



5-Methylcytosine RNA Methyltransferases-Related Long Non-coding RNA to Develop and Validate Biochemical Recurrence Signature in Prostate Cancer

OPEN ACCESS

Ke Wang^{1,2†}, Weibo Zhong^{1†}, Zining Long^{1†}, Yufei Guo^{1†}, Chuanfan Zhong¹, Taowei Yang¹, Shuo Wang¹, Houhua Lai¹, Jianming Lu¹, Pengxiang Zheng^{1,3*} and Xiangming Mao^{1*}

Edited by:

Jia Qu,
Changzhou University, China

Reviewed by:

Luo Yuhao,
Southwest Medical University, China
Nana Guan,
Guizhou University of Finance and
Economics, China
Min Liang,
Fifth Affiliated Hospital of Guangzhou
Medical University, China

*Correspondence:

Xiangming Mao
mxm631221@126.com
Pengxiang Zheng
zpx365@163.com

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

Specialty section:

This article was submitted to
Protein and RNA Networks,
a section of the journal
Frontiers in Molecular Biosciences

Received: 13 September 2021

Accepted: 01 November 2021

Published: 01 December 2021

Citation:

Wang K, Zhong W, Long Z, Guo Y,
Zhong C, Yang T, Wang S, Lai H, Lu J,
Zheng P and Mao X (2021) 5-
Methylcytosine RNA
Methyltransferases-Related Long
Non-coding RNA to Develop and
Validate Biochemical Recurrence
Signature in Prostate Cancer.
Front. Mol. Biosci. 8:775304.
doi: 10.3389/fmolb.2021.775304

¹Department of Urology, Zhujiang Hospital, Southern Medical University, Guangzhou, China, ²Department of Urology, The Hospital of Trade-Business in Hunan Province, Changsha, China, ³Department of Urology, Fujing City Hospital Affiliated with Fujian Medical University, Fuzhou, China

The effects of 5-methylcytosine in RNA (m5C) in various human cancers have been increasingly studied recently; however, the m5C regulator signature in prostate cancer (PCa) has not been well established yet. In this study, we identified and characterized a series of m5C-related long non-coding RNAs (lncRNAs) in PCa. Univariate Cox regression analysis and least absolute shrinkage and selector operation (LASSO) regression analysis were implemented to construct a m5C-related lncRNA prognostic signature. Consequently, a prognostic m5C-lnc model was established, including 17 lncRNAs: *MAFG-AS1*, *AC012510.1*, *AC012065.3*, *AL117332.1*, *AC132192.2*, *AP001160.2*, *AC129510.1*, *AC084018.2*, *UBXN10-AS1*, *AC138956.2*, *ZNF32-AS2*, *AC017100.1*, *AC004943.2*, *SP2-AS1*, *Z93930.2*, *AP001486.2*, and *LINC01135*. The high m5C-lnc score calculated by the model significantly relates to poor biochemical recurrence (BCR)-free survival ($p < 0.0001$). Receiver operating characteristic (ROC) curves and a decision curve analysis (DCA) further validated the accuracy of the prognostic model. Subsequently, a predictive nomogram combining the prognostic model with clinical features was created, and it exhibited promising predictive efficacy for BCR risk stratification. Next, the competing endogenous RNA (ceRNA) network and lncRNA-protein interaction network were established to explore the potential functions of these 17 lncRNAs mechanically. In addition, functional enrichment analysis revealed that these lncRNAs are involved in many cellular metabolic pathways. Lastly, *MAFG-AS1* was selected for experimental validation; it was upregulated in PCa and probably promoted PCa

Abbreviations: AUC, area under the ROC curve; BCR, biochemical recurrence; CCK-8, Cell Counting Kit-8; ceRNA, competing endogenous RNA; CIT, Cartes d'Identité des Tumeurs; DCA, decision curve analysis; FPKM, Fragments Per Kilobase Million; GEO, Gene Expression Omnibus; GO, Gene Ontology; GS, Gleason score; HR, hazard ratio; KEGG, Kyoto Encyclopedia of Genes and Genomes; lncRNA, long non-coding RNA; LASSO, least absolute shrinkage and selector operation; m5C, 5-methylcytosine; m6A, N6-methyladenosine; mCRPC, metastatic castration-resistant prostate cancer; mCSPC, metastatic castration-sensitive prostate cancer; NGS, next-generation sequencing; PCa, prostate cancer; PRAD, Prostate Adenocarcinoma; PSA, prostatic specific antigen; ROC, receiver operating characteristic; RT-qPCR, real-time quantitative PCR; siRNA, small interfering RNA; SD, standard deviation; TCGA, The Cancer Genome Atlas.

proliferation and invasion *in vitro*. These results offer some insights into the m5C's effects on PCa and reveal a predictive model with the potential clinical value to improve the prognosis of patients with PCa.

Keywords: 5-methylcytosine in RNA (m5C), lncRNA, biochemical recurrence, prostate cancer, prognostic model

INTRODUCTION

Prostate cancer (PCa) ranks as the second most commonly diagnosed tumor and the fifth leading cause of mortality in men worldwide, with an increasing trend in incidence (Sung et al., 2021). Moreover, it is reported that the estimated new cases of PCa for 2021 in the United States will probably hit 248,530 with its estimated deaths up to about 34,130, leading this common male malignancy to be a considerable challenge and economic burden to both health services and society (Siegel et al., 2021).

Thankfully, 5-year relative survival rates between localized PCa and metastatic one could differ; the 5-year relative survival rate of localized PCa is >99%, but when it comes to distant/metastatic PCa, the unfavorable 5-year relative survival rate decreases to about 30.6% (National Cancer Institute, 2021), suggesting that metastasis mainly accounts for poor prognoses of patients. Collectively, clinical predicaments that PCa has brought include distinguishing whether a PSA (prostatic specific antigen) detection-based localized PCa presents as an indolent or aggressive one, determining the optimal collocation sequence of systemic therapies for both metastatic castration-sensitive and castration-resistant PCa (mCSPC and mCRPC), and developing promising biomarkers to guide treatment options and improve the prognoses of patients with metastatic PCa (Ku et al., 2019).

High heterogeneity is commonly considered to be one of the significant hallmarks of PCa. The high heterogeneity, mainly characterized by multiple genomic alterations, contributes to cancer initiation, progression and metastasis, and difficulty for the diagnosis, prognosis, and treatment (Dinescu et al., 2019). In addition to genomic aberrations, epigenetic modifications that recently have risen to fame have been reported to be associated with cancer progression and might provide new insights to innovate novel diagnostic and therapeutic strategies. For decades, epigenetics has been prominently acquainted with various and reversible chemical modifications of DNA and histones influencing gene expression while bypassing genomic alteration at the same time. However, in recent years, another layer of gene regulation at the RNA level, known as epitranscriptomics (Deng et al., 2018), has attracted increasing attention and interest in the scientific community with advanced molecular technologies emerging. Compared with the relatively limited spectrum of DNA modifications (six kinds), the abundance of modifications in RNA presents much higher, with over 170 types identified so far. Generally, 5-methylcytosine in DNA (5mC) and its oxidized derivatives, including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and as 5-carboxylcytosine (5caC) are the most well-documented ones at the DNA level (Bohnsack et al., 2019), functioning epigenetically as gene expression regulators via

plenty of diverse mechanisms. However, 5-methylcytosine is also detected in several RNA species (m5C), emerging as a critical modulator in many aspects of gene expression, including RNA export, ribosome assembly, translation, and RNA stability (Bohnsack et al., 2019; Dou et al., 2020; Rong et al., 2021).

Long non-coding RNAs (lncRNAs), one category of the non-coding RNA family, have gained intensive attention ever since they were re-classified as critical regulators of gene expression at both transcriptional and post-transcriptional levels rather than useless non-coding materials, functioning in multiple fundamental cellular processes, even in cancer states. Diverse functions, along with the large abundance of lncRNAs, indicate a promising future for lncRNAs to act as both promoters and suppressors of tumor, and accordingly, cumulative evidence has significantly emerged greatly in the past few years (Liang et al., 2019; Hong et al., 2020; Goodall and Wickramasinghe, 2021). For example, the lncRNA *HOTAIR* has been found highly expressed in primary breast tumors and metastases, promoting cancer progression (Gupta et al., 2010), while Second Chromosome Locus Associated with Prostate-1, *LINC00913*, tends to be a promising biomarker given that its expression level is independently related to poor prognosis (Prensner et al., 2013). In our previous study, *SNHG1* has also been proved to be a novel lncRNA, contributing to the malignant progression in PCa (Tan et al., 2021). Previously, methylation of 5-cytosine at the post-transcriptional level was found only in tRNAs and rRNAs. Still, in the era of high-throughput sequencing, m5C is now validated to widely exist in coding RNAs plus other noncoding RNAs such as lncRNAs, microRNAs, etc. Despite accumulating studies focusing on m5C in mRNA, the role of m5C in lncRNA remains rarely elucidated.

Since the mid-2000s, the sequencing technologies named next-generation sequencing (NGS) have continued to evolve, making great strides in the speed and cost of sequencing cancer (Mardis, 2013; Roychowdhury and Chinnaiyan, 2016). In parallel, the volumes of biological data generated by these high-throughput sequencing technologies are so vast that bioinformatics analysis and high-performance computing are necessarily introduced to data processing (Singer et al., 2019). Thus, more and more researchers utilize bioinformatics analysis to construct models related to diagnosis, treatment, and prognosis of cancer.

In this study, we processed the dataset from The Cancer Genome Atlas (TCGA) database via utilizing the least absolute shrinkage and selector operation (LASSO) algorithm (Tibshirani, 1996) to construct a prognostic model with a m5C-related lncRNA signature to predict the biochemical recurrence (BCR) status of patients with PCa. We verified its efficacy and authenticity with the GSE54460 dataset from the Gene Expression Omnibus (GEO) database. Then the competing

endogenous RNA (ceRNA) network and lncRNA–protein network were also established, followed by functional enrichment analysis for further exploration. Finally, we chose lncRNA *MAFG-AS1* for preliminary experimental validation, and a set of functional enrichment analyses were performed to investigate its potential molecular function.

MATERIALS AND METHODS

Data Processing

TCGA dataset for Prostate Adenocarcinoma (PRAD), including a total of 460 patients with prognostic information, was obtained from PCaDB (Li et al., 2021a; Li et al., 2021b) (<http://bioinfo.jialab-ucr.org/PCaDB/>), a comprehensive and interactive database for transcriptomes from PCa population cohorts. In addition, the complete clinical data of 460 patients were downloaded from TCGA (<https://portal.gdc.cancer.gov/>) for subsequent analyses. The GSE54460 dataset was downloaded from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). R software version: 4.1.0, website tool “catRAPID omics v2.0,” and website tool “OmicShare tools” were used for all analyses in the study.

