Submitted: 18 July, 2022 Accepted: 19 August, 2022 Published: 15 April, 2023 DOI:10.22514/ejgo.2023.024

ORIGINAL RESEARCH

European Journal of Gynaecological Oncology

N6-methyladenosine-related genes contribute to malignant progression, have clinical prognostic and neoadjuvant treatments response impact for breast cancer

Jun Shen[1], Hongfang Ma[2], Yongxia Chen[3], Cong Chen[1], Zhaoqing Li[1], Jianguo Shen[1,*]

[1]Department of Surgical Oncology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 310016 Hangzhou, Zhejiang, China
[2]Department of Plastic Surgery, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 310016 Hangzhou, Zhejiang, China
[3]Laboratory of Cancer Biology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 310016 Hangzhou, Zhejiang, China

***Correspondence**
drshenjianguo@zju.edu.cn
(Jianguo Shen)

Abstract

N6-methyladenosine (m6A) methylation dysregulation contributes to tumorigenesis and breast cancer development. This study intends to conduct a comprehensive analysis for determining the clinical significance of m6A-related genes and establishing m6A-related gene-based risk signature to predict the clinical outcomes and neoadjuvant treatments response for breast cancer patients. The m6Avar database was utilized for downloading the m6A-regulated genes. The Cancer Genome Atlas (TCGA) database was utilized for downloading breast cancer patients' RNA-Seq data and clinicopathological information. For determining the differentially expressed m6A-related gene, a one-way analysis of variance (ANOVA) was conducted. The interaction and correlation of m6A-related genes were evaluated using search tool for the retrieval of interacting genes/proteins (STRING) and Spearmen test. For determining clusters of breast cancer patients with different clinical outcomes, a consensus clustering analysis was conducted. We screened differentially expressed genes and functional enrichment pathways between subgroups utilizing gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG). We constructed and verified a prognostic signature utilizing Cox regression analysis as well as a least absolute shrinkage and selection operator (LASSO) regression model. 286 genes were detected as significantly differentially expressed in different stages, including 3 m6A RNA methylation regulators, Wilms tumor 1-associating protein (*WTAP*), YT521-B homology (YTH) domain containing 2 (*YTHDC2*), and YTH domain family 2 (*YTHDF2*). A 13-gene prognostic signature was constructed and could predict the overall survival and the neoadjuvant treatments response in breast cancer patients. We categorized breast cancer patients into four groups based on m6A-associated RNAs expression. Significant differences were found in the overall survivals among the four clusters of patients. The biological processes and the key signaling pathways closely related to breast cancer have a close connection to the four clusters. This study confirmed that the m6A-related genes expression levels were highly associated with prognosis and neoadjuvant treatment response in breast cancer and constructed an effective m6A-related gene-based risk signature for predicting the prognosis of patients with breast cancer.

Keywords

Breast cancer; N6-methyladenosine methylation; Prognostic signature; Overall survival; Neoadjuvant treatment

1. Introduction

In the United States, the leading cancer among women is breast cancer, with approximately 276,480 new cases and 42,170 deaths in 2020 [1]. Previous studies have suggested lymph node status, tumor size, and tissue grade as predictors for the clinical outcomes of breast cancer patients [2–4], although these factors' accuracy varied when applied to different populations [5]. It is critical to determine accurate and reliable prognostic markers for breast cancer.

In breast cancer, considerable genetic and epigenetic alterations could be utilized as biomarkers for cancer detection, treatment and prognosis [5, 6]. The most prevalent epigenetic modification that occurs in the N6-position of adenosine in eukaryotic mRNA is N6-methyladenosine (m6A) [7]. m6A methylation is enriched in the RRACH motif (R: A/G, H: A/C/U) near the stop codon, 3' untranslated regions, and

the internal long exons [8], being installed by methyltransferases (writers), determined by methyl-binding proteins (readers), and removed by demethylases (erasers) [9, 10]. The m6A pattern affects the expression of the modified RNA, thus influencing the corresponding biological processes and functions. Dysregulation of m6A methylation has a role in pathogenesis and development of various human diseases, including breast cancer, bladder cancer, and endometrial cancer [11, 12]. Breast cancer progression was reported to associate with the alterations in m6A writers such as methyltransferase like 3 (METTL3) [13] and WT1-associated protein (WTAP) [12], with the reader such as YT521-B homology (YTH) N6-methyladenosine RNA-binding proteins (YTHDF3) [14], and with the eraser such as alkB homolog 5 (ALKBH5) [15]. Even so, m6A role in breast cancer and m6A-related genes prognostic value in breast cancer remains obscure. In this study, using the Cancer Genomic Atlas (TCGA), we determined m6A-related genes expression is significantly differential among different breast cancer stages. The interaction and correlation among these m6A-related genes were evaluated. Taking advantage of the consensus clustering analysis, a list of genes that could classify breast cancer patients with different clinical outcomes was obtained. Then, a 13-gene risk signature was simulated within TCGA breast cancer cohort and it showed promising performance of this risk signature for prognosis prediction. Furthermore, we validated this prognostic signature in a TCGA validation cohort and an independent external breast cancer cohort.

2. Materials and methods

2.1 Data collection

For analyzing m6A role in breast cancer, the TCGA database was utilized to acquire the RNA-Seq expression profile and clinical characteristics of 1218 breast cancer samples.

To verify the prognostic signature in TCGA for neoadjuvant treatment response in patients with breast cancer, we excluded the databases with small sample size, lack of survival result or unitary molecular type. Finally, we collected data from two independent cohorts using the Gene Expression Omnibus (GEO) database, which included data of 310 breast cancer patients who received neoadjuvant treatments from GSE25055 and 182 from GSE25065. The candidate m6A-related genes associated with breast cancer were obtained from review articles published from 2017 to 2019 and from the m6avar database (http://m6avar.renlab.org/).

2.2 Consensus clustering Analysis

We discovered differentially expressed m6A-related genes ($p < 0.05$) in different breast cancer stages according to the RNA-Seq expression profile of breast cancer samples from the TCGA database. These genes were utilized as classifiers for recognizing breast cancer subtypes. Consensus clustering was performed to classify the breast cancer samples using the R package of "Consensus Cluster Plus". We computed gap statistics [15] from K = 2 to 10 for determining the optimal of clusters number. The optimal K was considered as the K where the proportional area change under the cumulative distribution function (CDF) curve reaches its maximum.

2.3 Construction of protein-protein interaction (PPI) network

We constructed a PPI network and mapped differentially expressed m6A-related genes (n = 286) to the Search Tool for the Retrieval of Interaction Genes database (STRING, http://string.embl.de/). The p-value was less than 0.01, and the minimum required interaction score was 0.3.

2.4 Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway analysis

The Database for Annotation Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov/) was utilized to conduct GO annotation and KEGG pathway enrichment analyses. For biological pathway analysis, p-value $<$ 0.001 and gene counts >2 were set as threshold. For KEGG pathway analysis, p-value < 0.05 and gene counts >2 were set.

