection into the fertilized eggs. We report this negative result to signal in future experiments that the RNase enzyme should be removed from the aliquot prior to the microinjection step.

Finally, co-injection with sperm RNAs from ND could compromise obesity (**Figure 2**). Previously, we reported that microinjection of total sperm RNA (aqueous phase free RNA fraction) from males raised on HFD into fertilized mouse eggs (one-cell embryos) transmitted HFD-MPs to the offspring. We investigated how to counterbalance the activity of HFD sperm RNAs. To examine this question, we combined founder sperm RNAs raised on HFD with sperm RNAs from animals raised on ND in a 1:1 ratio. As shown in **Figures 2**, **3**, ND sperm RNAs (sRNA-ND) affected the activity of sperm RNA (sRNA-HFD) of animals on HFD. In fact, the animals weighed less but maintained the difference throughout the growing period. However, the activity of HFD RNA is not completely suppressed by sRNA-ND.

Total RNAs of sperm from males raised on normal food withdraw miR-19b activity in early embryos

In studies reported from our laboratory, we have shown that microinjection of miR-19b-5p is sufficient to recapitulate HFD-mediated phenotypes (HFD-MPs) (Grandjean et al., 2015). Indeed, we have shown that a high number of miR-19b-5p molecules microinjected into fertilized mouse eggs leads to phenotype like HFD-MPs (see **Figure 3**). After microinjection into fertilized mouse eggs of a high copy number of miRNAs, most (>90%) of them are released from eggs within 24 h (our unpublished results). This high copy number scenario is however difficult to imagine and if to occur would be very rare *in vivo* except under extreme conditions. Under natural conditions this scenario does not occur for a given miRNA of spermatozoa from animals reared in HFD. It is unlikely that a small amount of changes in individual RNAs is immediately responsible for the complete phenotype. On the other hand, several ncRNAs

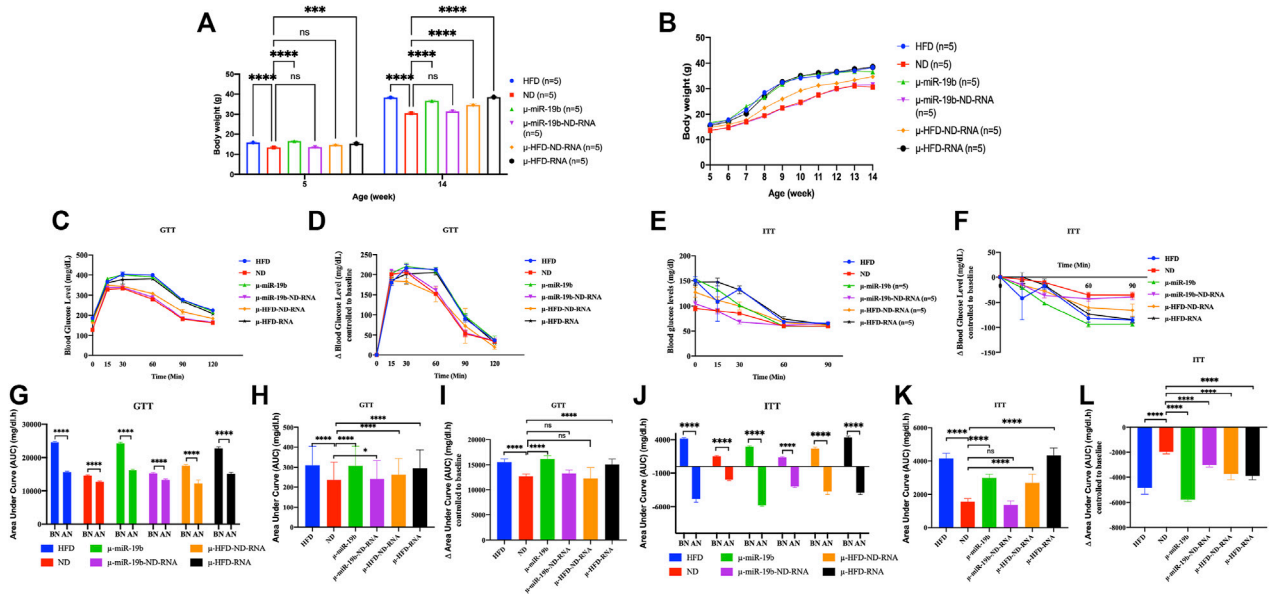

FIGURE 3 | Induction of obesity and diabetes in mice by microinjection of miR-19b-5p or total RNAs of sperm into fertilized mouse eggs. Microinjection of miR-19b-5p (synthetic RNA) into fertilized mouse eggs (one-cell embryos) leads to obesity and diabetes in derived adult mice. Here, we compared the activity of miR-19b-5p RNA alone or in combination with total RNA as carrier from spermatozoa from animals raised on ND. All male animals were fed an ND and monitored for body weight, fasting blood glucose (GTT) and an insulin tolerance test (ITT) at 15 weeks. Total RNA from spermatozoa from males raised on a normal diet (ND) suppresses the activity of miR 19b-5p in the induction of obesity and type 2 diabetic (T2D) phenotype. We also compared the activity of the total RNA of sperm from animals fed with HFD alone (1 g/ml) or in combination with total RNA from sperm (1 g/ml) as carrier from animals raised in ND. Total sperm RNAs from ND animal do not completely suppress the induction of HFD-Mediated Phenotypes (HFD-MPs) of total RNA from sperm from HFD animals but a reduction is observed in the expression of HFD-MPs. n = number of mice in each group. **(A)** Body weight of males born after microinjection of miR-19b-5p in each group at the age of 5 weeks to 14 weeks. Statistical analysis was performed by two-tailed, one-way analysis of variance (ANOVA), uncorrected Fisher's least significant difference (LSD). ****$p < 0.0001$ (ND vs. each group); ***$p < 0.001$, ****$p < 0.0001$. **(B)** Body weight of males in each group from 6–14 weeks of age. **(C)** Blood glucose during the glucose tolerance test (GTT). Statistical analysis was performed by two-tailed, two-way ANOVA, uncorrected Fisher's LSD. ****$p < 0.0001$ (ND vs. HFD); $^{\#\#\#\#}p < 0.0001$ (ND vs. μ-miR-19b); $^{\star\star\star}p < -0.001$ (ND vs. μ-HFD-ND-RNA), $^{\times\times\times\times}p < 0.0001$(ND vs. μ-HFD-RNA). The area under curve (AUC) (Virtue and Vidal-Puig, 2021) were also calculated after normalization based on the glucose levels of each time point **(D,F)**. **(E)** Relative blood glucose during the insulin tolerance test (ITT). Statistical analysis was performed by two-tailed, two-way ANOVA, uncorrected Fisher's LSD. ****$p < 0.0001$ (ND vs. HFD); $^{\#\#\#\#}p < 0.0001$, $^{\#}p < 0.01$ (ND vs. μ-miR-19b); $^{\star\star\star\star}p < 0.0001$, $^{\star\star\star}p < 0.0001$ (ND vs. μ-HFD-ND-RNA), $^{\times\times\times\times}p < 0.0001$(ND vs. μ-HFD-RNA). $^{\bullet}p < 0.01$, $^{\bullet\bullet\bullet\bullet}p < 0.0001$ (ND vs. μ-miR-19b-ND-RNA). ns, not significant. All data are plotted as mean ± S.E.M. All of the raw data in **Figure 3** are presented in **Supplementary Table S3** and individual raw data in Supplementary Table for **Figure 3**.