Identification of m5C-Related lncRNAs

Thirteen m5C-related genes were collected from literature mining (Takai et al., 2014; Archer et al., 2016; Blanco et al., 2016; Yamashita et al., 2017; Li Y. et al., 2018; Cheng et al., 2018; García et al., 2018; Chen et al., 2019; Gao et al., 2019; Janin et al., 2019; Carella et al., 2020; He et al., 2020; Mei et al., 2020; Sato et al., 2020). Then, expression data of the 13 m5C-related genes and all lncRNAs were retrieved from the TCGA dataset. Pearson correlation analysis (Schober et al., 2018) was used to explore the correlation between lncRNAs and the 13 m5C-related genes; the criteria were $|\text{Pearson } R| > 0.4$ and $p < 0.05$, and eventually 678 lncRNAs were filtered out. Next, logistic regression analysis (Tolles and Meurer, 2016) and univariate Cox regression analysis (Cox, 1972) were implemented on these 678 lncRNAs for further screening, and eventually, 102 overlapped lncRNAs with p -value < 0.01 were obtained.

Construction and Verification of the m5C-Related Prognostic Model

We applied the LASSO regression model (R package “glmnet”) to shrink the candidate genes and, meanwhile, to establish the predictive model. Ultimately, 17 lncRNAs and their respective coefficients were calculated, and the minimum criteria chose the penalty parameter (λ). The m5C-lnc score was calculated by the formula: $m5C - \text{lnc Score} = \sum_{i=1}^N (\text{Coef}_i \times \text{Expression level of lncRNA}_i)$ where N (17) denotes the number of lncRNAs included in the predictive model, Coef_i refers to a specific lncRNA’s coefficient, and $\text{Expression level of lncRNA}_i$ represents the relative expression level of a specific lncRNA. The TCGA PCa patients were separated into low- and high-score subgroups according to the median score (–1.713). The BCR-free time was compared between the two subgroups via Kaplan–Meier (KM) survival analysis (Kaplan and Meier, 1958)

in the “survminer” package. The “survivalROC” package was employed to perform 1-, 3-, and 5-year ROC (receiver operating characteristic) curve analyses (Mandrekar, 2010) for assessing the predictive power of the prognostic signature, and we compared the AUC (area under the ROC curve) of m5C-lnc score with the AUCs of other clinicopathological features to determine its clinical value. A cohort from the GEO database (GSE54460) was included for validation of the predictive model. The m5C-lnc score for each patient in the GSE54460 cohort was also calculated by the same formula above. Likewise, patients were separated into two subgroups—low- and high-score subgroups—for KM analysis. Finally, the two subgroups were then compared to validate the predictive model, too. We constructed the nomogram and calibration plots based on our previous study (Zhong et al., 2021).

Construction of ceRNA Network and lncRNA–Protein Interaction Network and Functional Enrichment Analysis

The “GDCRNAtools” package was employed to construct the ceRNA network (Salmena et al., 2011), and the website tool “catRAPID omics v2.0” (http://service.tartagliolab.com/page/catrapid_omics2_group) was utilized to construct the lncRNA–protein interaction network. The website tool “OmicShare tools” (<https://www.omicshare.com/tools/>) was then used to explore the functional enrichments of the prognostic lncRNAs.

Cell Culture, RNA Extraction, and Real-Time Quantitative PCR

PCa cell lines, DU145 and PC-3, were both obtained from the National Collection of Authenticated Cell Cultures. Both cell lines were cultured in Dulbecco’s modified Eagle’s medium. Culture media contained 10% fetal bovine serum and 1% double antibiotics (penicillin and streptomycin). All cell lines were cultivated at 37°C and 5% CO₂. Total RNAs were extracted from PC-3 and DU145 cells using Trizol reagent (15596018, Takara). Total RNAs were then reverse-transcribed into cDNA using TransScript All-in-one First-Strand cDNA Synthesis SuperMix for qPCR (AT341-01, TransGen). RT-qPCR (real-time quantitative PCR) was carried out using the PerfectStart Green (AQ601-02, TransGen) in Applied Biosystems 7500 Real-Time PCR System. Finally, the relative expression of *MAFG-AS1* was calculated based on the internal reference glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*). All experiments were carried out with three replicates. The primers of *MAFG-AS1* and *GAPDH* are listed in **Supplementary Table S3**.

RNA Interference and Loss of Function Assays

GenePharm Company synthesized small interfering RNAs (siRNAs) targeting *MAFG-AS1*. RT-qPCR validated the transfection efficiency after the transfection of siRNAs along with siRNA-Mate (GenePharm) for 72 h. CCK-8 (Cell Counting Kit-8, MA0218-5, Meilunbio) viability assay and

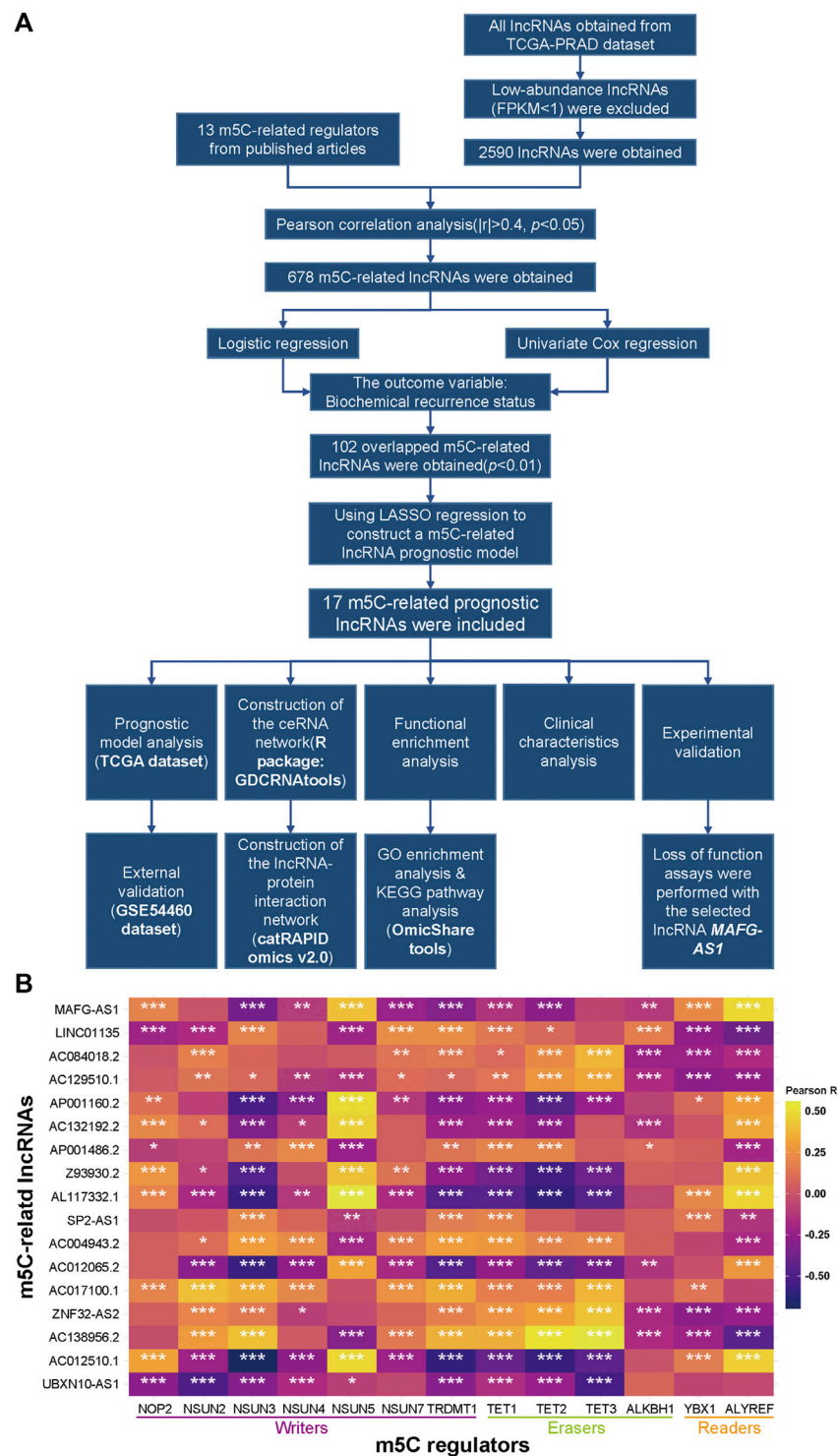


FIGURE 1 | Identification of m5C-related lncRNAs in TCGA-PRAD. **(A)** Flow chart. **(B)** Heatmap of the correlations between m5A-related genes and the 17 prognostic m5C-related lncRNAs. * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

colony formation assay were applied to examine the proliferative ability of PCa cell lines after knocking down *MAFG-AS1*. Transwell assay was conducted to examine the invasiveness of

MAFG-AS1 downregulated cells. Detailed procedures of the above assays are available in our previous study (Zhong et al., 2021). All experiments were implemented with three replicates.

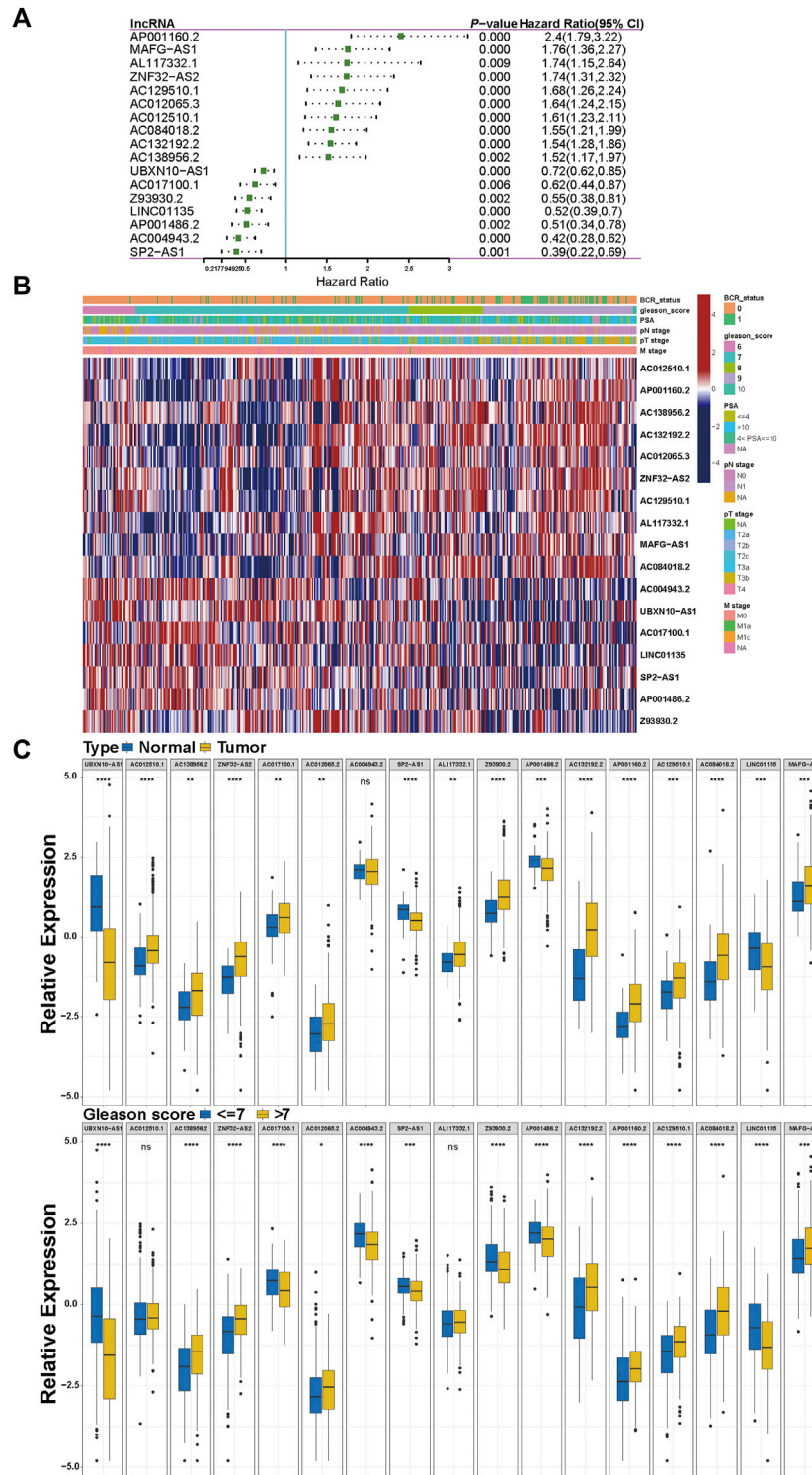


FIGURE 2 | Clinical characteristics of the 17 prognostic m5C-related lncRNAs. **(A)** The result of univariate Cox regression for the 17 prognostic m5C-related lncRNAs. **(B)** The heatmap of clinical signature of the 17 prognostic m5C-related lncRNAs. **(C)** The correlations between the 17 prognostic m5C-related lncRNAs and clinical features. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