2.5 Construction and validation of a prognostic signature based on m6A-related genes

The differentially expressed m6A-related genes (n = 286) in various stages of breast cancer were analyzed utilizing the R package 'Survival' related to patient's prognosis. The relation between differentially expressed genes and overall patient survival was determined utilizing multivariate Cox regression analysis in the TCGA training dataset (n = 720). The genes with $p < 0.05$ were defined as candidate survival-related genes. For identifying survival-related genes, we constructed a least absolute shrinkage and selection operator (LASSO) regression model utilizing the R package glmnet. We calculated each patient's risk score as follows:

$$\text{Risk score} = \sum_{i=1}^{n} \text{Coef}_i \times \text{Exp}_i$$

In compliance with the median level of risk score, we classified patients into high-risk and low-risk groups. We established Kaplan-Meier survival curves. The gene-based prognostic model's sensitivity and specificity in prediction of patients' clinical outcomes were assessed using receiver operating characteristic (ROC) analysis. The risk scores in the TCGA validation (n = 480) cohort were determined for evaluating the prognostic model's strength. Additionally, the 13-gene-based risk scores were computed for patients in GSE25065 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25065) and GSE25055 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE25055) databases to further verify the prognostic and predictive values of neoadjuvant treatment response of the risk signature. Since, the panel, which is used in the combined dataset of GSE25055 and GSE25065, is not contained 4 of the 13 genes. We assigned the four missing genes values as zeros.

TABLE 1. Clinicopathological characteristics of patients with breast cancer from the TCGA database.

parameter	subtype	n
Age (yr)		
	>50	838
	≤50	380
Gender		
	female	1202
	male	13
	Unknown	3
Pathologic stage		
	I	202
	II	691
	III	277
	IV	22
	Unknown	26
Pathologic-M		
	M0	1021
	M1	24
	MX	173
	Unknown	8
Pathologic-N		
	N0	561
	N1	416
	N2	132
	N3	83
	Unknown	26
Pathologic-T		
	T1	310
	T2	705
	T3	150
	T4	47
	Unknown	6
PAM50 Call-RNAseq		
	Basal	142
	LumA	434
	LumB	194
	Her2	67
	Normal	119
	Unknown	262
Tumor status		
	Distant Metastasis	67
	Locoregional Disease	9
	Locoregional Recurrence	13
	New Primary Tumor	21
	Unknown	1108

PAM50: Prediction Analysis of Microarray 50.

2.6 Statistical analysis

Statistical analysis was done utilizing the R packages listed above under R software (version 3.6.0, The Free Software Foundation, Inc., Boston, MA, USA). The m6A-related gene expressions in different breast cancer stages were compared utilizing univariate ANOVA. $p < 0.05$ was the threshold. Clinical data were treated as categorical variables. We used Chi-

square test when detecting the relationship between categorical variables. PPI network was analyzed using Spearman's rank correlation coefficient, followed by rank-sum test.

3. Results

3.1 The differentially expressed m6A-related genes in different stages of breast cancer

The normalized RNA-Seq data of 1218 breast cancer samples were downloaded through the TCGA database. Table 1 illustrates the clinical characteristics. Among these samples, 202 were classified in stage I, 691 were in stage II, 277 were in stage III, and 22 were in stage IV. The 26 samples without pathologic stage information were excluded.

The process of bioinformatic analysis is shown in Fig. 1A. By data mining from published literatures, 16 candidate m6A-related genes associated with breast cancer were selected, comprising methyltransferase 3, N6-adenosine-methyltransferase complex catalytic subunit (*METTL3*), insulin like growth factor 2 mRNA binding protein 1 (*IGF2BP1*), methyltransferase 14, N6-adenosine-methyltransferase subunit (*METTL14*), alkB homolog 5, RNA demethylase (*ALKBH5*), vir like m6A methyltransferase associated (*KIAA1429*), zinc finger CCCH-type containing 13 (*ZC3H13*), YTH domain containing 1 (*YTHDC1*), RNA binding motif protein 15 (*RBM15*), *YTHDC2*, *YTHDF2*, heterogeneous nuclear ribonucleoprotein C (*HNRNPC*), FTO alpha-ketoglutarate dependent dioxygenase (*FTO*), *WTAP*, methyltransferase 16, N6-methyladenosine (*METTL16*), YTH N6-methyladenosine RNA binding protein 3 (*YTHDF3*), and YTH N6-methyladenosine RNA binding protein 1 (*YTHDF1*) [16–18]. Additionally, 1630 m6A-related genes associated with breast cancer was found by the m6avar database examination (http://m6avar.renlab.org/). After excluding the genes without expression values in the breast cancer samples or expressed in fewer than 80% of samples, we obtained 1331 candidate genes.

For determining m6A-related genes associated with breast cancer staging, we utilized univariate analysis and obtained 286 m6A-related genes that were differentially expressed across the 4 stages ($p < 0.05$), comprising 3 m6A RNA methylation regulators: *WTAP*, *YTHDC2*, and *YTHDF2*. For visualizing 286 genes expression in the TCGA samples, a heatmap was constructed (Fig. 1B).

3.2 m6A-related genes consensus clustering identifies four breast cancer samples clusters with various clinical outcomes

Following the determination of the 286 m6A-related genes that were differentially expressed across various breast cancer stages, consensus clustering was performed to classify the 1218 patients from TCGA. We computed gap statistics for K = 2 to 10 (Fig. 2A–B). The proportional area change under the CDF curve peaked at K = 4, suggesting that four subtypes were optimal for categorizing the patient samples. Therefore, we classified the 1218 breast cancer samples into type of malignancy (TM) 1 (n = 304), TM2 (n = 398), TM3 (n = 510), and TM4 (n = 6) subgroups (Fig. 2C). By comparing the clinical data of these subgroups, we observed significant differences in gender, N stage, T stage, and Prediction Analysis of Microarray 50 (PAM50) subtype among these subgroups (all $p < 0.05$; Fig. 2D). In the M stage or pathologic stage, no significant difference was observed. Furthermore, Kaplan-Meier survival analysis demonstrated that the overall survivals differed significantly across the subgroups in descending order of TM1, TM2, TM3, and TM4 ($p = 0.0020$; Fig. 2E), suggesting that the 286 genes could classify the breast cancer samples with different clinical outcomes.

3.3 The correlation and interaction among the m6A-related genes

For interactions examination among the 286 m6A-related genes, the protein-protein interactions were predicted utilizing the STRING database. The interaction network is displayed in Fig. 3A, including the interactions among YTHDC2, WTAP, and YTHDF2. Spearman correlation analysis revealed 8781 pairs of significantly correlated m6A-related genes (Fig. 3B).

3.4 Functional annotation of m6A-related genes

Next, to understand the m6A-related genes' biological significance in breast cancer, functional annotation was conducted. We selected 239 m6A-related genes that were differentially expressed across TM1–4 subgroups (one-way ANOVA, $p < 0.05$). As shown in Fig. 4A, these genes were mainly annotated with biological processes, including chromatin remodeling, signal transduction, positive regulation of protein tyrosine kinase activity, histone H3-K4 methylation, and cytokine production. According to KEGG pathway analysis, these genes are mainly enriched in microRNA in cancer, Phosphoinositide 3-Kinase (PI3K)- protein kinase B (PKB/Akt) (PI3K-Akt) pathway, and calcium signaling pathway (Fig. 4B). Furthermore, ATP binding cassette subfamily B member 5 (*ABCB5*), adaptor related protein complex 4 subunit epsilon 1 (*AP4E1*), Rho guanine nucleotide exchange factor 19 (*ARHGEF19*), Rho guanine nucleotide exchange factor 3 (*ARHGEF3*), cholinergic receptor nicotinic alpha 6 subunit (*CHRNA6*), PHD and ring finger domains 1 (*PHRF1*), PR/SET domain 10 (*PRDM10*), TBC1 domain family member 15 (*TBC1D15*), tubulin alpha 3d (*TUBA3D*), and Wnt family member 3A (*WNT3A*) were identified as independent prognostic factors, serving as therapeutic targets and potential biomarkers for breast cancer prognosis and treatment.