(microRNAs, piRNAs, siRNAs, cRNAs, tRNAs, lncRNAs, etc.) are also increased during HFD; as such, we expect a cumulative role of ncRNAs in the inheritance of diet-induced metabolic phenotypes. To answer the question of the specificity of miR-19b-5p, we used the same amount of miR-19b-5p (see Materials and methods) as that used previously, now in combination with total sperm RNAs (sRNA) as carrier from animals raised in an ND. As shown in **Figure 3**, the activity of miR-19b-5p was completely suppressed by sRNA-ND in the induction of metabolic disorders. With ND sRNAs, the induction of HFD-MPs pathways by miR-19b-5p is compromised. This indicates that a certain amount of miR-19b-5p is required, and here sRNA-ND carrier mixture (mRNAs, lncRNAs, sncRNAs, piRNAs, cRNAs, etc.) affects the activity of miR-19b.

Under these conditions sRNA-HFD mixed with sRNA-ND always maintains a detectable level of activity. RNA-mediated transfer of phenotype is confirmed by variation in the phenotype expression of weight gain and GTT and ITT tests (with or without normalization). Microinjection of a synthetic miR-19b is completely abolished by sRNA-ND; on the other hand, under the same conditions, sRNA-HFD maintains detectable levels of activity. We predict that synthetic miR-19b-5p is less stable than endogenous sRNA-ND or sRNA-HFD. The stability of RNA molecules would secure their activity.

miRNA-19 Expression profiles of spermatozoa from male (raised on HFD) compared to controls

From a previous study, we found a short list of deregulated miRNAs in the sperm of males raised on HFD. By quantitative RT-PCR analysis, we confirmed the deregulation of miR-182, miR-19a, miR-19b, miR-29a, and miR-340 in testis and sperm RNA of the HFD males compared to ND males. Therefore, we focused our study here on the analysis of one miRNA (miR-19b) given its HFD-MPs inducing activity. **Figure 4A** is a confirmation of our previous results (Grandjean et al., 2015) and shows that miR-19b-5p in sperm RNA samples was consistently up-regulated in males raised in HFD compared to ND in two inbred lines (*C57BL/6 and Balb/c*), while that in the *B6/D2* F1 hybrid miR-19b-3p is increased. In contrast, in $Dnmt2^{-/-C57BL/6}$ males, none of the mir-19b was increased under HFD conditions.

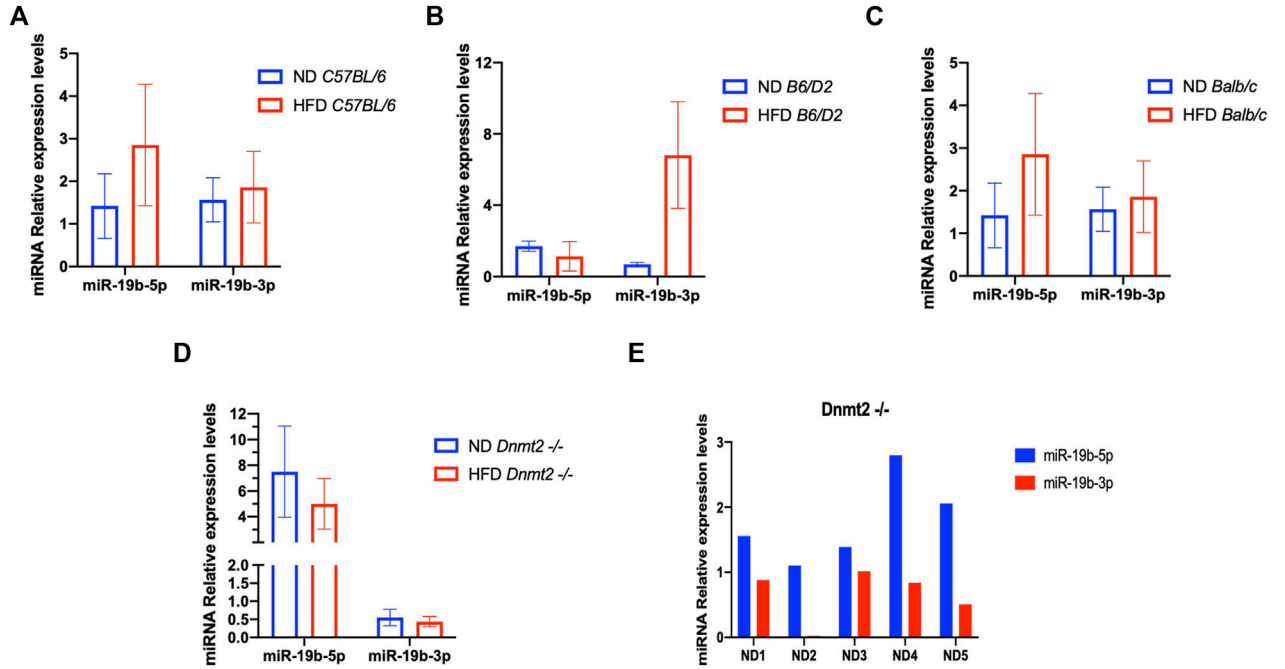

FIGURE 4 | Level of miR 19b in the total sperm RNAs of different lines founders mice (*C57BL/6*, *Balb/c*, *B6/D2*, and *Dnmt2*$^{-/-C57BL/6}$) fed with HFD (21% butter). **Figure 4** shows the expression levels of miRNA from four males in each group. By quantitative RT-PCR analysis, total sperm RNAs of two ND males compared to two HFD males, in *C57BL/6* **(A)**, *B6/D2* **(B)**, *Balb/c* **(C)**, *and Dnmt2*$^{-/-C57BL/6}$ **(D)**. **(E)** RT-PCR results for total sperm RNAs from male-to-male in five independent males in ND.

Supplementary Figure S14 shows deregulation of sperm miR-19b in *Dnmt2*$^{-/-C57BL/6}$ mice from male to male, regardless of food. To exclude effects of litter in the expression levels of miR-19 in sperm of *Dnmt2*$^{-/-C57BL/6}$ under ND, we tested five males from two separate litters and the results are shown from male to male in **Figure 4B**. We confirmed varying levels of miR-19 in sperm RNA of *Dnmt2*$^{-/-C57BL/6}$ mice in ND. This means that in *Dnmt2*$^{-/-C57BL/6}$ mice, an alteration of miRNA-19b already exists in males raised on ND, and this marker suggests sporadic changes in the proportion of miR-19b-5p and -3p in its sperm RNA. We also confirm, see below, by RNA-seq analysis with four different *Dnmt2*$^{-/-C57BL/6}$ RNA samples from sperm, the presence of other transcripts variations.

Sperm transcripts: analysis of free and DNA-associated RNA fraction reveal differences between Dnmt2$^{-/-}$ and control

To get an overall view of sperm RNA profiles, we addressed the issue of examining different sperm RNA fractions. In particular, the RNA fraction associated with the genome in hybrid DNA/RNA structures. Recently, we reported that with a conventional sperm RNA extraction protocol, two sperm RNA fractions could be obtained (Kianmehr et al., 2019). Unlike the classic free RNA fraction extracted with the TRIzol protocol in an aqueous phase (annotated total RNA extract), a sperm RNA fraction consisting of DNA/RNA hybrid molecules is closely associated with the genome (Kianmehr et al., 2019; Rassoulzadegan et al., 2021). The RNA fraction associated with the DNA in the phenol/chloroform interface phase is isolated and purified after treatment with the enzyme *Dnase*. The RNAs associated with the DNA and the free RNA fraction from the sperm are recovered in equivalent quantities, with the hypothesis that the comparison between the proportion of strongly maintained RNA with the DNA fraction on free RNA of the spermatozoa could reveal a change of the hereditary epigenetic marks.

Two populations of sperm RNA were isolated from the founder animals (*C57BL/6* wild-type and *Dnmt2*$^{-/-C57BL/6}$): free (F) and fraction associated with DNA (D) [for more details see Methods and references (Kianmehr et al., 2019; Rassoulzadegan et al., 2021)]. With deep RNA sequencing, we generated data (see **Figure 5** and **Supplementary Table S4** and **Supplementary Figure S15**). All sperm transcripts are present in both fractions (F and D).