siRNAs targeting sites in *MAFG-AS1* are shown in **Supplementary Table S3**. The original scans of the plate colony formation assays and Transwell assays are shown in **Supplementary Figure S3**.

Statistical Analyses

All bioinformatics analyses were performed by R software version 4.1.0 (The R Project for Statistical Computing, Vienna, Austria). Pearson correlation analysis was carried out to analyze correlations between m5C-related genes and lncRNAs. The “survival” and “survminer” packages were applied for KM plots and Cox regression analysis. GraphPad Prism 7.0 (GraphPad, La Jolla, CA, United States) was utilized to analyze the results of RT-qPCR and cell functional assays. All statistical results were showed as mean \pm SD (standard deviation) with two-sided test, and the results with a p -value < 0.05 were regarded as statistically significant ones.

RESULTS

Identification of m5C-Related lncRNAs With Significant Prognostic Value

The workflow chart is presented in **Figure 1A**. Firstly, we searched and identified 13 m5C- regulators from published articles, and then TCGA and GEO datasets of prostate adenocarcinoma were downloaded for further analysis. After extracting the expression matrixes of these m5C regulators and all lncRNAs in the TCGA dataset and excluding lncRNAs with FPKM < 1 (Fragments Per kilobase Million), we performed a Pearson correlation analysis between these m5C-related genes and 2,590 lncRNAs to determine whether a lncRNA was correlated with the m5C modification ($|\text{Pearson } R| > 0.4$ and $p < 0.05$), and consequently we included 678 qualified lncRNAs into logistic regression analysis and univariate Cox regression analysis for further screening (**Supplementary Tables S1 and S2**). Subsequently, 102 lncRNAs, whose p -values were both less than 0.01 in the two analyses mentioned above, were chosen for a LASSO regression analysis to build a m5C-related prognostic model to forecast the BCR in PCa (**Supplementary Figure S1A**). Eventually, we constructed a model that contains 17 m5C-related lncRNAs (**Supplementary Figure S1B**), and the correlations between the 17 lncRNAs and the 13 m5C regulators in the TCGA dataset are shown in **Figure 1B**.

Clinical Characteristics of the 17 m5C-Related lncRNAs With Prognostic Significance

Given the LASSO regression analysis, we focused on these 17 lncRNAs' clinical value. The univariate Cox analysis results of the 17 lncRNAs are shown in **Figure 2A**; 7 lncRNAs (*UBXN10-AS1*, *AC017100.1*, *Z93930.2*, *LINC01135*, *AP001486.2*, *AC004943.2*, and *SP2-AS1*) tend to be protective factors with HR (hazard ratio) < 1 , while another 10 lncRNAs' HRs are more than 1 indicating their high-risk relation with the BCR of PCa. We also separated patients into high-expression and low-expression

subgroups based on the median of each lncRNA's expression value. We performed KM survival analysis to detect a significant difference in BCR status between two subgroups. As a result, all 17 KM survival analyses revealed that the high-expression subgroups belonging to the 7 protective lncRNAs mentioned above harbored favorable BCR status. In contrast, the high-expression subgroups of another 10 high-risk lncRNAs gained poorer BCR status, and meanwhile, all the KM analyses hit a p -value < 0.05 , as shown in **Supplementary Figure S2**. Then we extracted 460 patients' clinical data, including BCR status, Gleason score (GS), PSA, pathological T stage, pathological N stage, and metastatic stage from the TCGA-PRAD dataset to inspect the 17 lncRNAs' signatures, and the heatmap is displayed in **Figure 2B**. Moreover, we implemented a further analysis to explore the correlation between 17 m5C-related modulators and clinical features after classifying these patients into several subgroups, as shown in **Figure 2C** and **Supplementary Figure S1E**. The expression of all lncRNAs except *AC004943.2* in tumors exhibited significance, compared with that in adjacent normal tissue. Regarding the T stage, N stage, and GS subgroups, five, four, and two lncRNAs did not show significant difference, respectively. Consequently, most of these lncRNAs revealed their predictive value.

Construction and Validation of the m5C-Related lncRNA Prognostic Model

Following the analyses above, we then constructed the m5C-related lncRNA prognostic model named m5C-lnc score with these 17 lncRNAs, and their coefficients calculated by LASSO regression analysis are shown in **Figure 3A**. Patients in the TCGA cohort were divided into low-score and high-score subgroups based on the m5C-lnc score attained by the formula above and its median. KM survival curves demonstrated that patients in the low-score subgroup had better clinical outcomes (longer BCR-free time), as shown in **Figure 3B**. m5C-lnc score and BCR status distribution of TCGA-PRAD dataset are depicted in **Figure 3C**. Subsequently, ROC analysis showed that the predictive m5C-lnc score model whose 1-, 3-, and 5-year AUCs were all about 0.8 (12-month AUC = 0.785, 36-month AUC = 0.825, and 60-month AUC = 0.815; **Figure 3D**) had excellent sensitivity and specificity, exhibiting better power to predict BCR status in the TCGA cohort; moreover, compared with the AUCs of other prognostic factors such as PSA, age, GS, and pathological T stage (AUC of PSA = 0.612, AUC of age = 0.51, AUC of GS = 0.728, and AUC of pathological T stage = 0.685), the AUC of the m5C-lnc score hit the highest score (0.793; **Figure 3E**), suggesting the model's better sensitivity, specificity, and clinical predictive value. The decision curve analysis (DCA) of the TCGA cohort indicated that the clinical benefit rate of the m5C-lnc score model was significantly higher than that of other clinical characteristics based on GS and T stage, respectively (**Supplementary Figure S1C**). Moreover, we performed univariate Cox regression with m5C-lnc score and other clinical pathological features including age, PSA, GS, and pathological T stage, respectively. As a result, all factors except age were significantly associated with patients' BCR status, as shown in **Supplementary Figure S4A**; PSA, GS,