3.5 Validation and identification of gene-based prognostic signature

To evaluate and test m6A-related genes prognostic value in breast cancer, we collected 1200 breast cancer samples with the survival data through the TCGA database and randomly split them into 5 cohorts, 3 cohorts for training (n = 720), and 2 cohorts for validation (n = 480). The clinical characteristics are summarized in Table 2. After multivariate Cox regression analysis in the training dataset, we noticed that 20 out of 286 m6A-related genes were related significantly to the

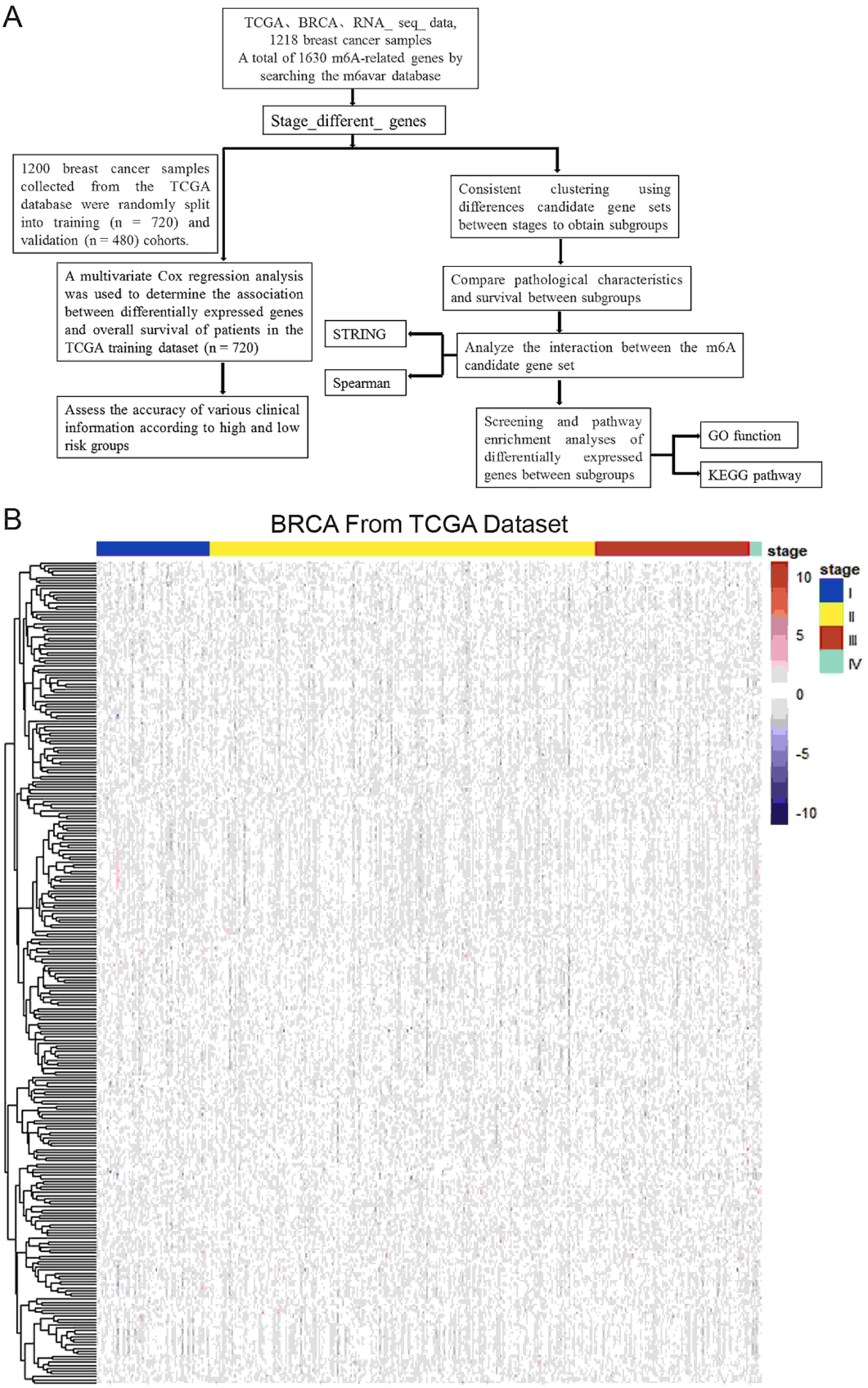

FIGURE 1. The differentially expressed m6A-related genes in different stages of breast cancer. (A) An overview of the process of bioinformatic analysis. (B) The heatmap of the expression of 286 differentially expressed m6A-related genes in breast cancer samples from TCGA database. TCGA: The Cancer Genome Atlas; GO: gene ontology; KEGG: Kyoto encyclopedia of genes and genomes; STRING: search tool for the retrieval of interacting genes/proteins; BRCA: breast cancer.