To identify significant peaks, we searched for peaks throughout the genome, and a peak annotation process was used to relate the location of the peaks/regions to neighboring genes (see bioinformatic analysis methods). This analysis aimed to efficiently assign peaks in terms of important genomic characteristics. Regions of the genome were annotated by sperm RNA fractions of DNA-bound sperm RNAs versus free RNA. The analysis was performed to obtain information if a peak is found in the promoter TSS (transcription start site), 5′UTR, exon (coding), intron, 3′UTR, TTS (transcription termination site), and intergenic and non-coding regions. In normal mice, there was no difference between the number of counts in the

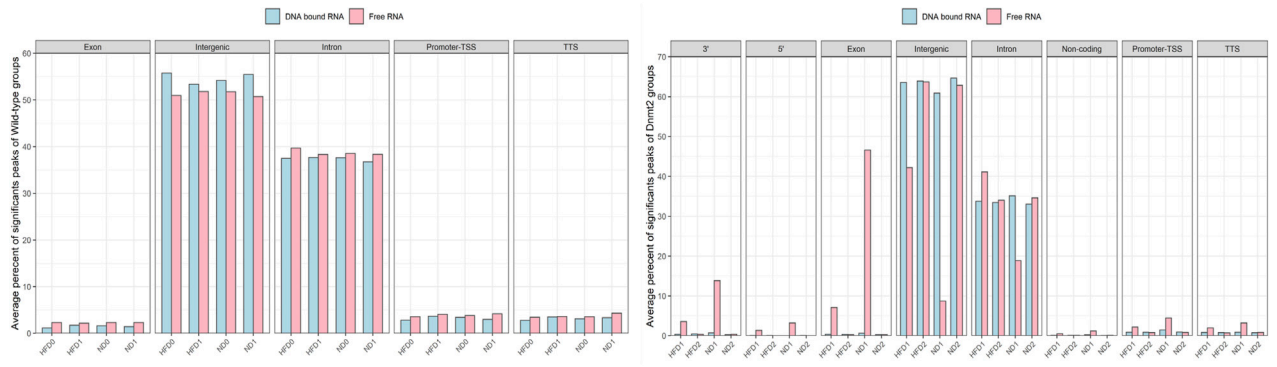

FIGURE 5 | Average percentage of significant peaks per genomic regions of RNA fractions (DNA bound and Free RNA) in spermatozoa from C57BL/6 and *Dnmt2*$^{-/-C57BL/6}$ maintained in high-fat diet (21% butter) and normal diet. Two RNA fractions (DNA-bound and free) are prepared from each male, four wild-type and four *Dnmt2*$^{-/-C57BL/6}$. For each group of samples, a fraction corresponds to the RNAs released from the sperm during extraction in the aqueous phase which we have called the free RNA fraction. Another fraction is released from the purification of genomic DNA after treatment with *Dnase* which we call DNA-bound fraction RNA. The two samples used as template for the RNA-seq analysis and the data were analyzed (see Methods). After obtaining sets of peaks, peak annotation was performed using Homer annotate peaks to relate each peak to nearby genes and genomic features. A false discovery rate (FDR) cut off 0.001 was used for identifying significant peaks. The positions annotated for promoter-TSS, exons, introns, and other features were based on the mm10 transcripts. Males on high-fat diet (HFD) and normal diet (ND), as well as for Dnmt2 samples.

samples of ND/HFD and the results are homogenous between the replicates of each sample group **Figure 5** (four samples with F and four D fractions, altogether eight RNA-sequences). This means that diet probably does not affect overall peak ratios. More peaks are enriched in F relative to D in the region of the promoter TSS, 5′UTR, exon, intron, 3′UTR, and TTS, and except for the intergenic, more peaks are observed in fraction D. However, in the samples (four samples with F and four D fractions altogether eight RNA-sequences) of *Dnmt2*$^{-/-C57BL/6}$ animals, the profiles are more complex, as a sporadic variation in all regions is observed between samples regardless of the diet conditions **Figure 5**.

The variations in the peaks of the intergenic region are particularly important, on the other hand, the average percentage of significant peaks of the intergenic region is higher in the samples Dnmt2$^{-/-C57BL/6}$ > 60% than in the samples with wild-type around 50%. In contrast, transcripts from intron regions are somewhat lower in Dnmt2$^{-/-C57BL/6}$ samples (35%) compared to normal animals (38%) regardless of diet conditions (**Figures 5A,B**). This may suggest that RNA signals induced by HFD should be sought among RNAs derived from intron regions and resistance to diet in intergenic regions.

In addition, transcripts are more abundant as expected in free fractions than in those associated with DNA in all samples. Evaluation of sperm transcripts here by deep sequencing revealed that the amount of transcripts is generally decreased in mature sperm of ND/HFD *Dnmt2*$^{-/-C57BL/6}$ compared to wild-type sperm (**Supplementary Table S4**).

The loss of two cytosine-5 RNA methyltransferases (Dnmt2 and NSun2) has been reported to be associated with reduced rates of protein synthesis (Tuorto et al., 2012) and genome integrity (Genenncher et al., 2018). In addition, methylation of tRNA at the 38th position cytosine improves the stability of certain tRNAs under altered environmental conditions (Schaefer et al., 2010) or improves the viability of the virus in host cells (Dev et al., 2017). What`s more Dnmt2 activity is required for the establishment of an inherited epigenetic change (altered RNA levels) with "paramutation in c-kit loci" in mice (Kiani et al., 2013). Finally, whereby Dnmt2 is maintained throughout evolution, a role in adaptation to a new environment is assigned (Ashapkin et al., 2016). It is also found that tRNA fragments are in higher levels in the sperm of HFD males (Peng et al., 2012). Small tsRNA fragments of 30–34 nucleotides are derived from the 5′ end of tRNAs. In *Dnmt2*$^{-/-C57BL/6}$ sperm; the profile of tRNA-derived sncRNAs and rRNA-derived sncRNAs is not altered with HFD (Zhang et al., 2018). In other words, in the absence of the Dnmt2 protein, the levels of 30–40 nt RNA fractions in spermatozoa are not changed by HFD.

Here, we focus on the set of abundant mouse transcripts presented in **Figure 6**, the AY036118 gene (ETS-linked transcription factor *Mus musculus* ERF, Erf1), Gm42418, and Gm26917 (pre-ribosomal 45S RNA, serving as a precursor for rRNA 28, 18 and 5.8S). The transcripts are visualized using the UCSC genome browser. The expression signal is very high and comparable between samples according to the scaling of the data view on the left (**Figure 6**), and overlaps with repetitive elements. The pre-ribosomal 45S RNA transcript profiles in the free fraction of *Dnmt2*$^{-/-}$ show changes compared to the wild-type samples (**Figure 6**). We reveal sporadic changes in the levels of transcripts in the F in *Dnmt2*$^{-/-C57BL/6}$ compared to wild-type sperm. In addition, **Figure 6** shows variation in the expression signal of *Mus Musculus* leucyl-tRNA synthetase, mitochondrial (Lars2), transcript variant 1 on chromosome 9 of *Dnmt2*$^{-/-C57BL/6}$ to wild-type *C57BL/6* in HFD/ND.