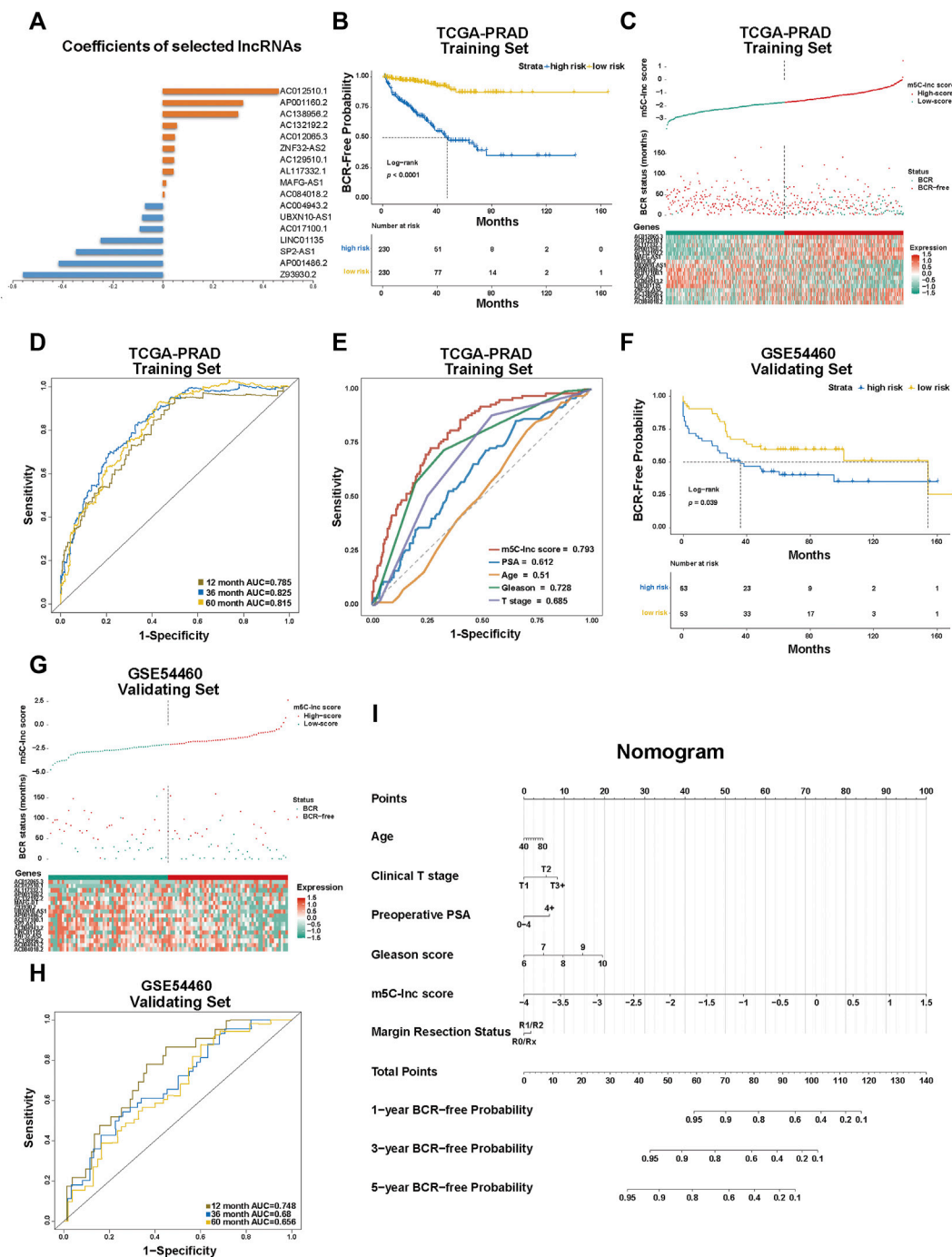
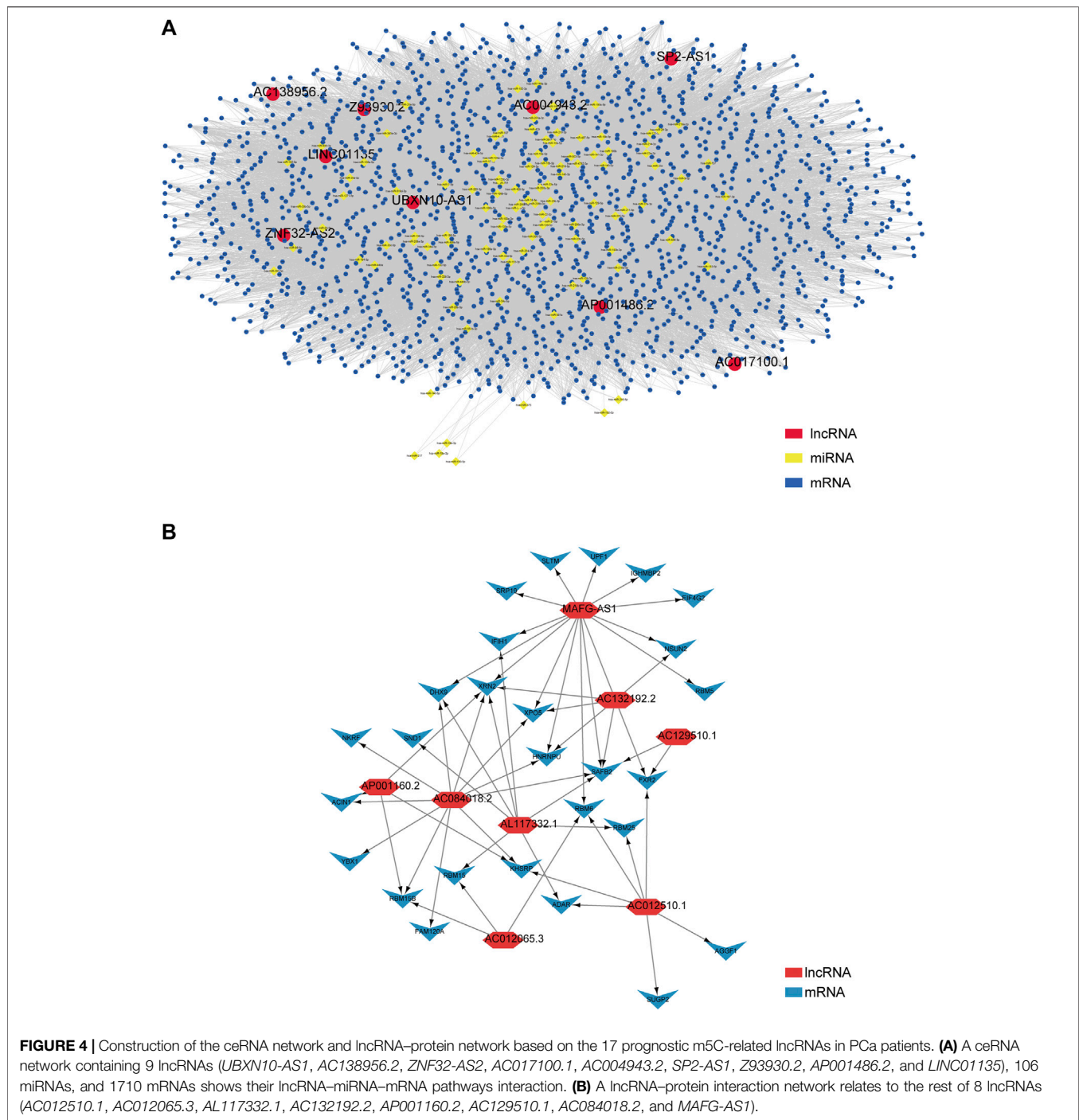


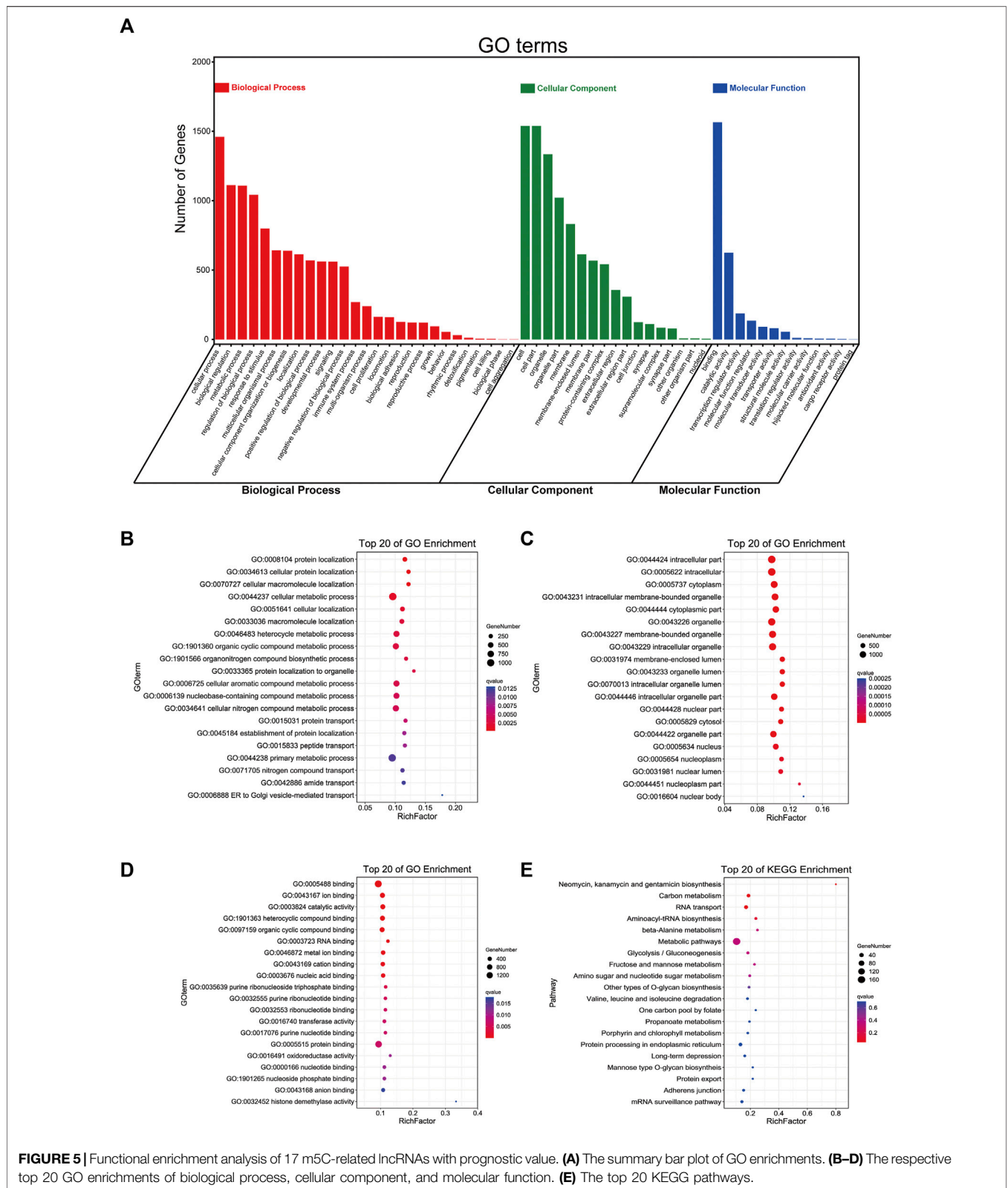
FIGURE 3 | Construction and validation of the m5C-Related lncRNA prognostic model in patients with prostate cancer. **(A)** The coefficients of the prognostic 17 m5C-related lncRNAs calculated by LASSO regression analysis. **(B)** Kaplan–Meier survival analysis shows that BCR-free time of patients with high m5C-Inc scores is significantly shorter than of those with low m5C-Inc scores. **(C)** The separation of low and high m5C-Inc scores in patients with PCA; scatter plot depicts the distribution of BCR status between two groups divided by the median m5C-Inc score; the heatmap shows that patients with shorter BCR-free time displayed higher gene expression level of risk factors (*AP001160.2*, *MAFG-AS1*, *AL117332.1*, *ZNF32-AS2*, *AC129510.1*, *AC012065.3*, *AC012510.1*, *AC084018.2*, *AC132192.2*, and *AC138956.2*), while patients with longer BCR-free time displayed higher gene expression level of protective factors (*UBXN10-AS1*, *AC017100.1*, *AC004943.2*, *SP2-AS1*, *Z93930.2*, *AP001486.2*, and *LINC01135*). **(D)** ROC curves illustrated the sensitivities of the prognostic m5C-Inc score model predicting 1-, 3- and 5-year BCR status. **(E)** ROC sensitivity comparison between m5C-Inc score and other clinical features including PSA level, age, Gleason score, and T stage. **(F–H)** The validation for the constructed prognostic model using GSE54460 dataset. **(I)** The nomogram containing the m5C-Inc score and other clinicopathological features for prognosis prediction.



m5C-lnc score, and pathological T stage were then included in further multivariate Cox regression, and consequently, m5C-lnc score presented as the primary risk factor with the highest HR of about 3, as shown in **Supplementary Figure S4B**.

To validate the predictive value of the established m5C-related lncRNA signature, we applied the GSE54460 dataset to verify the results; likewise, patients in the GSE54460 dataset were also separated into low-score and high-score groups by the median m5C-lnc score. Consequently, the survival analysis results were

similar to those in the TCGA dataset: patients with higher m5C-lnc scores had a lower BCR-free rate and a shorter BCR-free time (**Figure 3F**). Likewise, m5C-lnc score and BCR status distribution are illustrated in **Figure 3G**. The ROC analysis indicated that these m5C-related regulators' signature also had a promising prognostic value for patients in the GSE54460 dataset given that the 1-, 3-, and 5-year AUCs were all around 0.7 (12-month AUC = 0.748, 36-month AUC = 0.68, and 60-month AUC = 0.655; **Figure 3H**). Nomogram is always used to



quantitatively predict the prognosis of patients in clinical practice by calculating the relative points on a set of scales. In our study, we illustrated our predictive model in the form of a nomogram as

shown in **Figure 3I** to predict 1-, 3-, and 5-year BCR-free probabilities, respectively, for patients with PCA according to the “Total Points.” A total point considers six variables including

PSA, age, T-stage, GS, m5C-lnc score, and resection status. With the scale bar “Points” to be a reference, the primary value of each variable could correspond to a point on the “Points” scale, and all the six relative points are added together to make the total points. Ultimately, the probabilities of 1-, 3-, or 5-year BCR-free time for each patient could be quantified respectively by the points corresponding to the scale bar “Total Points.” Besides, we also carried out a calibration curve analysis to demonstrate that the estimated 3- and 5-year BCR-free time were in accordance with the actual scenario (**Supplementary Figure S1D**).