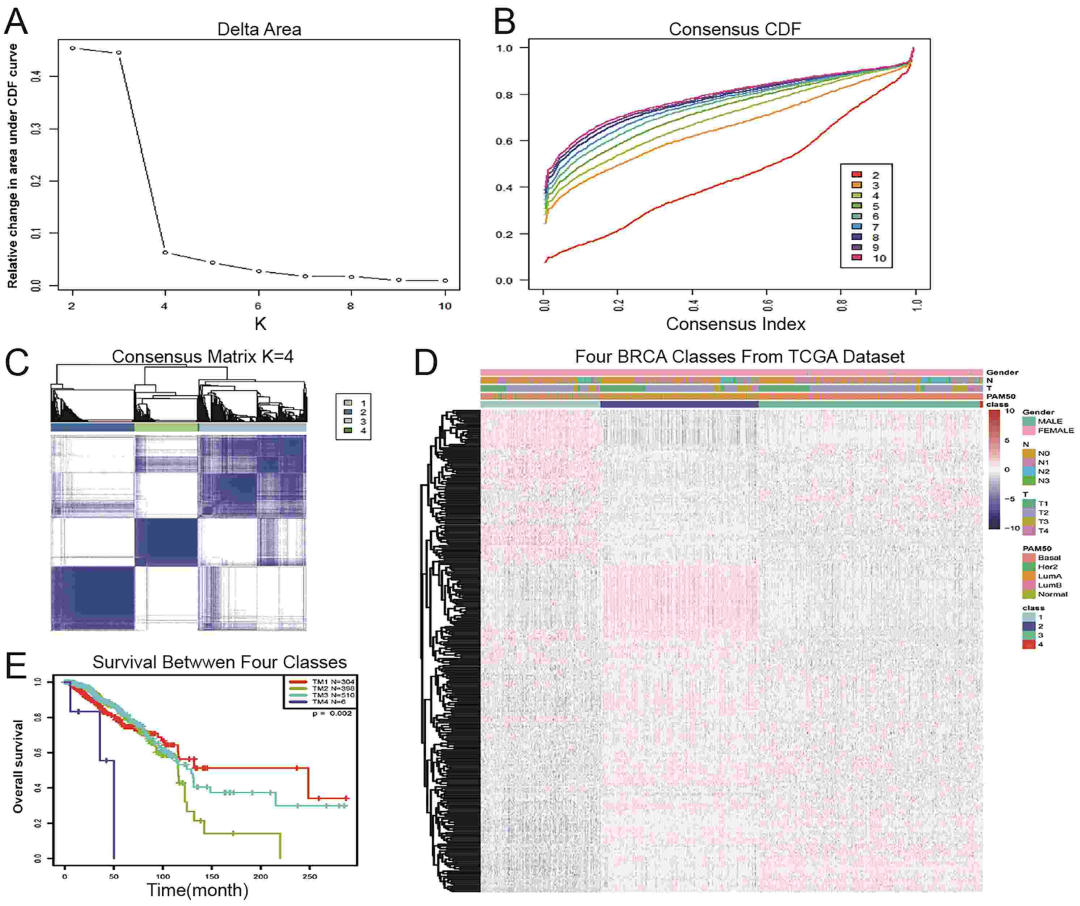

FIGURE 2. The Consensus clustering analysis identified four genetic distinct subgroups in breast cancer patients. (A) Relative change in area under CDF curve for K = 2 to 10. (B) Empirical CDF for K = 2 to 10. (C) Consensus clustering matrix for K = 4. (D) The m6A-related genes consensus expression identified clinicopathologic characteristics and heatmap of the four subgroups (TM1/2/3/4). (E) Kaplan-Meier overall survival curves for various subgroups. TCGA, the Cancer Genome Atlas; CDF, cumulative distribution function. k, the number of categories.

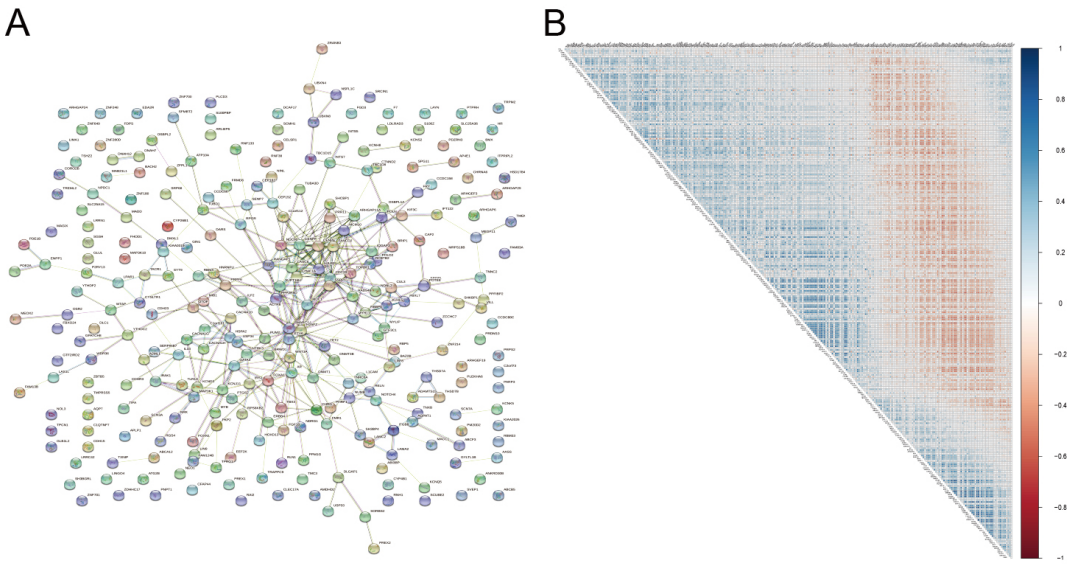

FIGURE 3. The interaction and correlation among the m6A-related genes. (A) Protein-protein interaction graph. Proteins are represented as nodes, and possible relationships are represented by lines. (B) Correlations among 286 m6A-related genes. Spearman's correlations are color-coded to reveal negative (red) or positive (blue) associations.

overall survival (all $p < 0.05$). Fig. 5A displays the p-values and hazard ratios (HR). The LASSO regression model further

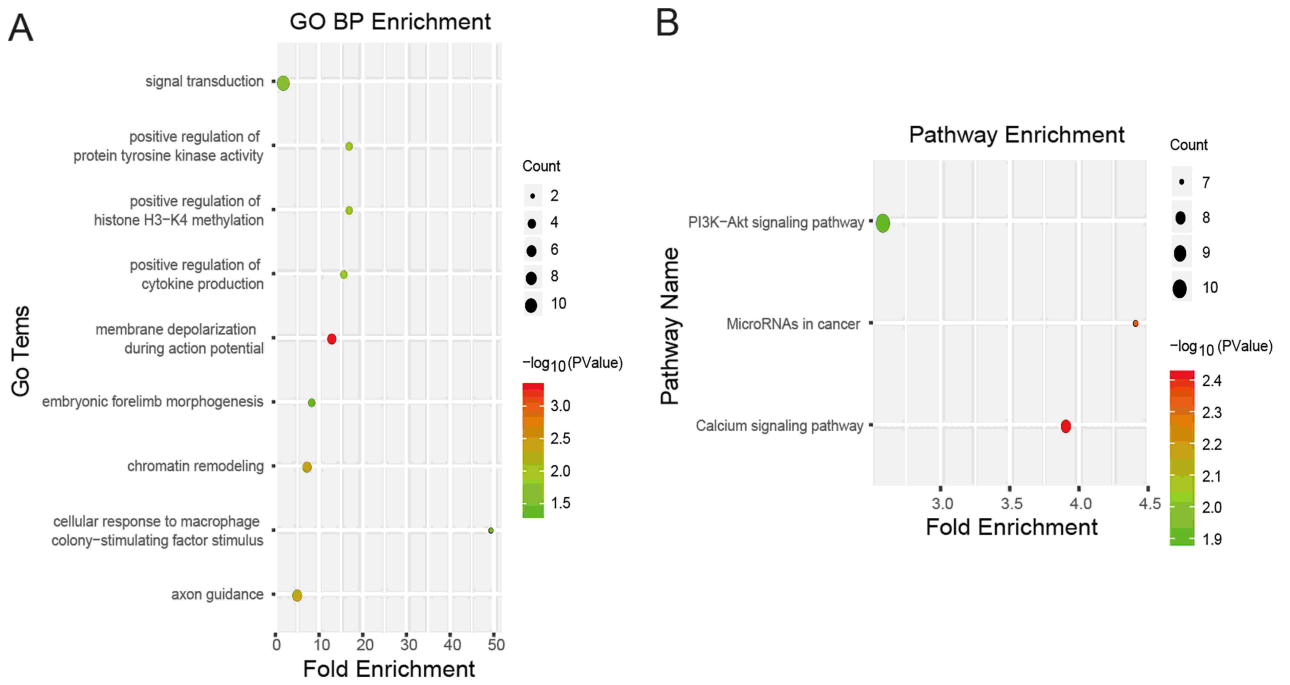

FIGURE 4. Functional annotation of the m6A-related genes. GO analysis (A) and KEGG pathway analysis (B) were conducted on a total of 239 m6A-related genes that were differentially expressed among TM1–4 subgroups. GO, gene ontology; KEGG, Kyoto encyclopedia of genes and genomes.

identified 13 independent prognostic genes (Fig. 5B–C). Thus, for estimating the patients' survival risk, a 13-gene prognostic model was developed.