Unlike the wild type (four samples), each biological replicate of the *Dnmt2*$^{-/-C57BL/6}$ samples exhibits a unique profile and shows variation in transcripts among the four samples. The unpredictable state of transcript levels from *Dnmt2*$^{-/-C57BL/6}$ animals would suggest altered RNA stability as a cause of

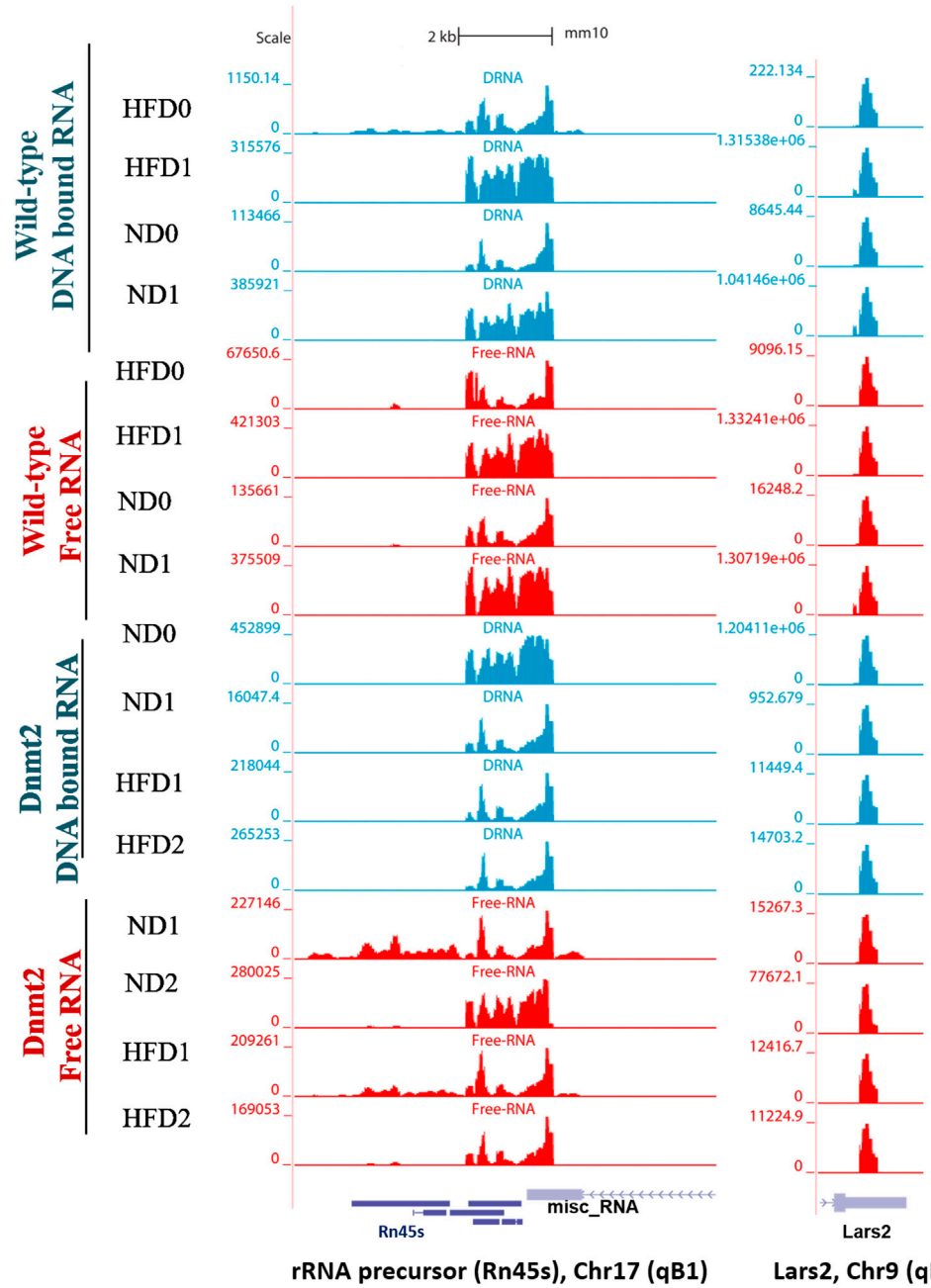

FIGURE 6 | Visualization of highly-expressed signals of DNA-bound and free-RNA moleculees in GRCm38/mm10 in the sperm of wild- type C57BL/6 and Dnmt2-/- mice line in ND and HFD (21% butter, high fat diet). Expression signal tracks of different highly expressed transcripts are visualized which are demonstrated at the bottom of each panel.

variation in *Dnmt2*$^{-/-C57BL/6}$ sperm transcripts. Our working hypothesis is that the loss of the Dnmt2 protein triggers random changes in the overall stability of the transcripts and thus influences the half-life dynamics of DNA/RNA hybrids, which may be one of the causes of loss of acquired inheritance in *Dnmt2*$^{-/-C57BL/6}$ line.

The snapshot of the genome browser of pre-rRNA transcripts and leucyl-tRNA synthetase, mitochondrial (lars 2) of normalized BigWig files of two types of sperm RNA fractions (DNA bound RNA and free RNA), shows a normalized expression RNA-seq signal in two mouse lines, wild-type and *Dnmt2*$^{-/-C57BL/6}$ in normal diet (ND) and 21% butter (HFD) through generations. The transcript set comprising the mouse gene AY036118 (transcription factor linked to the ETS of Mus musculus ERF (Erf1), Gm42418, and Gm26917 (A 45S pre-rRNA transcripts) serves as a precursor for 28, 18, and 5.8S rRNA on chromosome

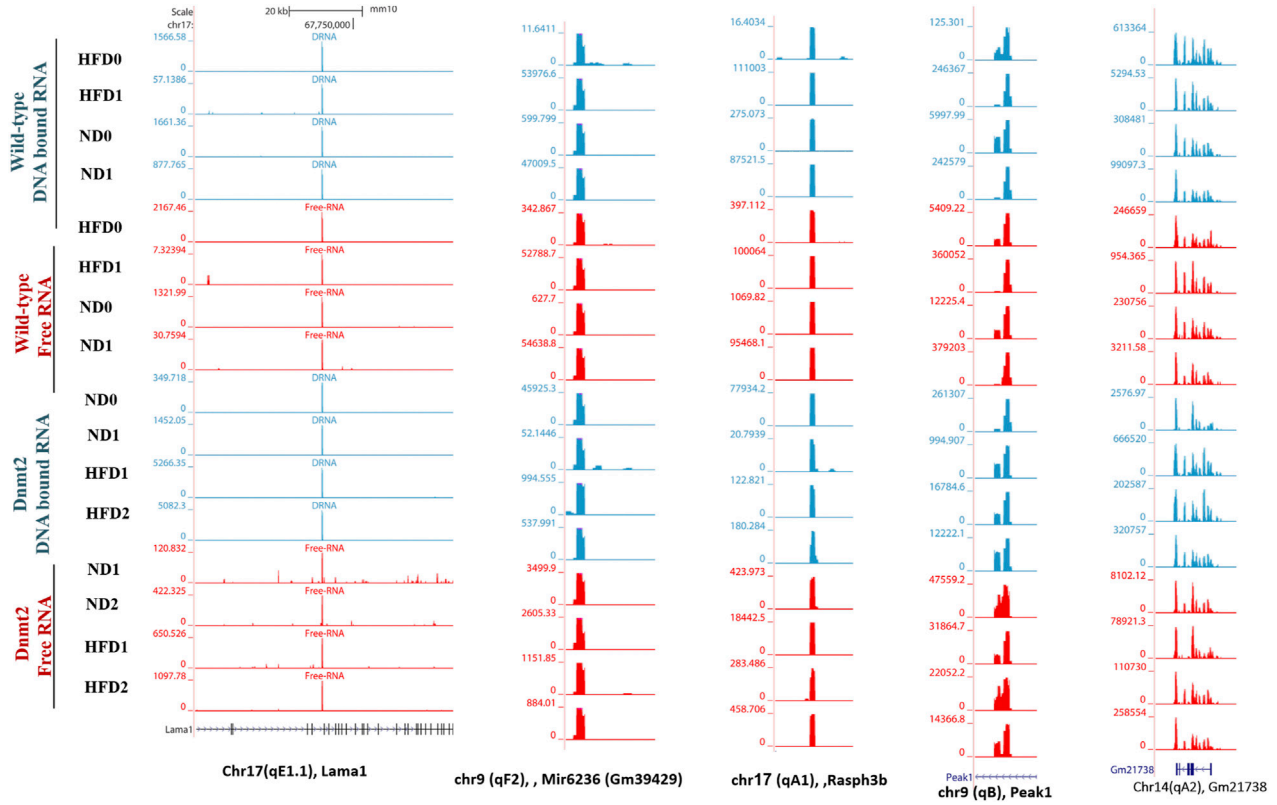

FIGURE 6 | (*Continued*).