Construction of the Competing Endogenous RNA Network and the lncRNA–Protein Interaction Network

Next, we explored how these lncRNAs function in biological processes by using the R package called “GDCRNAtools” to construct a ceRNA network and visualizing it with the help of the software Cytoscape (Shannon et al., 2003). Consequently, 9 lncRNAs (*UBXN10-AS1*, *AC138956.2*, *ZNF32-AS2*, *AC017100.1*, *AC004943.2*, *SP2-AS1*, *Z93930.2*, *AP001486.2*, and *LINC01135*), 106 miRNAs, and 1,710 mRNAs were included in the ceRNA network with their lncRNA–miRNA–mRNA pathways interacting mutually, as shown in **Figure 4A**. For the rest of lncRNAs (*AC012510.1*, *AC012065.3*, *AL117332.1*, *AC132192.2*, *AP001160.2*, *AC129510.1*, *AC084018.2*, and *MAFG-AS1*), we utilized an online website tool called catRAPID omics v2.0 (Agostini et al., 2013) to inspect the lncRNA–protein interaction. After inputting each lncRNA’s information, the tool automatically predicted the candidate proteins and presented the result list. Given the build-in ranking system (0 star to 3 stars), we selected 27 candidate proteins with 2.5 and more stars to construct the interaction network. Likewise, the lncRNA–protein interaction network was visualized by Cytoscape, as shown in **Figure 4B**. Detailed data can be found in **Supplementary Tables S4 and S5**.

Functional Enrichment Analysis of 17 m5C-Related lncRNAs With Prognostic Value

Furthermore, we used these 17 lncRNAs to perform functional enrichment analysis with an online tool called OmicShare tools. The summary of GO (Gene Ontology) enrichment and details about its three parts are shown in **Figures 5A–D**, and the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment is also demonstrated in **Figure 5E**. Accordingly, we found that the top five GO Biological Processes that these genes were primarily enriched in are heterocycle metabolic process, nucleobase-containing compound metabolic process, organic cyclic compound metabolic process, cellular nitrogen compound metabolic process, and cellular aromatic compound metabolic process (**Figure 5B**); neomycin, kanamycin, and gentamicin biosynthesis, carbon metabolism, and RNA transport were significantly enriched in KEGG Pathway (**Figure 5E**). These data may bring us some insights into the

potential functions of these m5C-related lncRNAs in PCa. Complete information are shown in **Supplementary Table S6**.

Selecting *MAFG-AS1* for Experimental Validation and Individual Analysis

With the univariate Cox regression analysis mentioned above, 10 lncRNAs’ HRs are greater than 1, indicating their potential promoting PCa. Among them, the abundance of *MAFG-AS1* is the highest (**Figure 2C**), and its HR ranks second (**Figure 2A**). Therefore, we chose to implement the experimental validation with *MAFG-AS1*. Based on the data provided on LNCipedia (Volders et al., 2019), a website with comprehensive information of human lncRNAs, *MAFG-AS1*, known as *MAFG-DT*, is located at q25.3 of chromosome 17 with a transcript size of 1,895 bp. Subsequently, we used an online tool called iRNAm5C-PseDNC to identify 5-methylcytosine modified sites of *MAFG-AS1*, and the predicted sites included positions 4, 6, 10, 11, 13, and 17, as shown in **Figure 6A**.

Firstly, we detected the basal expression level of *MAFG-AS1* in PCa cell lines by RT-qPCR, and it was found that *MAFG-AS1* was highly expressed in all PCa cell lines with the benign prostate cell line RWPE1 to be the reference (**Figure 6B**). The expressions of *MAFG-AS1* in PC-3 and DU145 cell lines are much higher than that of others, so we chose these two cell lines to construct the *MAFG-AS1* knock-down phenotypes with two siRNAs, respectively. The results of RT-qPCR confirmed that both siRNA fragments significantly silenced the expression of *MAFG-AS1* in PC-3 cell line and DU145 cell line (**Figure 6C**). After examining the transfection efficiency in PC-3 and DU145 cell lines, we performed a set of assays to verify this specific gene’s function on cancer cells. CCK-8 assay and plate colony formation assay suggested that down-regulated *MAFG-AS1* significantly inhibited PCa proliferation ability (**Figures 6D, E**). Furthermore, knockdown of *MAFG-AS1* attenuated PCa cell’s invasion ability via transwell assay (**Figure 6F**). These results indicated that *MAFG-AS1* acts as a high-risk predictive factor in PCa, and somehow its high expression promotes cancer progression.

After determining *MAFG-AS1*’s promotion in PCa’s proliferation and invasion, we attempted to inspect its molecular function further. First of all, we checked out *MAFG-AS1*’s expression profile with a pan-cancer analysis on GEPIA 2 (Tang et al., 2017) as shown in **Figure 7A**, and it turned out that *MAFG-AS1* is highly expressed in most cancers. Additionally, we detected a significant difference in the expression of *MAFG-AS1* in tumor tissue and adjacent normal tissue in another two datasets, CIT dataset and Cambridge Dataset (GSE70768), as shown in **Figures 7B, C**; the expression of *MAFG-AS1* is higher in tumor tissue than in the adjacent normal tissue. Meanwhile, we also explored the relation between *MAFG-AS1*’s expression and clinical characteristics in patients with PCa. It was found that patients with high GS tended to have high *MAFG-AS1* expression level (GS = 6: 1.29 ± 0.70 , GS = 7: 1.58 ± 0.84 , and GS ≥ 8 : 1.84 ± 0.92 , $p < 0.001$), and the significant difference was also observed in the pathological T stage category, lymph node metastasis category, and BCR status

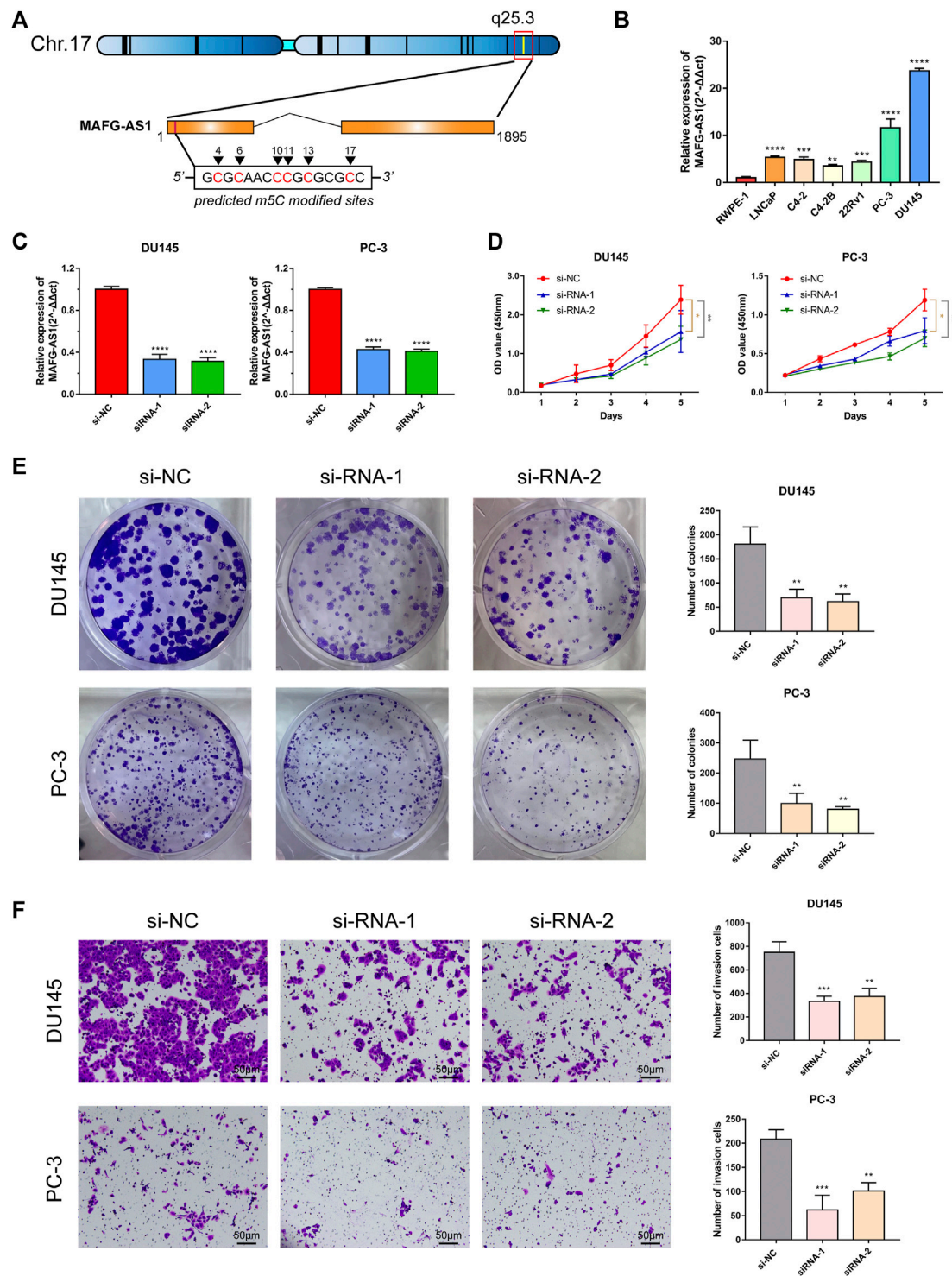
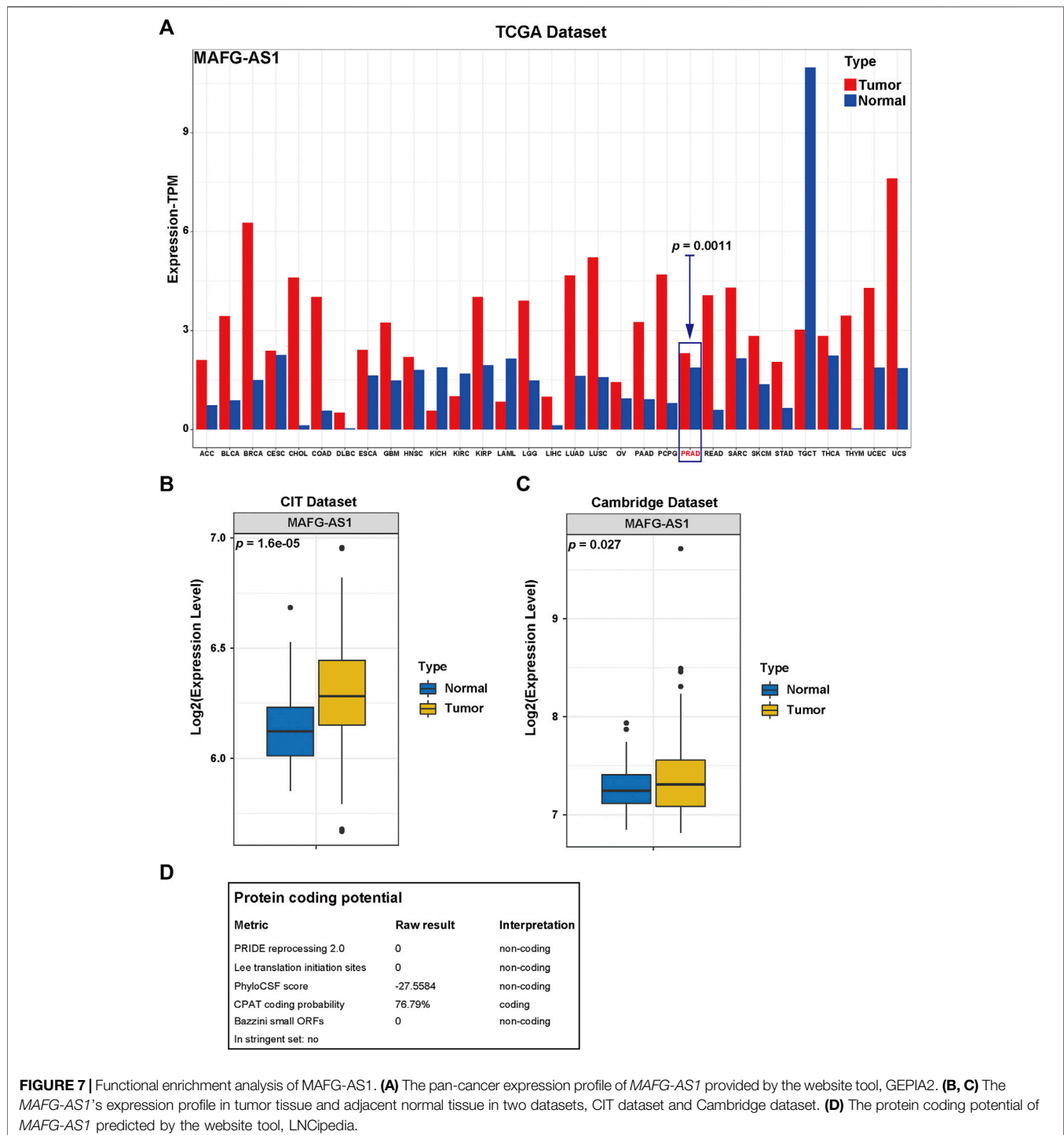