Then, the risk score was calculated for each patient, and the training and validation cohorts were divided into high- and low-risk groups, respectively, by the median risk score utilization as the cut-off value. Fig. 5D demonstrated that in the training cohort, overall survival in the high-risk group was significantly lower than in the low-risk group (p = 2.42 × 10^{-7}). The validation cohort yielded consistent results (p = 0.0041; Fig. 5E). Fig. 5F–G display the distributions of the overall survivals and risk scores of the two cohorts.

3.6 The prognostic signature-based risk scores can predict the clinical outcomes and the response to neoadjuvant treatment in breast cancer patients

After comparing the entire cohort's clinical data (n = 1200), we observed significant differences in different PAM50 subtypes (p = 4.55 × 10^{-21}) or pathologic stages (p = 0.0402) (Fig. 6A–C), but not in gender, age, pathologic stage, M stage, N stage, T stage, or the mutation in breast cancer 1 (BRCA1) or breast cancer 2 (BRCA2).

Next, we determined whether the 13-gene risk signature has prognostic value in patients with breast cancer. As illustrated in Fig. 6D–H, the risk model performed well in the entire cohort, with the area under curves (AUCs) of ROC curves at 0.641, 0.705, 0.709, 0.683, and 0.725 at 1, 2, 3, 4, and 5 years, respectively. In comparison with long-term disease-free survivors (n = 1090), patients with distant metastasis (n = 67) had significantly lower overall survival, as detailed in Fig. 6I.

Finally, we identified 13-gene risk signature's value for predicting responsiveness to neoadjuvant treatments in patients with breast cancer. Data of 310 breast cancer patients who administered neoadjuvant treatments from GSE25055 and 182 from GSE25065 were acquired and combined into one dataset. The dataset contained RNA-Seq data and clinical information, such as stage, response to neoadjuvant treatments (pathological complete response (pCR) or residual disease (RD)), and survival information. Notably, the panel, which is used in the combined dataset of GSE25055 and GSE25065, is not contained 4 of the 13 genes. We assigned the four missing genes values as zeros, when the 13-gene-based risk score was calculated. For visualizing 9-gene expression, a heatmap was constructed and shown in Fig. 7A. Consistent with the findings presented earlier, the 13-gene-based risk scores were significantly different in different responses to neoadjuvant treatments (p = 0.0082) and pathologic stages (p = 0.0005) in the cohort (Fig. 7B–C). Overall survival of the high-risk group was shorter than that of the low-risk group (Fig. 7D, p = 0.0138). Above all, the prognostic signature-based risk score could be a robust predictor of the clinical outcomes and responsiveness to neoadjuvant treatments in breast cancer patients.

4. Discussion

In this study, m6A-related gene expression was assessed in different breast cancer stages. According to 286 m6A-related gene expression pattern that were differentially expressed across different breast cancer stages, the TCGA breast cancer dataset was classified into four subgroups (TM1–4) with significant differences in the overall survival as well as other

TABLE 2. Clinicopathological characteristics of patients in the training and validation cohorts.

Vairable	Training cohort (n = 720)	Validation cohort (n = 480)
Age (yr)		
>50	492	333
≤50	228	147
Gender		
female	713	474
male	7	6
Pathologic stage		
I	129	68
II	415	279
III	147	117
IV	14	10
Unknown	15	6
Pathologic-M		
M0	595	411
M1	16	8
Unknown	109	61
Pathologic-N		
N0	338	213
N1	248	164
N2	77	55
N3	44	38
Unknown	13	10
Pathologic-T		
T1	197	113
T2	412	284
T3	84	62
T4	24	20
Unknown	2	1
PAM50Call-RNAseq		
Basal	86	54
LumA	250	179
LumB	122	68
Her2	36	28
Normal	67	50
Unknown	159	101
Tumor status		
Distant Metastasis	39	28
Locoregional Disease	6	3
Locoregional Recurrence	5	8
New Primary Tumor	9	12
Unknown	661	429

clinicopathologic features including gender, N stage, T stage, and PAM50 subtypes. Moreover, a 13-gene risk signature was