17, overlapping elements repeated by Repeat Masker on the mouse assembly (GRCm38/mm10).

DISCUSSION

The complexity of the genetic contribution to a common phenotype of weight gain with a diet high in sugar and fat poses many challenges. One question raised here is whether the RNA alterations produced in the sperm of laboratory mouse lines reared on the same HFD (high-fat diet) vary from line to line. In other words, what are the common signals (RNAs) produced for the transmission of high-fat diet-mediated phenotypes (HFD-MPs) between different genotypes? Here, we show that RNA variations are produced in male mice maintained on HFD and are transported by sperm regardless of the genetic background, at least in the four mouse lines tested.

The data indicate that four laboratory lines (*C57BL/6*, *B6/D2*, *Balb/c* and *Dnmt2*$^{-/-C57BL/6}$) nevertheless under the same conditions respond differently in the expression of HFD-MPs and that all founders even with the wild-type genome could be resistant, like the *B6/D2* hybrid, and yet pass the changes on to their offspring. Two of the four founder mouse lines do not develop HFD-MPs as indicated previously for the *Dnmt2*$^{-/-C57BL/6}$ mice, and here the hybrid *B6/D2*. *C57BL/6* is a commonly used inbred mouse line that responds effectively to HFD by increasing fat and body weight, being insulin-resistant (Type 2 diabetes-like), and transmitting all metabolic alterations to at least two generations as previously indicated (Winzell and Ahre´n, 2004). *Balb/c* (founder) is another laboratory inbred line which, in HFD condition that does not increase body weight but develops Type 2 diabetes-like phenotypes and passes on HFD-MPs to the F1 and F2 generations. Thus, with *C57BL/6* and *Balb/c* two inbred lines maintained in the same homing, the same food and the same conditions respond differently to HFD. *B6/D2* is a hybrid between *C57BL/6* and *DBA2*; under the same feeding and environmental conditions as the above inbred lines, founder males appears to be resistant to HFD, but in their offspring certain aspects of the phenotype are found as predicted by genetic segregation. Transmission of the phenotype from *B6/D2*-resistant male founders in backcrossing is reminiscent of genetic anticipation (Carpenter, 1994; Martinez-Delgado et al., 2011). This means that despite the founders' *B6/D2* resistance to HFD, their germ cells have retained a diet-induced change which is then passed on to the next generation. In the F1 generation from *B6/D2* (F0 founder), the Type 2 diabetes-like phenotype is detectable, and in the F2 generation with different genotypic segregation, obesity is displayed. This means that the body weight and Type 2 diabetes-like characters are separately segregating. Finally, a *Dnmt2* null mutation in the *C57BL/6* background abolished the sensitivity to HFD (21–33% butter), and the animals do not gain weight and do not exhibit Type 2 diabetes-like traits, but unlike

B6/D2 the transmission of obesity and diabetes in the offspring is also well abolished. Likewise, using a higher concentration of butter (33% fat), the $Dnmt2^{-/-C57BL/6}$ founder mice maintained the same weight, but without symptoms, profiles in both conditions glucose and insulin tolerance tests are changed (more pronounced changes are visible by normalization of the data, **Figure 1**). Based on previous reports (Zhang et al., 2018), with a diet with a higher concentration of butter (60%), $Dnmt2^{-/-}$mice become also obese thus exceeding the controls by the Dnmt2 protein. The Dnmt2 protein methylate tRNAs (Lyko, 2018), and these results demonstrate that the involvement of RNA methylation is an important parameter in the control of health (such as weight and type 2 diabetes like phenotypes) (Lyko, 2018). RNA methylation is a means of modulating RNA stability and enhancing a phenotype (Jeltsch et al., 2017; Lyko, 2018).

Taken together, these results suggest that an additional mechanism governs the development of phenotypes in addition to genetic mechanisms, and this mechanism persists during transfer to the next generation.

Heterogeneities in transcripts could be a way to escape the disease

Ribosomal heterogeneity is reported in several studies with environmental stresses and is generally accepted as a source of differential ribosome selectivity for the translation circuits of gene expression (Xue and Barna, 2012; Shi et al., 2017). In addition to the protein composition and multiple modifications, heterogeneity is also observed in the composition of ribosomes and ribosomal RNA (rRNA) as diversity generates an impact on translational activities. HFD and protein restriction in mice also affects the levels of ncRNAs in mature sperm, including increased amounts of 5′ glycine tRFs (Cassman et al., 2016). Additionally, here we report the heterogeneity of the levels of not only miR-19b in $Dnmt2^{-/-C57BL/6}$ mice but also the activity of miR-19b which is compromised by mixing with sperm RNA from animals raised in ND. This result indicates that overall RNA activities are modifiable by methylation or compensation with total sperm RNAs. Likewise, total RNAs from ND sperm modulates active RNAs from HFD sperm when co-injected in the mixed buffer into fertilized mouse eggs. Changes in the distribution of RNAs could be responsible for changes in activities among sperm RNAs from ND to HFD. The case of the heterogeneous distribution of miR-19b transcripts levels in $Dnmt2^{-/-C57BL/6}$ animals indicates that variation is one of the factors which globally influence regulation and signal transmission. These heterogeneities are not specific for miR-19b but are also observed for the highly expressed transcripts revealed here by the analysis of the peak of the two RNA fractions in samples of sperm RNAs from $Dnmt2^{-/-C57BL/6}$ mice. In addition, abundant transcripts which were previously reported to be altered in $Dnmt2^{-/-C57BL/6}$ mice (Chen et al., 2016a; Zhang et al., 2016) show more dispersed profiles in the free fraction compared to wild-type. The changes in the levels of the free RNA fraction (F) relative to RNA attached to the genome (D) revealed here strongly suggests that hybrid DNA/RNA signals are important in the transmission of epigenetic information from generation to generation.

The free RNA fraction of sperm from four lines induces HFD phenotypes described here. The DNA-bound RNA fraction is not yet tested *in vivo*; since global transcripts are also present in both (the free and DNA bound fraction), it is expected that both fractions are active in microinjection tests in fertilized mouse eggs. We suggest that the activities of RNAs could be dynamic and dependent on the half-life of each RNA as DNA-bound or circulating molecules. The robustness and mechanisms underlying these hereditary changes will require further investigation.

METHODS

Mice

All animals (mice) and experiments were carried out in compliance with the ARRIVE guidelines (https://arriveguidelines.org). Mice were maintained in building with accord to European and international guidelines and policies in accordance with the animal welfare laws in Turkey in an animal facility at Genkok Institute, Kayseri, Turkey. The experiments were approved by the Turkish ethics committee (see file number below). 1) Institutional and/or licensing committee Erciyes University, Kayseri, Turkey; 2) we confirm all experiments were performed in accordance with relevant guidelines and regulations with permission from Erciyes University committee in an animal facility at Genkok Institute, Kayseri, Turkey. 3) All mice were maintained according to European and international guidelines and policies according to the animal welfare laws in Turkey in an animal facility at Genkok Institute, Kayseri, Turkey.

All protocols in this study were approved by the Committee on the Ethics of Animal Experiments of Erciyes University, Kayseri, Turkey, on Animal Care, permit number: number 13/58, date: 13.03.2013 and permit number 14/116, date: 13.08.2014 from the committee on the Ethics of Animal Experiments of Erciyes University, in compliance with the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH publication no.85-23, revised 1996).

The mice were maintained in a facility under controlled conditions (light from 06:00 to 18:00, 22°C temperature, 55% humidity). The animals were cared for and treated according to the Principles of Laboratory Animal Care (European rules). All tests were performed between 10:00 and 16:00 in isolated rooms.