FIGURE 6 | Characteristics of *MAFG-AS1* and its functional effect on PCa. **(A)** The basic information of *MAFG-AS1* and its m5C sites predicted by the website tool, iRNAm5C-PseDNC. **(B)** Basal expression level of *MAFG-AS1* in PCa cell lines. **(C)** RT-PCR verified the transfection efficiency of siRNAs targeting to *MAFG-AS1* in PC-3 and DU145. **(D)** CCK-8 assay revealed that knockdown of *MAFG-AS1* inhibited PCa cells viability. **(E)** Knockdown of *MAFG-AS1* inhibited colony formation in PCa cells. **(F)** Knockdown of *MAFG-AS1* attenuated PCa cells' invasive ability. Error bar indicates mean \pm SD. * $p < 0.05$, ** $p < 0.01$, and **** $p < 0.0001$.



category ($\leq T2c$: 1.52 ± 0.81 , $T3/T4$: 1.75 ± 0.92 , $p = 0.005$; lymph node metastasis-positive: 1.89 ± 0.99 , lymph node metastasis-negative: 1.62 ± 0.87 , $p = 0.018$; BCR-positive: 1.94 ± 0.91 , BCR-negative: 1.58 ± 0.86 , $p < 0.001$), as shown in **Supplementary Table S8**. Although it was not included in the ceRNA network (**Figure 4A**), it interacts with 15 proteins in the lncRNA–protein network, ranking first, which indicates it might play an essential role in cancer progression. Then we

implemented Spearman's correlation analysis between *MAFG-AS1* and all other genes in TCGA-PRAD dataset to screen out the correlated genes of *MAFG-AS1* for further functional enrichment analysis. The criteria were $|\text{Spearman } R| > 0.4$ as well as $p\text{-value} < 0.05$, and 833 correlated genes were found. Next, we utilized these genes to perform functional enrichment analysis via the website tool OmicSharetools. Consequently, most genes were enriched in these terms: RNA processing, mRNA metabolic

process, ribonucleoprotein complex biogenesis, protein targeting, and mRNA catabolic process (GO Biological Process) and ribosome, spliceosome, and RNA transport (KEGG Pathway) (Supplementary Figures S5A–E). Complete functional enrichment documents can be found in Supplementary Table S7. These results suggested that *MAFG-AS1* might influence cancer development by interacting with proteins to interrupt the activities of RNAs, especially mRNAs. Lastly, we used LNCipedia to predict *MAFG-AS1*'s protein-coding potential. The website would carry out its prediction with various applications such as CPC, HMMER, PRIDE, PhyloCSF, CPAT, and Ribosome-profiling. As a result, only CPAT provided a coding probability of 76.79% (Figure 7D). It seems that *MAFG-AS1* encodes protein with little potential.

DISCUSSION

In the United States, PCa's incidence has already ranked first in males, with its estimated death cases ranking second. About 35% of patients will go through BCR after radical prostatectomy within a decade (Freedland et al., 2005). It could heavily threaten patients' quality of life and even lifespan when it comes to BCR because patients with BCR have higher chances of developing metastatic PCa. Thus, risk stratification of patients via PSA, GS, and pathological stage, etc., becomes significant for prognosis prediction (Kreuz et al., 2020). However, with a deeper understanding of PCa, they gradually fail to provide more precise prediction. Therefore, novel predictive models are urgent to be discovered and applied in clinical practice.

More recently, with well-improved methods via both analytical chemistry and high-throughput sequencing for detection, RNA modification has come to prominence (Roundtree et al., 2017). And numerous pieces of evidence suggest that dysfunction of RNA epigenetic pathway is strongly associated with human diseases, especially cancer (Barbieri and Kouzarides, 2020). In terms of RNA modifications, apart from the famous N6-methyladenosine (m6A) modification, 5-methylcytosine (m5C) has also gradually attracted great interest with the development of detective techniques; three effective techniques have become dominant: detection at single nucleotide resolution with bisulfite conversion, also known as bisulfite sequencing (bsRNA-seq), 5-azacytidine-mediated RNA immunoprecipitation (Aza-IP), and methylation individual-nucleotide-resolution cross-linking and immunoprecipitation (miCLIP), making precise m5C mapping more possible (Dinescu et al., 2019). The dynamic activities of RNA modifications to adjust RNAs' functions are technically controlled by three elements (Barbieri and Kouzarides, 2020): enzymes that catalyze the modified site to form ("writers"), reading proteins that detect and bind to the modified sites ("readers"), and enzymes that carry out removing the modified base, ensuring the modifying process is reversible ("erasers"). There are two significant groups of m5C writers: seven members from the NOP2/SUN RNA methyltransferase (NSUN) family and DNA methyltransferase-2 (DNMT2), previously considered

to relate to DNA methylation. While writers for m5C are now well documented, the existence of m5C erasers are still unclear given that no currently existing protein can fully reverse m5C to cytosine (Rong et al., 2021).

Previous studies showed the Aly/REF export factor (ALYREF) and Y-box binding protein 1 (YBX1) may act as significant readers for m5C, taking part in cancer-related biological processes (Yang et al., 2017; Chen et al., 2019). NSUN2 (NOP2/Sun domain family, member 2), a major m5C-modifying methyltransferase (writer) of RNA, is highly expressed in various tumors such as the esophagus, liver, pancreas, cervix, prostate, kidney, bladder, thyroid, and breast cancers (Okamoto et al., 2012; Huang et al., 2021). Recently, NSUN2 was reported to promote gastric cancer cell proliferation in a m5C-dependent manner (Mei et al., 2020). These researches demonstrated that m5C modification of RNA plays a vital role in cancer.

Notably, lncRNAs take a significant part of non-coding RNAs; increasing research unveiled those various chemical modifications, including m5C, are found in lncRNAs, and their possible involvement in different types of cancer are also reported previously. For instance, m5C modifications in lncRNAs such as XIST (X-inactive specific transcript), TERC (telomerase RNA component), RPPH1 (ribonuclease P RNA component H1), and ANRIL (antisense non-coding RNA in the INK4 locus) may have an association with leukemia and colorectal cancer, PCa, breast cancer, and PCa as well as lung cancer, respectively (Squires et al., 2012; Amort et al., 2013; Baena-Del Valle et al., 2018; Zhao et al., 2018; Pan et al., 2021). Research reported that an m5C-modified H19 lncRNA might enhance tumorigenesis and progression (Sun et al., 2020). However, researches about m5C-modified lncRNAs impacting on PCa are rarely found.

In our study, we firstly collected the m5C-related genes from published articles, and consequently, 13 m5C-related regulators were found. Then we screened out the lncRNAs related to m5C modification in PCa with TCGA dataset for LASSO regression analysis and tried to construct a prognostic model with 17 lncRNAs obtained by the analysis. After checking out the predictive model's clinical characteristics, we verified the model's value, and it turned out to be a better predictor for the prognosis of PCa patients. Subsequently, a series of bioinformatics analyses, including construction of ceRNA network and lncRNA–protein network and functional enrichment analyses, were carried out to explore the lncRNAs' functional profiles. Given functional analysis outcomes, cell metabolic processes may be involved in these lncRNAs, among which cellular nitrogen compound metabolic process enriched the most genes.

Moreover, we noticed *MAFG-AS1* given its highest abundance and second-highest HR among 10 lncRNAs with their HRs > 1. *MAFG-AS1*, known as *MAFG-DT*, is found to get involved in several tumors such as bladder cancer (Xiao et al., 2020), colorectal cancer (Ruan et al., 2020), lung cancer (Sui et al., 2019), liver cancer (Chen et al., 2021), ovarian cancer (Bai et al., 2021), breast cancer (Jia et al., 2021), and pancreatic cancer (Ye et al., 2020). However, the role of *MAFG-AS1* in PCa has not been unveiled yet. Therefore, we picked *MAFG-AS1* for experimental

validation. The results revealed that *MAFG-ASI* could promote proliferation and invasion in PCa, indicating a cancer-promoting effect. At the same time, we inspected its expression levels in both tumor tissue and adjacent normal tissue, via bioinformatics analysis, and its relation with various clinical characteristics. As a result, the expression of *MAFG-ASI* is higher in tumor tissue than in adjacent normal tissue; *MAFG-ASI*'s high expression relates to patient's poor clinical status given the analysis mentioned above. Subsequently, we tried to explore *MAFG-ASI*'s molecular function. As shown in previous results, we use the R package "GDCRNAtools" (Li R. et al., 2018) to construct the ceRNA network by putting the expression matrixes of these 17 lncRNAs, all miRNAs and all mRNAs from TCGA dataset, in the "miRcode" database for exploration. Unluckily, *MAFG-AS-1* was not included in the ceRNA network above, but it was found to rank first in the lncRNA–protein network, interacting with 15 proteins, all scored with 2.5 or more stars in the build-in system mentioned above. This indicated that *MAFG-ASI* might impact the cancer progression in other ways instead of ceRNA network in PCa. Then the functional enrichment analysis with the 833 correlated genes of *MAFG-ASI* revealed that most genes were found to get involved in RNA processing. Hence, it is likely that *MAFG-ASI* might influence cancer development via interrupting RNAs' activities like biogenesis or degradation. With LNCipedia to predict *MAFG-ASI*'s protein coding potential, only one application called CPAT provided a coding probability of 76.79%. Seemingly, *MAFG-ASI* encodes a protein with little potential.