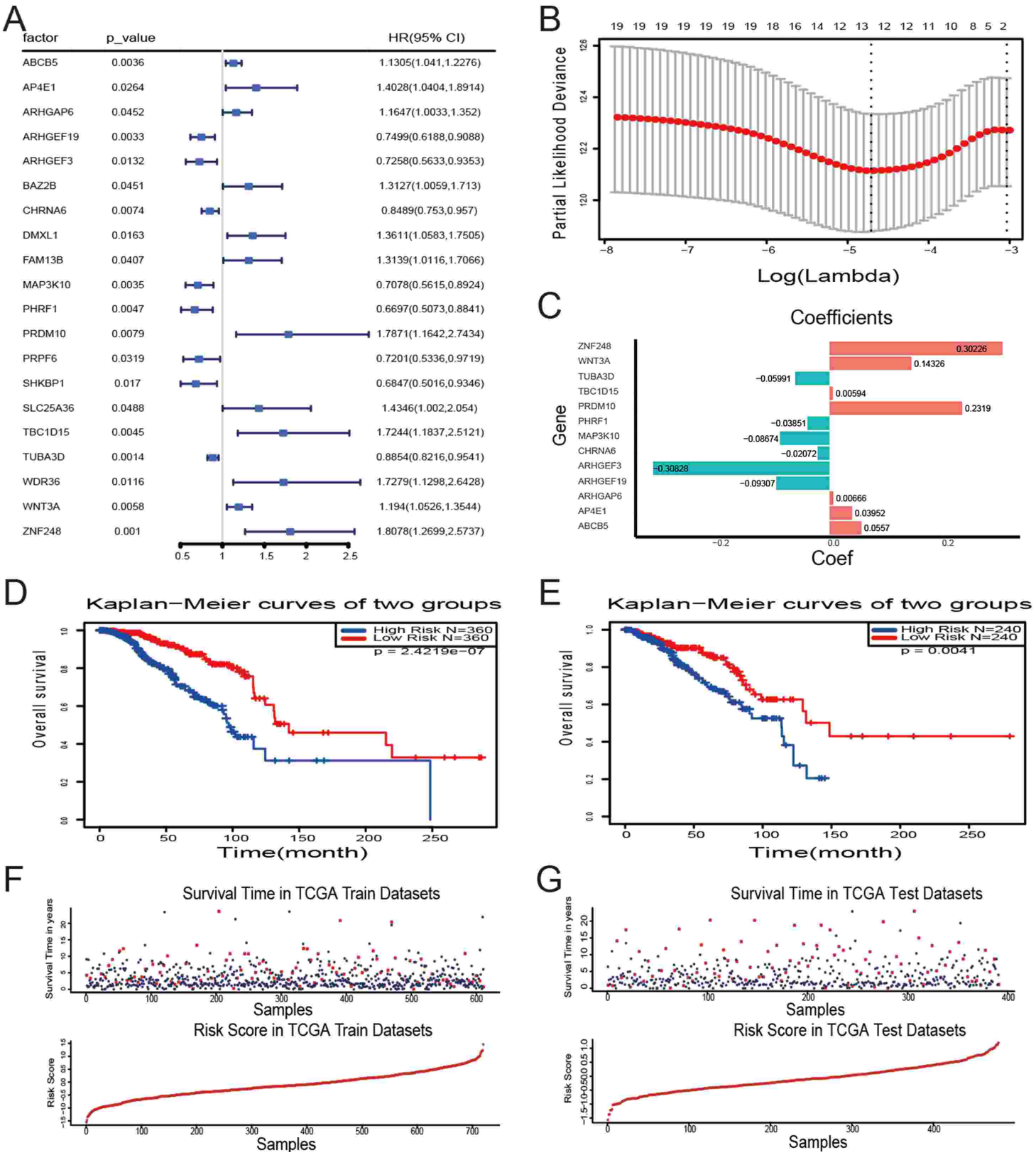

FIGURE 5. Prognostic signature validation and construction based on breast cancer cohorts from TCGA. (A) Univariate analysis identified 20 m6A-related genes that significantly correlated with overall survivals of breast cancer patients from TCGA. (B) A LASSO regression model of the 20 m6A-related genes. (C) The regression coefficients of 13 independent prognostic genes. (D and E) Kaplan-Meier survival curves for participants allocated to high- or low-risk groups based on risk scores in the training and validation cohorts. (F and G) The distributions of the overall survival and risk scores of the two cohorts. TCGA: The Cancer Genome Atlas.

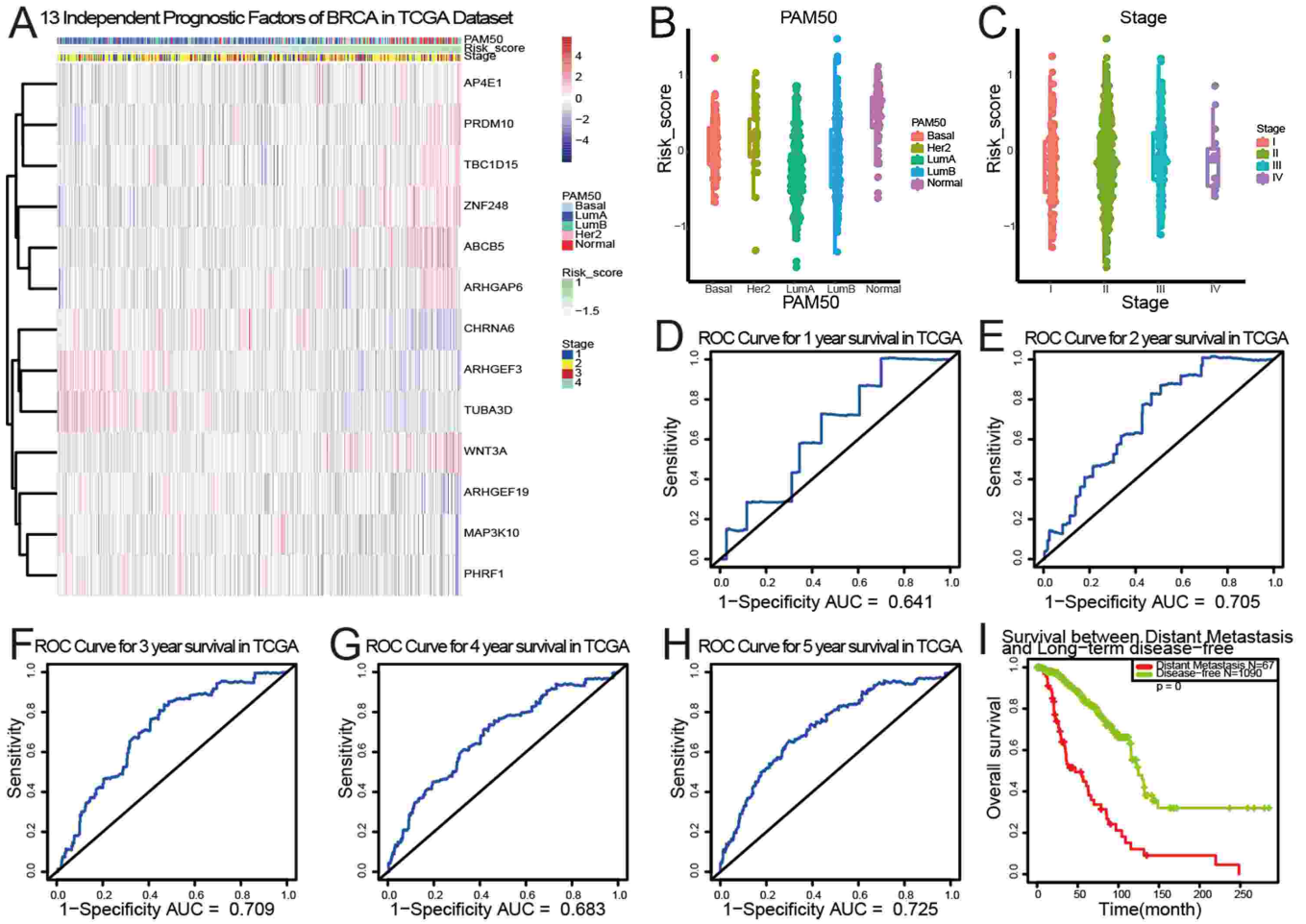

FIGURE 6. The prognostic signature-based risk scores can predict the clinical outcomes of breast cancer patients. (A) The heatmap reveals the 13 m6A-related genes expression levels within patients arranged according to risk scores. We compared risk scores distribution among the patients with different PAM50 (B) and pathologic stages (C). (D–H) ROC curves indicate the 13-gene risk signature's predictive efficiency for 1-, 2-, 3-, 4-, and 5-year survival rates. (I) Kaplan-Meier survival curves for patients with distant metastasis (n = 67) and long-term disease-free survivors (n = 1090). TCGA: The Cancer Genome Atlas; ROC: receiver operating characteristic; AUC: area under curves; PAM50: Prediction Analysis of Microarray 50.

constructed according to breast cancer training dataset from TCGA, and patients were categorized into high- and low-risk groups. This risk signature performed well in predicting overall survival in breast cancer patients and was verified further in an independent TCGA breast cancer cohort as well as an external dataset, suggesting that this prognostic model is reliable for prognosis prediction. Additionally, the risk signature demonstrated the predictive value of neoadjuvant treatment response in breast cancer in external cohort.

m6A regulators are aberrantly expressed in different cancer types, being encompassed in tumorigenesis and tumor progression. For instance, *WTAP* expression is significantly reduced in breast cancer in comparison with normal controls. Reduced *WTAP* expression is related to a poor prognosis in breast cancer patients [12]. In contrast, *WTAP* is overexpressed in hepatocellular carcinoma (HCC), serving as an independent predictor for poor prognosis in HCC patients [19]. In colon cancer, *YTHDC2* expression is positively correlated with the tumor stage [20]. Similarly, *YTHDF2* expression is upregulated in pancreatic cancer tissue compared with normal tissue and is greater in clinical patients with advanced cancer stages [21]. In our study, we found that *WTAP*, *YTHDC2*, and *YTHDF2* were significantly differentially expressed across different breast cancer stages, suggesting that these m6A regulators might be related to breast cancer progression and prognosis. Both YTHDC2 and YTHDF2 are m6A "reader" proteins of the YTH family, harboring a conserved hydrophobic binding pocket specific for m6A [22]. The PPI network analysis in this study provided detailed interaction among WTAP, YTHDC2, and YTHDF2, which has not been reported elsewhere and needs further investigation.