Four lines of mice were maintained: *C57BL/6*, *B6/D2*, *Balb/c* and $Dnmt2^{-/-C57BL/6}$ (Charles River, Wilmington, MA, USA). After weaning, 4-week-old males were divided into two groups. One group was maintained on a normal diet (ND), a standard chow diet (7% simple sugars, 3% fat, 50% polysaccharide, 15% protein (w/w), energy 3.5 kcal/g) and the other was fed a Western diet containing, in addition, 17% casein, 0.3% DL-methionine, 34% sucrose, 14.5% cornstarch, 0.2% cholesterol, 5% cellulose, 7% CM 205B, 1% vit200, and 21% or 35% butter (Diet Western 1,635, Safe, Augy; France).

Glucose tolerance and insulin tolerance tests

Overnight fasted mice were injected with glucose or insulin in saline buffer. The baseline glucose measurements were analyzed

from tail blood before intraperitoneal (ip) glucose injection (2 mg/g body weight) using the OneTouch Vita (LifeScan, Johnson & Johnson, Milpitas, CA, USA) system.

Insulin (NovoRapid®, Novo Nordisk, Bagsværd, Denmark) diluted in 0.08 mU/µl saline was injected to a final delivery of 0. 8 mU/g body weight.

Blood glucose measurements were taken using tail blood at 0, 15, 30, 45, 90, and 120 min after injection.

RNA preparation and microinjection

RNA from purified sperm (centrifugation) was extracted using TRIzol reagent (Invitrogen, Life Technologies). All total RNA preparation from mouse sperm were treated with proteinase K 400 microg and Dnase 10–100 ng. The quality of the RNA preparations was verified by spectrometry on an Agilent Bioanalyzer 2,100 apparatus (Agilent, Santa Clara, CA, USA) (Grandjean et al., 2015). Microinjection into fertilized eggs was performed according to established transgenesis methods. Sperm RNA solutions were adjusted to a concentration of 1 µg/ml, and 1–2 pl was microinjected into the male pronuclei of fertilized eggs isolated after normal ovulation and mating to a *Balb/c* strain. Synthetic miRNA (purchased from Sigma, St. Louis, MO, USA) was prepared in a filtered microinjection buffer (10 mM Tris, pH 7.4; 0.1 mM EDTA) at a concentration of 3-5 × 10 (Pomp et al., 2008) molecules/pl.

RT-PCR and quantitative RT-PCR

miRNA quantitation was performed using miScript Primer Assays (Qiagen, Hilden, Germany), following the manufacturer-recommended protocols. Real-time PCR was performed with a Light Cycler Instrument (Roche Diagnostics, Indianapolis, IN, USA) using the miScript SYBR Green PCR kit (Qiagen, Hilden, Germany).

Isolation of RNA in R-loop structure as DNA/RNA hybrid for RNA sequencing

A classic TRIzol protocol produces two fractions of RNAs: one in aqueous phase (free fraction) and the other at the junction of the phenol-chloroform DNA fraction. The DNA fraction consists of DNA/RNA hybrid molecules and RNAs which are tightly associated with the genome (Kianmehr et al., 2019; Rassoulzadegan et al., 2021). The RNA fraction associated to DNA is then released from the column by *Dnase* enzyme treatment. Here, we have released sperm cells from epididymis and washed and separated from somatic cells by centrifugation and overnight incubation at 56°C in Tris buffer 20 mM pH8, EDTA 50 mM, with 0.5% SDS, 20 µM dithiothreitol, and 400 µg/ml Proteinase K. Total nucleic acids after enzymatic removal of the proteins was followed by fractionation of RNA and DNA-bound RNA. Extracts were fractionated by the ZR-Duet™ DNA/RNA MiniPrep Plus protocol according to the specifications of the manufacturer (www.zymoresearch.com). After digestion by DNase, 10–100 ng of RNA was prepared and sent to Eurofins for high-throughput sequencing on Illumina Hiseq 2,500 or Illumina MiSeq. Libraries of all samples corresponding to DNA-associated RNA and free-RNA molecules were generated. Small RNA high-throughput sequencing was performed by Eurofins (Europe). Number of mapped reads are summarized in **Supplementary Table S4**. Considerably, the map rate of $Dnmt2^{-/-C57BL/6}$ samples is lowered than HFD/ND samples.

Bioinformatics analysis

FastQC version 0.11.5 was used to do some quality control checks on the RAW sequence data. RNA-seq reads adapters removed by cutadapt v1.16. Based on the FastQC results, trimming of bad-quality reads was performed using Trimmomatic version 0.36 (10 nucleotides were cropped from the 5′ end of each read, bases were trimmed with Phred score lower than 20 from heading and trailing of each read, and the trimmed reads with a size less than 30 bp were removed) (Bolger et al., 2014). Transcript quantification was performed using Salmon v0.12.0 (Patro et al., 2017) and imported into tximport for downstream analysis. High enriched transcripts were visualized to show difference of expression in HFD/ND and $Dnmt2^{-/-C57BL/6}$ samples.

To align trimmed reads to the reference assembly (GRCm38) Hisat2 was used (Pertea, 2016 #10321). Quality of alignment was assessed using stats and plot-bamstats utilities of samtools (Li, 2011 #10323). Bam files were converted into BigWig format using deep tools bamCoverage with the following parameters: -of = bigwig - 50- normalizeUsing RPKM. Visualization of normalized BigWig files were performed using the UCSC genome browser (http://genome.ucsc.edu) (Kent et al., 2002).

HOMER version 4.9 was used for peak finding {Heinz, 2010 #10325}. Statistically significant peaks of expression were identified using HOMER (10,000 size of region used for local filtering, fourfold over local region, Poisson p-value over local region <0.0001, false discovery rate (FDR) rate threshold (cut off) < 0.001). After obtaining sets of peaks, peak annotation was performed using Homer "annotate_peaks" to relate each peak to nearby genes and genomic features. Homer (annotate_peaks) is widely used to assign peaks to positions annotated for gene, exon, intron, promoter, 5′ untranslated region (UTR), 3′ UTR, and other genomic features. The promoter-TSS, exons, introns and other features were based on the mm10 transcripts.

Statistical analysis

Statistical analysis was performed by two-tailed, one-way analysis of variance (ANOVA), uncorrected Fisher's least significant difference (LSD). Data are expressed as the mean with sem. $p < 0.05$ was considered to be statistically significant.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/geo/, GSE166636.

ETHICS STATEMENT

All protocols in this study were approved by the Committee on the Ethics of Animal Experiments of Erciyes University Kayseri

Turkey on Animal Care, permit number: number 13/58, date: 13.03.2013, and permit number 14/116 date:13.08.2014 from Erciyes Üniversitesi Hayvan Deneyleri Yerel Etik Kurul, in compliance with the Guide for the Care and Use of Laboratory Animals published by the US National Institutes of Health (NIH publication no.85-23, revised 1996)." Written informed consent was obtained from the owners for the participation of their animals in this study.

AUTHOR CONTRIBUTIONS

Mice experiments were performed by GS, EM, ET, and KK. miRNA expression was performed by EM; MM, figures. Bioinformatic analysis for RNA-Seq were performed by LK. Data analysis was performed by EM and LK. ST and FK provided help to start the project. This study was conceived and conducted by MR. The project and manuscript was written by MR. Data analysis was performed by EM, ZY, and LK.

FUNDING

This work is supported by Gant 2019–2020 of La Fondation Nestlé France to MR. In addition, this project was supported by Sanofi-Aventis and Turkish Diabetes Foundation 07-07-2014 to GB (grant no. 275).