Due to lncRNA's broad definition, lncRNA's mechanisms function heterogeneously. Spatially, lncRNA's functions could be divided into two significant parts: the nucleus and the cytoplasm. Three fundamental mechanisms are highlighted in the nucleus: scaffolding and integrating regulatory proteins to form macromolecular complexes, localizing to specific sites on genomic DNA, and organizing three-dimensional nuclear structure (Engreitz et al., 2016). One of the best illuminated examples of lncRNA regulating gene expression in the nucleus is Xist (X inactive specific transcript). In the cytoplasm, lncRNAs mainly regulate gene expression by competing for endogenous RNAs (ceRNAs) to competitively bind miRNAs, thereby indirectly affecting the target mRNAs' expression (Salmena et al., 2011). Based on the definition, a lncRNA usually refers to a transcript with a sequence longer than 200 bp and lacks protein-coding potential (Goodall and Wickramasinghe, 2021). However, with the newly emerging advanced techniques like full-length translating mRNA sequencing and ribosome profiling, the long debate whether non-coding RNAs taking account for 98% of the human genome can encode functional proteins besides short peptides seems to be ended by recent studies. Research surprisingly confirmed a hidden human functional proteome encoded by specific lncRNAs, suggesting that the annotation of lncRNAs used before is outdated in some way (Lu et al., 2019). Therefore, we inspected the coding potential of

MAFG-ASI even though it does not seem to encode functional proteins according to the result. Additionally, *MAFG-ASI* is not included in our constructed ceRNA network but interacts with several proteins to involve some mechanisms inside the nucleus.

Our current study is just one small step into exploring the valuable predictive model based on RNA m5C modification in PCa. Hence, in-depth validation with a clinical trial is definitely needed to confirm its exact value, and further experimental confirmation is also required to illuminate lncRNA *MAFG-ASI*'s effects on PCa. Despite the above limitations, we believe that our predictive model based on m5C-related lncRNA signatures will offer some valuable information to the current clinical predicament in PCa.

To sum, our study revealed that the m5C-related lncRNA signatures could precisely predict BCR in patients with PCa by m5C-lnc score. Besides, the nomogram integrated with m5C-related lncRNA signatures and other clinicopathological features presented the customized BCR-free probability prediction. *MAFG-ASI* was ultimately chosen for further experimental verification, and it was upregulated in PCa and probably promoted PCa aggression *in vitro*.

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**.

AUTHOR CONTRIBUTIONS

XM and PZ conceived and designed the study. SW and JL collected the public data. KW, HL, YG, and TY analyzed and interpreted the data. KW, ZL, WZ, and CZ carried out the experimental validation. KW, ZL, and WZ drafted and revised the article. All authors have read and approved the manuscript.

FUNDING

This research was supported by grants from China's National Natural Science Foundation (82173039, 81773277, 82003271), Science and Technology Program of Guangzhou (201803010014), and Guangdong Province Basic and Applied Basic Research Fund Project (2021A1515010659).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.775304/full#supplementary-material>

REFERENCES

- Agostini, F., Zanzoni, A., Klus, P., Marchese, D., Cirillo, D., and Tartaglia, G. G. (2013). catRAPID Omics: a Web Server for Large-Scale Prediction of Protein-RNA Interactions. *Bioinformatics* 29 (22), 2928–2930. doi:10.1093/bioinformatics/btt495
- Amort, T., Soulière, M. F., Wille, A., Jia, X.-Y., Fiegl, H., Wörle, H., et al. (2013). Long Non-coding RNAs as Targets for Cytosine Methylation. *RNA Biol.* 10 (6), 1002–1008. doi:10.4161/rna.24454
- Archer, N. P., Perez-Andreu, V., Scheurer, M. E., Rabin, K. R., Peckham-Gregory, E. C., Plon, S. E., et al. (2016). Family-based Exome-wide Assessment of Maternal Genetic Effects on Susceptibility to Childhood B-Cell Acute Lymphoblastic Leukemia in Hispanics. *Cancer* 122 (23), 3697–3704. doi:10.1002/cncr.30241
- Baena-Del Valle, J. A., Zheng, Q., Esopi, D. M., Rubenstein, M., Hubbard, G. K., Moncaliano, M. C., et al. (2018). MYC Drives Overexpression of Telomerase RNA (hTR/TERC) in Prostate Cancer. *J. Pathol.* 244 (1), 11–24. doi:10.1002/path.4980
- Bai, Y., Ren, C., Wang, B., Xue, J., Li, F., Liu, J., et al. (2021). lncRNA MAFG-AS1 Promotes the Malignant Phenotype of Ovarian Cancer by Upregulating NFKB1-dependent IGF1. *Cancer Gene Ther.* doi:10.1038/s41417-021-00306-8
- Barbieri, I., and Kouzarides, T. (2020). Role of RNA Modifications in Cancer. *Nat. Rev. Cancer* 20 (6), 303–322. doi:10.1038/s41568-020-0253-2
- Blanco, S., Bandiera, R., Popis, M., Hussain, S., Lombard, P., Aleksic, J., et al. (2016). Stem Cell Function and Stress Response Are Controlled by Protein Synthesis. *Nature* 534 (7607), 335–340. doi:10.1038/nature18282
- Bohnsack, K., Höbartner, C., and Bohnsack, M. (2019). Eukaryotic 5-methylcytosine (m5C) RNA Methyltransferases: Mechanisms, Cellular Functions, and Links to Disease. *Genes* 10 (2), 102. doi:10.3390/genes10020102
- Carella, A., Tejedor, J. R., García, M. G., Urduingio, R. G., Bayón, G. F., Sierra, M., et al. (2020). Epigenetic Downregulation of TET3 Reduces Genome-wide 5hmC Levels and Promotes Glioblastoma Tumorigenesis. *Int. J. Cancer* 146 (2), 373–387. doi:10.1002/ijc.32520
- Chen, T., Huang, B., and Pan, Y. (2021). Long Non-coding RNA MAFG-AS1 Promotes Cell Proliferation, Migration, and EMT by miR-3196/STRN4 in Drug-Resistant Cells of Liver Cancer. *Front. Cell Dev. Biol.* 9, 688603. doi:10.3389/fcell.2021.688603
- Chen, X., Li, A., Sun, B.-F., Yang, Y., Han, Y.-N., Yuan, X., et al. (2019). 5-methylcytosine Promotes Pathogenesis of Bladder Cancer through Stabilizing mRNAs. *Nat. Cell Biol.* 21 (8), 978–990. doi:10.1038/s41556-019-0361-y
- Cheng, J. X., Chen, L., Li, Y., Cloe, A., Yue, M., Wei, J., et al. (2018). RNA Cytosine Methylation and Methyltransferases Mediate Chromatin Organization and 5-azacytidine Response and Resistance in Leukaemia. *Nat. Commun.* 9 (1), 1163. doi:10.1038/s41467-018-03513-4
- Cox, D. R. (1972). Regression Models and Life-Tables. *J. R. Stat. Soc. Ser. B (Methodological)* 34 (2), 187–202. doi:10.1111/j.2517-6161.1972.tb00899.x
- 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 (5), 507–517. doi:10.1038/s41422-018-0034-6
- Dinescu, S., Ignat, S., Lazar, A., Constantin, C., Neagu, M., and Costache, M. (2019). Epitranscriptomic Signatures in lncRNAs and Their Possible Roles in Cancer. *Genes* 10 (1), 52. doi:10.3390/genes10010052
- Dou, L., Li, X., Ding, H., Xu, L., and Xiang, H. (2020). Prediction of m5C Modifications in RNA Sequences by Combining Multiple Sequence Features. *Mol. Ther. - Nucleic Acids* 21, 332–342. doi:10.1016/j.omtn.2020.06.004
- Engreitz, J. M., Ollikainen, N., and Guttman, M. (2016). Long Non-coding RNAs: Spatial Amplifiers that Control Nuclear Structure and Gene Expression. *Nat. Rev. Mol. Cell Biol.* 17 (12), 756–770. doi:10.1038/nrm.2016.126
- Freedland, S. J., Humphreys, E. B., Mangold, L. A., Eisenberger, M., Dorey, F. J., Walsh, P. C., et al. (2005). Risk of Prostate Cancer-specific Mortality Following Biochemical Recurrence after Radical Prostatectomy. *JAMA* 294 (4), 433–439. doi:10.1001/jama.294.4.433
- Gao, Y., Wang, Z., Zhu, Y., Zhu, Q., Yang, Y., Jin, Y., et al. (2019). NOP2/Sun RNA Methyltransferase 2 Promotes Tumor Progression via its Interacting Partner RPL6 in Gallbladder Carcinoma. *Cancer Sci.* 110 (11), 3510–3519. doi:10.1111/cas.14190
- García, M. G., Carella, A., Urduingio, R. G., Bayón, G. F., Lopez, V., Tejedor, J. R., et al. (2018). Epigenetic Dysregulation of TET2 in Human Glioblastoma. *Oncotarget* 9 (40), 25922–25934. doi:10.18632/oncotarget.25406
- Goodall, G. J., and Wickramasinghe, V. O. (2021). RNA in Cancer. *Nat. Rev. Cancer* 21 (1), 22–36. doi:10.1038/s41568-020-00306-0
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long Non-coding RNA HOTAIR Reprograms Chromatin State to Promote Cancer Metastasis. *Nature* 464 (7291), 1071–1076. doi:10.1038/nature08975
- He, Y., Yu, X., Li, J., Zhang, Q., Zheng, Q., and Guo, W. (2020). Role of m5C-Related Regulatory Genes in the Diagnosis and Prognosis of Hepatocellular Carcinoma. *Am. J. Transl. Res.* 12 (3), 912–922. doi:10.18632/aging.102669
- 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, Z., Pan, J., Wang, H., Du, X., Xu, Y., Wang, Z., et al. (2021). Prognostic Significance and Tumor Immune Microenvironment Heterogeneity of m5C RNA Methylation Regulators in Triple-Negative Breast Cancer. *Front. Cell Dev. Biol.* 9, 657547. doi:10.3389/fcell.2021.657547
- Janin, M., Ortiz-Barahona, V., de Moura, M. C., Martínez-Cardús, A., Llinàs-Arias, P., Soler, M., et al. (2019). Epigenetic Loss of RNA-Methyltransferase NSUN5 in Glioma Targets Ribosomes to Drive a Stress Adaptive Translational Program. *Acta Neuropathol.* 138 (6), 1053–1074. doi:10.1007/s00401-019-02062-4
- Jia, H., Wu, D., Zhang, Z., and Li, S. (2021). Regulatory Effect of the MAFG-AS1/miR-150-5p/MYB axis on the P-roliferation and M-igration of B-reast C-ancer C-ells. *Int. J. Oncol.* 58 (1), 33–44. doi:10.3892/ijo.2020.5150
- Kaplan, E. L., and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* 53 (282), 457–481. doi:10.2307/2281868, 10.1080/01621459.1958.10501452
- Kreuz, M., Otto, D. J., Fuessel, S., Blumert, C., Bertram, C., Bartsch, S., et al. (2020). ProstateTrend-A Multivariable Prognostic RNA Expression Score for Aggressive Prostate Cancer. *Eur. Urol.* 78 (3), 452–459. doi:10.1016/j.euro.2020.06.001
- Ku, S.-Y., Gleave, M. E., and Beltran, H. (2019). Towards Precision Oncology in Advanced Prostate Cancer. *Nat. Rev. Urol.* 16 (11), 645–654. doi:10.1038/s41585-019-0237-8
- Li, R., Qu, H., Wang, S., Wei, J., Zhang, L., Ma, R., et al. (2018a). GDCRNATools: an R/Bioconductor Package for Integrative Analysis of lncRNA, miRNA and mRNA Data in GDC. *Bioinformatics (Oxford, England)* 34 (14), 2515–2517. doi:10.1093/bioinformatics/bty124
- Li, R., Wang, S., Cui, Y., Qu, H., Chater, J. M., Zhang, L., et al. (2021a). Extended Application of Genomic Selection to Screen Multiomics Data for Prognostic Signatures of Prostate Cancer. *Brief. Bioinformatics* 22 (3), bbaa197. doi:10.1093/bib/bbaa197
- Li, R., Zhu, J., Zhong, W.-D., and Jia, Z. (2021b). PCaDB - a Comprehensive and Interactive Database for Transcriptomes from Prostate Cancer Population Cohorts. *bioRxiv*. doi:10.1101/2021.06.29.449134
- 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
- Liang, Z.-X., Liu, H.-S., Wang, F.-W., Xiong, L., Zhou, C., Hu, T., et al. (2019). lncRNA RPPH1 Promotes Colorectal Cancer Metastasis by Interacting with TUBB3 and by Promoting Exosomes-Mediated Macrophage M2 Polarization. *Cell Death Dis* 10 (11), 829. doi:10.1038/s41419-019-2077-0
- Lu, S., Zhang, J., Lian, X., Sun, L., Meng, K., Chen, Y., et al. (2019). A Hidden Human Proteome Encoded by 'non-Coding' Genes. *Nucleic Acids Res.* 47 (15), 8111–8125. doi:10.1093/nar/gkz646
- Mandrekar, J. N. (2010). Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *J. Thorac. Oncol.* 5 (9), 1315–1316. doi:10.1097/JTO.0b013e3181ec173d
- Mardis, E. R. (2013). Next-generation Sequencing Platforms. *Annu. Rev. Anal. Chem.* 6, 287–303. doi:10.1146/annurev-anchem-062012-092628
- Mei, L., Shen, C., Miao, R., Wang, J.-Z., Cao, M.-D., Zhang, Y.-S., et al. (2020). RNA Methyltransferase NSUN2 Promotes Gastric Cancer Cell Proliferation by Repressing p57Kip2 by an m5C-dependent Manner. *Cell Death Dis* 11 (4), 270. doi:10.1038/s41419-020-2487-z

- National Cancer Institute (2021). SEER*Explorer SEER*Explorer: An Interactive Website for SEER Cancer Statistics [Internet]. *Surveillance Research Program, National Cancer Institute*[Online] Available from <https://seer.cancer.gov/explorer/>(Cited April 15, 2021).[Accessed]
- Okamoto, M., Hirata, S., Sato, S., Koga, S., Fujii, M., Qi, G., et al. (2012). Frequent Increased Gene Copy Number and High Protein Expression of tRNA (Cytosine-5-)-methyltransferase (NSUN2) in Human Cancers. *DNA Cel. Biol.* 31 (5), 660–671. doi:10.1089/dna.2011.1446
- Pan, J., Huang, Z., and Xu, Y. (2021). m5C-Related lncRNAs Predict Overall Survival of Patients and Regulate the Tumor Immune Microenvironment in Lung Adenocarcinoma. *Front. Cel Dev. Biol.* 9, 671821. doi:10.3389/fcell.2021.671821
- Prensner, J. R., Iyer, M. K., Sahu, A., Asangani, I. A., Cao, Q., Patel, L., et al. (2013). The Long Noncoding RNA SChLAP1 Promotes Aggressive Prostate Cancer and Antagonizes the SWI/SNF Complex. *Nat. Genet.* 45 (11), 1392–1398. doi:10.1038/ng.2771
- Rong, D., Sun, G., Wu, F., Cheng, Y., Sun, G., Jiang, W., et al. (2021). Epigenetics: Roles and Therapeutic Implications of Non-coding RNA Modifications in Human Cancers. *Mol. Ther. - Nucleic Acids* 25, 67–82. doi:10.1016/j.omtn.2021.04.021
- Roundtree, I. A., Evans, M. E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. *Cell* 169 (7), 1187–1200. doi:10.1016/j.cell.2017.05.045
- Roychowdhury, S., and Chinnaiyan, A. M. (2016). Translating Cancer Genomes and Transcriptomes for Precision Oncology. *CA: a Cancer J. clinicians* 66 (1), 75–88. doi:10.3322/caac.21329
- Ruan, Z., Deng, H., Liang, M., Xu, Z., Lai, M., Ren, H., et al. (2020). Downregulation of Long Non-coding RNA MAFG-AS1 Represses Tumorigenesis of Colorectal Cancer Cells through the microRNA-149-3p-dependent Inhibition of HOXB8. *Cancer Cel Int* 20, 511. doi:10.1186/s12935-020-01485-4
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P. P. (2011). A ceRNA Hypothesis: the Rosetta Stone of a Hidden RNA Language. *Cell* 146 (3), 353–358. doi:10.1016/j.cell.2011.07.014
- Sato, K., Tahata, K., and Akimoto, K. (2020). Five Genes Associated with Survival in Patients with Lower-Grade Gliomas Were Identified by Information-Theoretical Analysis. *Anticancer Res.* 40 (5), 2777–2785. doi:10.21873/anticancer.14250
- Schober, P., Boer, C., and Schwarte, L. A. (2018). Correlation Coefficients. *Anesth. Analgesia* 126 (5), 1763–1768. doi:10.1213/ANE.0000000000002864
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303
- Siegel, R. L., Miller, K. D., Fuchs, H. E., and Jemal, A. (2021). Cancer Statistics. *CA A. Cancer J. Clin.* 71 (1), 7–33. doi:10.3322/caac.21654, 2021
- Singer, J., Irmisch, A., Ruscheweyh, H.-J., Singer, F., Toussaint, N. C., Levesque, M. P., et al. (2019). Bioinformatics for Precision Oncology. *Brief. Bioinformatics* 20 (3), 778–788. doi:10.1093/bib/bbx143
- Squires, J. E., Patel, H. R., Nousch, M., Sibbritt, T., Humphreys, D. T., Parker, B. J., et al. (2012). Widespread Occurrence of 5-methylcytosine in Human Coding and Non-coding RNA. *Nucleic Acids Res.* 40 (11), 5023–5033. doi:10.1093/nar/gks144
- Sui, Y., Lin, G., Zheng, Y., and Huang, W. (2019). lncRNA MAFG-AS1 Boosts the Proliferation of Lung Adenocarcinoma Cells via Regulating miR-744-5p/MAFG axis. *Eur. J. Pharmacol.* 859, 172465. doi:10.1016/j.ejphar.2019.172465
- 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 (45), 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 (3), 209–249. doi:10.3322/caac.21660
- Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., et al. (2014). 5-Hydroxymethylcytosine Plays a Critical Role in Glioblastomagenesis by Recruiting the CHTOP-Methylosome Complex. *Cel Rep.* 9 (1), 48–60. doi:10.1016/j.celrep.2014.08.071
- Tan, X., Chen, W.-b., Lv, D.-j., Yang, T.-w., Wu, K.-h., Zou, L.-b., et al. (2021). lncRNA SNHG1 and RNA Binding Protein hnRNPL Form a Complex and Coregulate CDH1 to Boost the Growth and Metastasis of Prostate Cancer. *Cel Death Dis* 12 (2), 138. doi:10.1038/s41419-021-03413-4
- Tang, Z., Li, C., Kang, B., Gao, G., Li, C., and Zhang, Z. (2017). GEPIA: a Web Server for Cancer and normal Gene Expression Profiling and Interactive Analyses. *Nucleic Acids Res.* 45 (W1), W98–W102. doi:10.1093/nar/gkx247
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B (Methodological)* 58 (1), 267–288. doi:10.1111/j.2517-6161.1996.tb02080.x
- Tolles, J., and Meurer, W. J. (2016). Logistic Regression. *JAMA* 316 (5), 533–534. doi:10.1001/jama.2016.7653
- Volders, P.-J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., et al. (2019). LNCipedia 5: towards a Reference Set of Human Long Non-coding RNAs. *Nucleic Acids Res.* 47 (D1), D135–D139. doi:10.1093/nar/gky1031
- Xiao, M., Liu, J., Xiang, L., Zhao, K., He, D., Zeng, Q., et al. (2020). MAFG-AS1 Promotes Tumor Progression via Regulation of the HuR/PTBP1 axis in Bladder Urothelial Carcinoma. *Clin. Translational Med.* 10 (8), e241. doi:10.1002/ctm2.241
- Yamashita, T., Higashi, M., Momose, S., Morozumi, M., and Tamaru, J.-I. (2017). Nuclear Expression of Y Box Binding-1 Is Important for Resistance to Chemotherapy Including Gemcitabine in TP53-Mutated Bladder Cancer. *Int. J. Oncol.* 51 (2), 579–586. doi:10.3892/ijo.2017.4031
- Yang, X., Yang, Y., Sun, B.-F., Chen, Y.-S., Xu, J.-W., Lai, W.-Y., et al. (2017). 5-methylcytosine Promotes mRNA export - NSUN2 as the Methyltransferase and ALYREF as an m5C Reader. *Cel Res* 27 (5), 606–625. doi:10.1038/cr.2017.55
- Ye, L., Feng, W., Weng, H., Yuan, C., Liu, J., and Wang, Z. (2020). MAFG-AS1 Aggravates the Progression of Pancreatic Cancer by Sponging miR-3196 to Boost NFIX. *Cancer Cel Int* 20 (1), 591. doi:10.1186/s12935-020-01669-y
- Zhao, B., Lu, Y.-L., Yang, Y., Hu, L.-B., Bai, Y., Li, R.-Q., et al. (2018). Overexpression of lncRNA ANRIL Promoted the Proliferation and Migration of Prostate Cancer Cells via Regulating Let-7a/TGF-B1/Smad Signaling Pathway. *Cbm* 21 (3), 613–620. doi:10.3233/CBM-170683
- Zhong, C., Wu, K., Wang, S., Long, Z., Yang, T., Zhong, W., et al. (2021). Autophagy-related circRNA Evaluation Reveals Hsa_circ_0001747 as a Potential Favorable Prognostic Factor for Biochemical Recurrence in Patients with Prostate Cancer. *Cel Death Dis* 12 (8), 726. doi:10.1038/s41419-021-04015-w

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.

Copyright © 2021 Wang, Zhong, Long, Guo, Zhong, Yang, Wang, Lai, Lu, Zheng and Mao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.