Based on m6A-related genes expression following consensus clustering, we identified four breast cancer subgroups and observed significant differences in gender, N stage, T stage, and PAM50 subtype among these subgroups. Furthermore, Kaplan—According to Meier survival analysis, overall survivals were significantly different across the subgroups in descending order of TM1, TM2, TM3, and TM4, suggesting that m6A-related genes levels are highly associated with poor prognosis of breast cancer. It was also demonstrated that the differentially expressed m6A-related genes in these four subtypes were related to certain biological processes and signaling

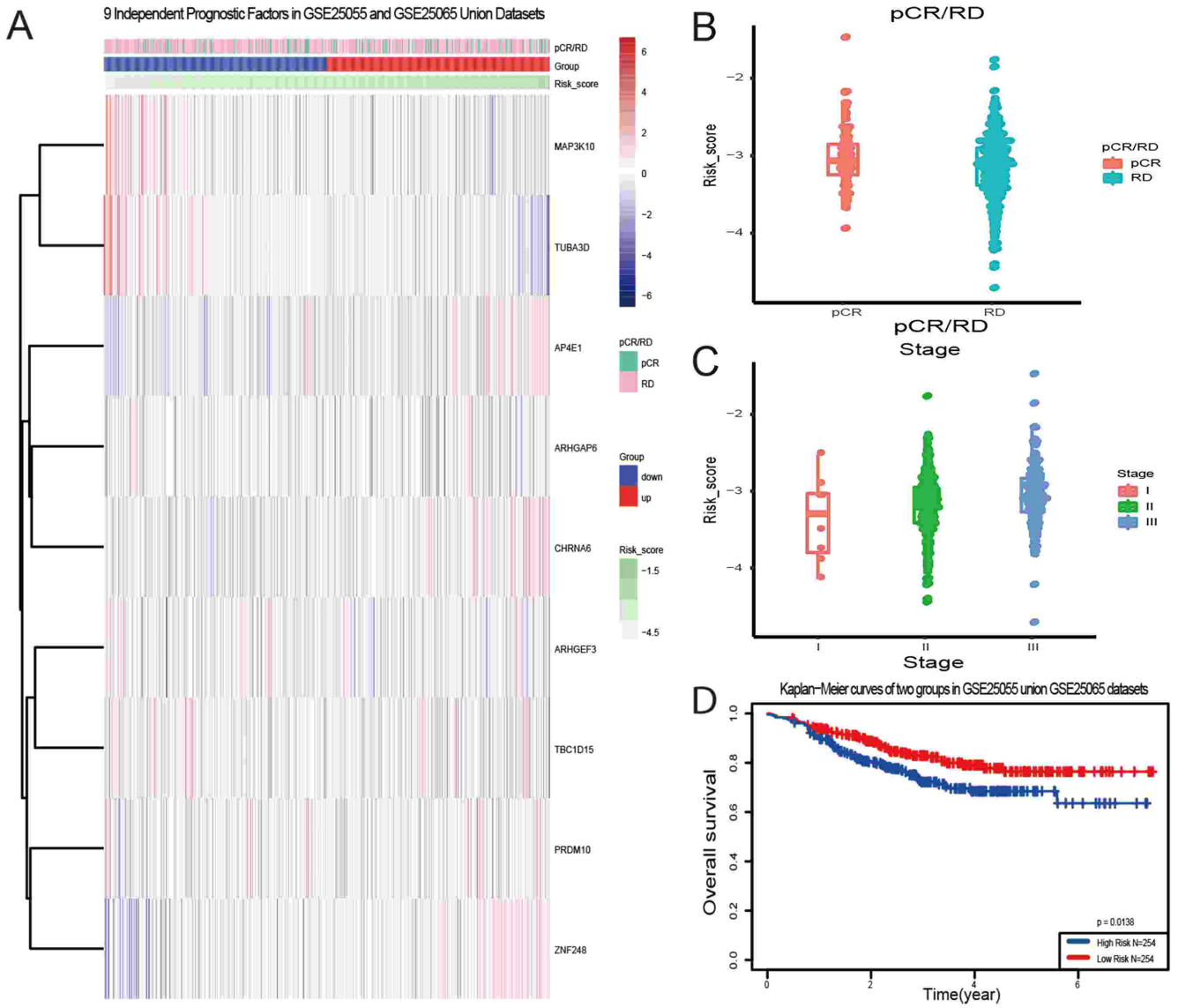

FIGURE 7. Prognostic signature validation in independent cohorts. (A) The heatmap reveals 9 of 13 m6A-related genes expression levels in patients classified according to risk scores. Risk scores distribution was compared among the patients with different response to neoadjuvant treatments (B) and pathologic stages (C). (D) Kaplan-Meier survival curves for individuals having high- or low-risk scores (n = 254/group). pCR: pathological complete response; RD: residual disease.

pathways. PI3K-Akt pathway is commonly activated in breast cancer [23], and its activation has been related to endocrine resistance and a poor prognosis in certain subgroups [23]. Coincidentally, the patients in TM4 are luminal type and had the poorest prognosis. Alterations in miRNA and calcium signaling pathways are linked to breast cancer diagnosis, metastasis, prognosis, and drug resistance [24, 25]. Our study indicated the potential biological process and signaling pathways linking m6A methylation and breast cancer development, representing a crucial step toward development of therapeutic strategies targeting m6A methylation for breast cancer treatment.

Besides the 16 candidates, m6A-related genes associated with breast cancer selected from articles, a 13-gene risk signature related to clinic-pathological characteristics, and clinical prognosis outcome of breast cancer patients were identified and validated in this study. Among these genes, *WNT3A* [26], *PRDM10* [27], *TUBA3D* [28], *PHRF1* [29], *ARHGAP6* [30], and *ABCB5* [31] implicated in breast cancer pathogenesis or prognosis, suggesting that our TCGA data-based analyses have predictive value.

5. Conclusions

This is the first report of a prognostic model of m6A-related genes for patients with breast cancer, which has a predictive value of the response to neoadjuvant treatment. The 13-gene-based risk score can predict prognosis and responsiveness to neoadjuvant treatment in breast cancer patients. These findings may provide important information for prognostic stratification and therapeutic strategies for breast cancer. Further research will be performed to research and verify the effects mechanism of risk genes on breast cancer development through m6A modification.

AUTHOR CONTRIBUTIONS

JS and HFM—conceptualized and designed the study, drafted the initial manuscript, and reviewed and revised the manuscript. YXC, CC, and ZQL—designed the data collection instruments, collected data, carried out the initial analyses, and reviewed and revised the manuscript. JGS—conceptualized and designed the study, coordinated and supervised data collection, and critically reviewed the manuscript for essential intellectual content. All authors endorsed the final manuscript and accepted to take responsibility for all work aspects.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

ACKNOWLEDGMENT

Not applicable.

FUNDING

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA: a Cancer Journal for Clinicians. 2020; 70: 7–30.

[2] Dai D, Jin H, Wang X. Nomogram for predicting survival in triple-negative breast cancer patients with histology of infiltrating duct carcinoma: a population-based study. American Journal of Cancer Research. 2018; 8: 1576–1585.

[3] Lee SK, Yang JH, Woo SY, Lee JE, Nam SJ. Nomogram for predicting invasion in patients with a preoperative diagnosis of ductal carcinoma in situ of the breast. The British Journal of Surgery. 2013;100:1756–1763.

[4] Zheng H, Luo L, Zhao W. Factors associated with level III lymph nodes positive and survival analysis of its dissection in patients with breast cancer. Laparoscopic, Endoscopic and Robotic Surgery. 2020; 3: 43–47.

[5] Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. The Lancet Oncology. 2015; 16: e173–e180.

[6] Cava C, Bertoli G, Castiglioni I. Integrating genetics and epigenetics in breast cancer: biological insights, experimental, computational methods and therapeutic potential. BMC Systems Biology. 2015; 9: 62.

[7] He L, Li H, Wu A, Peng Y, Shu G, Yin G. Functions of N6-methyladenosine and its role in cancer. Molecular Cancer. 2019;18:176.

[8] Ke S, Alemu EA, Mertens C, Gantman EC, Fak JJ, Mele A, *et al*. A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. Genes and Development. 2015; 29: 2037–2053.

[9] Dai D, Wang H, Zhu L, Jin H, Wang X. N6-methyladenosine links RNA metabolism to cancer progression. Cell Death & Disease. 2018; 9: 124.

[10] Batista PJ. The RNA modification N6-methyladenosine and its implications in human disease. Genomics, Proteomics & Bioinformatics. 2017; 15: 154–163.

[11] Maity A, Das B. N6-methyladenosine modification in mRNA: machinery, function and implications for health and diseases. Federation of European Biochemical Societies Journal. 2016; 283: 1607–1630.

[12] Wu L, Wu D, Ning J, Liu W, Zhang D. Changes of N6-methyladenosine modulators promote breast cancer progression. BMC Cancer. 2019; 19: 326.

[13] Cai X, Wang X, Cao C, Gao Y, Zhang S, Yang Z, *et al*. HBXIP-elevated methyltransferase METTL3 promotes the progression of breast cancer via inhibiting tumor suppressor let-7g. Cancer Letters. 2018; 415: 11–19.

[14] Liu L, Liu X, Dong Z, Li J, Yu Y, Chen X, *et al*. N6-methyladenosine-related genomic targets are altered in breast cancer tissue and associated with poor survival. Journal of Cancer. 2019; 10: 5447–5459.

[15] Zhang C, Zhi WI, Lu H, Samanta D, Chen I, Gabrielson E, *et al*. Hypoxia-inducible factors regulate pluripotency factor expression by ZNF217- and ALKBH5-mediated modulation of RNA methylation in breast cancer cells. Oncotarget. 2016; 7: 64527–64542.

[16] Lan Q, Liu PY, Haase J, Bell JL, Hüttelmaier S, Liu T. The critical role of RNA m6A methylation in cancer. Cancer Research. 2019; 79: 1285–1292.

[17] Lence T, Paolantoni C, Worpenberg L, Roignant J. Mechanistic insights into m6A RNA enzymes. Gene regulatory mechanisms BBA. Gene regulatory mechanisms. 2019; 1862: 222–229.

[18] Yang Y, Hsu PJ, Chen Y, Yang Y. Dynamic transcriptomic m6A decoration: writers, erasers, readers and functions in RNA metabolism. Cell Research. 2018; 28: 616–624.

[19] Chen Y, Peng C, Chen J, Chen D, Yang B, He B, *et al*. *WTAP* facilitates progression of hepatocellular carcinoma via m6A-HuR-dependent epigenetic silencing of ETS1. Molecular Cancer. 2019; 18: 127.

[20] Tanabe A, Tanikawa K, Tsunetomi M, Takai K, Ikeda H, Konno J, *et al*. RNA helicase *YTHDC2* promotes cancer metastasis via the enhancement of the efficiency by which HIF-1α mRNA is translated. Cancer Letters. 2016; 376: 34–42.

[21] Chen J, Sun Y, Xu X, Wang D, He J, Zhou H, *et al*. *YTH* domain family 2 orchestrates epithelial-mesenchymal transition/proliferation dichotomy in pancreatic cancer cells. Cell Cycle. 2017; 16: 2259–2271.

[22] Luo S, Tong L. Molecular basis for the recognition of methylated adenines in RNA by the eukaryotic *YTH* domain. Proceedings of the National Academy of Sciences. 2014; 111: 13834–13839.

[23] Verret B, Cortes J, Bachelot T, Andre F, Arnedos M. Efficacy of PI3K inhibitors in advanced breast cancer. Annals of Oncology. 2019; 30: x12–x20.

[24] Teoh SL, Das S. The role of microRNAs in diagnosis, prognosis, metastasis and resistant cases in breast cancer. Current Pharmaceutical Design. 2017; 23: 1845–1859.

[25] Makena MR, Rao R. Subtype specific targeting of calcium signaling in breast cancer. Cell Calcium. 2020; 85: 102109.

[26] He S, Lu Y, Liu X, Huang X, Keller ET, Qian C, *et al*. Wnt3a: functions and implications in cancer. Chinese Journal of Cancer. 2015; 34: 50.

[27] Sorrentino A, Federico A, Rienzo M, Gazzerro P, Bifulco M, Ciccodicola A, *et al*. PR/SET domain family and cancer: novel insights from the cancer genome atlas. International Journal of Molecular Sciences. 2018; 19: 3250.

[28] Liu L, Chen Z, Shi W, Liu H, Pang W. Breast cancer survival prediction using seven prognostic biomarker genes. Oncology Letters. 2019; 18: 2907–2916.

[29] Ettahar A, Ferrigno O, Zhang M, Ohnishi M, Ferrand N, Prunier C, *et al*. Identification of *PHRF1* as a tumor suppressor that promotes the TGF-β cytostatic program through selective release of TGIF-driven PML inactivation. Cell Reports. 2013; 4: 530–541.

[30] Chen WX, Lou M, Cheng L, Qian Q, Xu LY, Sun L, *et al*. Bioinformatics analysis of potential therapeutic targets among *ARHGAP* genes in breast cancer. Oncology Letters. 2019; 18: 6017–6025.

[31] Yao J, Yao X, Tian T, Fu X, Wang W, Li S, *et al*. ABCB5-ZEB1 axis promotes invasion and metastasis in breast cancer cells. Oncology Research. 2017; 25: 305–316.

How to cite this article: Jun Shen, Hongfang Ma, Yongxia Chen, Cong Chen, Zhaoqing Li, Jianguo Shen. N6-methyladenosine-related genes contribute to malignant progression, have clinical prognostic and neoadjuvant treatments response impact for breast cancer. European Journal of Gynaecological Oncology. 2023; 44(2): 67-78. doi: 10.22514/ejgo.2023.024.