ACKNOWLEDGMENTS

We are grateful to Yusuf Ozkul for open access to all of GenKok Institute, Kayseri, Turkey. We are indebted to Zeynep Soyer Sarıca for the initial setup of the animal facility at the Erciyes University of Kayseri, Turkey. We are thankful to animal facilities team of Genkok Institute, Celalettin Goktas, and to Fatma Aybuga for technical assistance. We thank K. Marcu for English editing of the last version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.839841/full#supplementary-material

Supplementary Figure 1 | Founder males *Balb/c* raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose, (GTT) and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 2 | Founder males *B6/D2* raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 3 | Founder males $Dnmt2^{-/-C57BL/6}$ raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 4 | Founder *males C57BL/6* raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 5 | *F1 generation* males *Balb/c* whose father was raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 6 | F1 generation males *B6/D2* whose father was raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 7 | *F1* generation males $Dnmt2^{-/-C57BL/6}$ whose father was raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 8 | *F1 generation males C57BL/6* whose father was raised with high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 9 | *F2 generation* males *Balb/c* whose grand-father was raised in high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 10 | *F2 generation* males *B6/D2* whose grand-father was raised in high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 11 | *F2 generation* males $Dnmt2^{-/-C57BL/6}$ *F2* whose grand-father was raised in high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 12 | *F2 generation* males *C57BL/6* whose grand-father was raised in high-fat diet of 21% butter (HFD); A: Body weight, B: glucose (GTT), and C: insulin (ITT) tolerance test (mg/dl).

Supplementary Figure 13 | Microinjection of total RNA of sperm from HFD-resistant males perturbs the health of susceptible offspring (*Balb/c*). Total sperm RNAs prepared from males raised on ND or HFD were microinjected into *Balb/c* fertilized eggs, as described in the Methods section. All mice were raised on an ND. Body weights at 15 weeks are presented for females born after microinjection of total sperm RNAs for the three lines, *Balb/c*, ND, HFD, Mixed ND/HFD, and HFD RNA added RNase before microinjection and finally HFD $Dnmt2^{-/-C57BL/6}$ RNA. *n* = number of mice in each group. Body weight is shown in each group from 6 to 14 weeks of age for females, respectively. Statistical analysis was performed by two-tailed, one-way analysis of variance (ANOVA), uncorrected Fisher's least significant difference (LSD). *p < 0.05, **p < 0.01, ***p < 0.001 ****p < 0.0001 (*ND* vs. each group). Ns, not significant. All data are plotted as mean ± S.E.M.

Supplementary Figure 14 | miR-19 expression in sperm of individual males. Each line corresponds to the results obtained from a male (repeated with three aliquots). The miR-19b-5p and -3p were systematically dysregulated in all males raised on HFD, but not in $Dnmt2^{-/-C57BL/6}$ males. In the latter, we found that miR-19b deregulation was heterogeneous between males, irrespective of the food type. Diet does not affect the level of miR19b in the sperm of $Dnmt2^{-/-C57BL/6}$.

Supplementary Figure 15 | RNA fractions (D and R) in the sperm of wild-type males in normal diet (ND) and 21% butter (high fat diet, HFD). Rn18s-rs5, n-R5s100 in sperm samples (13A) ND/HFD, wild-types, and (13B) $Dnmt2^{-/-C57BL/6}$.

Supplementary Table 1 | (Raw data of Figure 1) (A) Body weight (g), glucose (GTT), and insulin (ITT) tolerance test (mg/dl), males $Dnmt2^{-/-C57BL/6}$ (21% butter high fat diet). (B) Body weight (g), glucose (GTT), and insulin (ITT) tolerance test (mg/dl), males $Dnmt2^{-/-C57BL/6}$ mice (33% butter high fat diet).

Supplementary Table 2 | (Raw data of Figure 2) Body weight (g), glucose (GTT), and insulin (ITT) tolerance test (mg/dl) *Balb/c* animals born after microinjection with total sperm RNAs.

Supplementary Table 3 | (Raw data of Figure 3) Body weight (g), glucose (GTT), and insulin (ITT) tolerance test (mg/dl), *C57BL/6* F0 males born after total sperm RNA and miR-19b microinjection. Data were analyzed via one-way ANOVA test and the differences among the groups were determined by Tukey's post hoc test. There are no significant differences among the groups ($p > 0.05$). All results are expressed as mean and SEM.

Supplementary Table 4 | RNA-seq of fractions (D and R) in the sperm of two different mouse lines (*C57BL/6* and $Dnmt2^{-/-C57BL/6}$) in normal diet and 21% butter (HFD).

REFERENCES

Ashapkin, V. V., Kutueva, L. I., and Vanyushin, B. F. (2016). [Dnmt2 Is the Most Evolutionary Conserved and Enigmatic Cytosine DNA Methyltransferase in Eukaryotes]. *Genetika* 52, 269–282. doi:10.1134/s1022795416030029

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Botet, R., and Keurentjes, J. J. B. (2020). The Role of Transcriptional Regulation in Hybrid Vigor. *Front. Plant Sci.* 11, 410–419. doi:10.3389/fpls.2020.00410

Buettner, R., Schölmerich, J., and Bollheimer, L. C. (2007). High-fat Diets: Modeling the Metabolic Disorders of Human Obesity in Rodents*. *Obesity* 15, 798–808. doi:10.1038/oby.2007.608

Burl, R. B., Clough, S., Sendler, E., Estill, M., and Krawetz, S. A. (2018). Sperm RNA Elements as Markers of Health. *Syst. Biol. Reprod. Med.* 64, 25–38. doi:10.1080/19396368.2017.1393583

Carpenter, N. J. (1994). Genetic Anticipation: Expanding Tandem Repeats. *Neurol. Clin.* 1, 1. doi:10.1016/S0733-8619(18)30071-9

Cassman, K. G., Sharm, U., Conine, C. C., Shea, J. M., Boskovic, A., Derr, A. G., et al. (2016). Biogenesis and Function of tRNA Fragments during Sperm Maturation and Fertilization in Mammals. *Science* 351(6271):391–396. doi:10.1126/science.aad6780

Chen, Q., Yan, M., Cao, Z., Li, X., Zhang, Y., Shi, J., et al. (2016). Sperm tsRNAs Contribute to Intergenerational Inheritance of an Acquired Metabolic Disorder. *Science* 351, 397–400. doi:10.1126/science.aad7977

Chen, Q., Yan, W., and Duan, E. (2016). Epigenetic Inheritance of Acquired Traits through Sperm RNAs and Sperm RNA Modifications. *Nat. Rev. Genet.* 17, 733–743. doi:10.1038/nrg.2016.106

Clark, J. F., Dinsmore, C. J., and Soriano, P. (2020). A Most Formidable Arsenal: Genetic Technologies for Building a Better Mouse. *Genes Dev.* 34, 1256–1286. doi:10.1101/gad.342089.120

Dev, R. R., Ganji, R., Singh, S. P., Mahalingam, S., Banerjee, S., and Khosla, S. (2017). Cytosine Methylation by DNMT2 Facilitates Stability and Survival of HIV-1 RNA in the Host Cell during Infection. *Biochem. J.* 474, 2009–2026. doi:10.1042/bcj20170258

Easd (2015). Paternal Chronic High-Fat Diet Consumption Reprogrammes the Gametic Epigenome and Induces Transgenerational Inheritance of Metabolic Disorder. *Diabetologia* 1, 1. doi:10.1007/s00125-015-3687-4

Genenncher, B., Durdevic, Z., Hanna, K., Zinkl, D., Mobin, M. B., Senturk, N., et al. (2018). Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats. *Cel Rep.* 22, 1861–1874. doi:10.1016/j.celrep.2018.01.061

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., et al. (2006). Methylation of tRNA Asp by the DNA Methyltransferase Homolog Dnmt2. *Science* 311, 395–398. doi:10.1126/science.1120976

Grandjean, V., Fourré, S., De Abreu, D. A. F., Derieppe, M.-A., Remy, J.-J., and Rassoulzadegan, M. (2015). RNA-mediated Paternal Heredity of Diet-Induced Obesity and Metabolic Disorders. *Sci. Rep.* 5. doi:10.1038/srep18193

Hur, S. S. J., Cropley, J. E., and Suter, C. M. (2017). Paternal Epigenetic Programming: Evolving Metabolic Disease Risk. *J. Mol. Endocrinol.* 58, R159–R168. doi:10.1530/JME-16-0236

Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S., et al. (2017). Mechanism and Biological Role of Dnmt2 in Nucleic Acid Methylation. *RNA Biol.* 14, 1108–1123. doi:10.1080/15476286.2016.1191737

Kawano, M., Kawaji, H., Grandjean, V., Kiani, J., and Rassoulzadegan, M. (2012). Novel Small Noncoding RNAs in Mouse Spermatozoa, Zygotes and Early Embryos. *PLoS One* 7, e44542. doi:10.1371/journal.pone.0044542

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The Human Genome Browser at UCSC. *Genome Res.* 12 (6), 996–1006. doi:10.1101/gr.2291

Kc, C., Y, L., and Jt, C. (2018). The Areas under Curves (AUC) Used in Diabetes Research: Update View. *Integr. Obes. Diabetes* 4, 1–2. doi:10.15761/iod.1000212

Kiani, J., Grandjean, V., Liebers, R., Tuorto, F., Ghanbarian, H., Lyko, F., et al. (2013). RNA-mediated Epigenetic Heredity Requires the Cytosine Methyltransferase Dnmt2. *Plos Genet.* 9, e1003498. doi:10.1371/journal.pgen.1003498

Kianmehr, L., Khazali, H., Rajabi-Maham, H., Sharifi-Zarchi, A., Cuzin, F., and Rassoulzadegan, M. (2019). Genome-Wide Distribution of Nascent Transcripts in Sperm DNA, Products of a Late Wave of General Transcription. *Cells* 8 (10), 1196. doi:10.3390/cells8101196

Klastrup, L. K., Bak, S. T., and Nielsen, A. L. (2019). The Influence of Paternal Diet on sncRNA-Mediated Epigenetic Inheritance. *Mol. Genet. Genomics* 294, 1–11. doi:10.1007/s00438-018-1492-8

Li, J.-S., and YaoMicroRNAs, Z.-X. (2012). MicroRNAs: Novel Regulators of Oligodendrocyte Differentiation and Potential Therapeutic Targets in Demyelination-Related Diseases. *Mol. Neurobiol.* 45, 200–212. doi:10.1007/s12035-011-8231-z

Lyko, F. (2018). The DNA Methyltransferase Family: A Versatile Toolkit for Epigenetic Regulation. *Nat. Rev. Genet.* 19, 81–92. doi:10.1038/nrg.2017.80

Martinez-Delgado, B., Yanowsky, K., Inglada-Perez, L., Domingo, S., Urioste, M., Osorio, A., et al. (2011). Genetic Anticipation Is Associated with Telomere Shortening in Hereditary Breast Cancer. *Plos Genet.* 7, e1002182. doi:10.1371/journal.pgen.1002182

Massiera, F., Barbry, P., Guesnet, P., Joly, A., Luquet, S., Moreilhon-Brest, C., et al. (2010). A Western-like Fat Diet Is Sufficient to Induce a Gradual Enhancement in Fat Mass over Generations. *J. Lipid Res.* 51, 2352–2361. doi:10.1194/jlr.M006866

Montgomery, M. K., Hallahan, N. L., Brown, S. H., Liu, M., Mitchell, T. W., Cooney, G. J., et al. (2013). Mouse Strain-dependent Variation in Obesity and Glucose Homeostasis in Response to High-Fat Feeding. *Diabetologia* 56, 1129–1139. doi:10.1007/s00125-013-2846-8

Nätt, D., Kugelberg, U., Casas, E., Nedstrand, E., Zalavary, S., Henriksson, P., et al. (2019). Human Sperm Displays Rapid Responses to Diet. *Plos Biol.* 17, e3000559. doi:10.1371/journal.pbio.3000559

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* 14, 417–419. doi:10.1038/nmeth.4197

Peng, H., Shi, J., Zhang, Y., Zhang, H., Liao, S., Li, W., et al. (2012). A Novel Class of tRNA-Derived Small RNAs Extremely Enriched in Mature Mouse Sperm. *Cell Res* 22, 1609–1612. doi:10.1038/cr.2012.141

Pomp, D., Nehrenberg, D., and Estrada-Smith, D. (2008). Complex Genetics of Obesity in Mouse Models. *Annu. Rev. Nutr.* 28, 331–345. doi:10.1146/annurev.nutr.27.061406.093552

Rassoulzadegan, M., Sharifi-Zarchi, A., and Kianmehr, L. (2021). DNA-RNA Hybrid (R-Loop): From a Unified Picture of the Mammalian Telomere to the Genome-wide Profile. *Cells* 10, 1556. doi:10.3390/cells10061556

Rossmeisl, M., Rim, J. S., Koza, R. A., and Kozak, L. P. (2003). Variation in Type 2 Diabetes-Related Traits in Mouse Strains Susceptible to Diet-Induced Obesity. *Diabetes* 52, 1958–1966. doi:10.2337/diabetes.52.8.1958

Schaefer, M., Pollex, T., Hanna, K., Tuorto, F., Meusburger, M., Helm, M., et al. (2010). RNA Methylation by Dnmt2 Protects Transfer RNAs against Stress-Induced Cleavage. *Genes Dev.* 24, 1590–1595. doi:10.1101/gad.586710

Shi, Z., Fujii, K., Kovary, K. M., Genuth, N. R., Röst, H. L., Teruel, M. N., et al. (2017). Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cel* 67, 71–83. doi:10.1016/j.molcel.2017.05.021

Sims, E. K., Hatanaka, M., Morris, D. L., Tersey, S. A., Kono, T., Chaudry, Z. Z., et al. (2013). Divergent Compensatory Responses to High-Fat Diet between C57Bl6/J and c57BlKS/J Inbred Mouse Strains. *Am. J. Physiology-Endocrinology Metab.* 305, E1495–E1511. doi:10.1152/ajpendo.00366.2013

Tuorto, F., Liebers, R., Musch, T., Schaefer, M., Hofmann, S., Kellner, S., et al. (2012). RNA Cytosine Methylation by Dnmt2 and NSun2 Promotes tRNA Stability and Protein Synthesis. *Nat. Struct. Mol. Biol.* 19, 900–905. doi:10.1038/nsmb.2357

Virtue, S., and Vidal-Puig, A. (2021). GTTs and ITTs in Mice: Simple Tests, Complex Answers. *Nat. Metab.* 3, 883–886. doi:10.1038/s42255-021-00414-7

West, D. B., and York, B. (1998). Dietary Fat, Genetic Predisposition, and Obesity: Lessons from Animal Models. *Am. J. Clin. Nutr.* 67, 505S–512S. doi:10.1093/ajcn/67.3.505S

Winzell, M. S., and Ahre´n, B. (2004). The High-Fat Diet-Fed Mouse. *Diabetes* 53, S215–S219. doi:10.2337/diabetes.53.suppl_3.S215

Xue, S., and Barna, M. (2012). Specialized Ribosomes: A New Frontier in Gene Regulation and Organismal Biology. *Nat. Rev. Mol. Cel Biol* 13, 355–369. doi:10.1038/nrm3359

Zhang, X., Cozen, A. E., Liu, Y., Chen, Q., and Lowe, T. M. (2016). Small RNA Modifications: Integral to Function and Disease. *Trends Mol. Med.* 22, 1025–1034. doi:10.1016/j.molmed.2016.10.009

Zhang, Y., Shi, J., Rassoulzadegan, M., Tuorto, F., and Chen, Q. (2019). Sperm RNA Code Programmes the Metabolic Health of Offspring. *Nat. Rev. Endocrinol.* 15, 489–498. doi:10.1038/s41574-019-0226-2

Zhang, Y., Zhang, X., Shi, J., Tuorto, F., Li, X., Liu, Y., et al. (2018). Dnmt2 Mediates Intergenerational Transmission of Paternally Acquired Metabolic Disorders through Sperm Small Non-coding RNAs. *Nat. Cel Biol.* 20, 535–540. doi:10.1038/s41556-018-0087-2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership