Analysis of m7G-Related Signatures in the Tumor Immune Microenvironment and Identification of Clinical Prognostic Regulators in breast cancer

Qinghua Huang  
Guangxi Medical University Cancer Hospital

Jianlan Mo  
Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region

Huawei Yang  
Guangxi Medical University Cancer Hospital

Yinan Ji  
Guangxi Medical University Cancer Hospital

Rong Huang  
Guangxi Medical University Cancer Hospital

Yan Liu  
Guangxi Medical University Cancer Hospital

You Pan (✉ panyou126@126.com)  
Guangxi Medical University Cancer Hospital

Research Article

Keywords: N7-methylguanosine, Breast cancer, Immune microenvironment, Prognostic signature, qRT-PCR

Posted Date: December 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2290724/v1

License: ☺️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Breast cancer is a malignant tumor that seriously threatens the life and health of women and exhibits high inter-individual heterogeneity, emphasizing the need for more in-depth research on its pathogenesis. It is well-established that internal 7-Methylguanosine (m7G) modifications affect RNA processing and function and are thought to be involved in human diseases. However, little is currently known about the role of m7G modification in breast cancer. In this study, we elucidated the expression, copy number variation incidence, and prognostic value of the 24 m7G-related genes (m7GRGs) in breast cancer. Subsequently, based on the expression of 24 m7GRGs, consensus clustering divided tumor samples from the TCGA-BRCA dataset into 4 subtypes with significantly different immune cell infiltration and stromal scores. Differentially expressed genes between subtypes were mainly enriched in immune-related pathways such as Ribosome, TNF signaling pathway, and Salmonella infection. Support vector machines and multivariate cox regression analysis were performed based on these 24 m7GRGs, and 4 m7GRGs (AGO2, EIF4E3, DPCS, and EIF4E) were identified for constructing the prediction model. Next, we constructed a nomogram model based on the risk model and clinical factors. The ROC curve indicated that the nomogram model had a strong ability to predict the prognosis of breast cancer. The prognosis of patients in the high- and low-TMB groups was significantly different (p = 0.03). Moreover, the 4-gene signature could predict response to chemotherapy. Therefore, this study found that the m7GRGs signature in the immune microenvironment of breast cancer affects the prognosis of breast cancer patients. Our study provides a reference for applying m7GRGs signature to the personalized treatment of breast cancer patients and exploring the mechanism of breast cancer progression.

Introduction

Breast cancer is a malignant tumor that seriously threatens the life and health of women. It is estimated that there are about 1.3 million new breast cancer cases worldwide each year, and 450,000 patients die of breast cancer[1]. The etiology of breast cancer remains unclear and can be roughly divided into genetic, environmental, and hormonal factors, among which genetic factors occupy a very important part[2–4]. Epidemiological surveys have found that 5–10% of breast cancers are familial and are related to genetic mutations inherited from the mother or father[5]. At present, the treatment of breast cancer is mainly based on clinical staging, pathological typing, and molecular typing. It is well-established that breast cancer is a highly heterogeneous disease[6, 7]. The traditional clinicopathological diagnostic approach to breast cancer has been associated with low accuracy and poor specificity. Therefore, more in-depth studies on the molecular mechanisms in the progression of breast cancer are needed.

Current evidence suggests that the tumor immune microenvironment (TIME) plays an important role in the occurrence, development, invasion and metastasis. The past decade has witnessed significant inroads achieved in better understanding the TIME of breast cancer people have a deeper understanding of the relationship between breast cancer and TIME. The tumor immune microenvironment includes immune cells in various tumor microenvironments (such as T/B lymphocytes, macrophages, dendritic cells, giant cells), stromal cells, tumor cells and their corresponding expressed and secreted immune
molecules[8, 9]. Their interaction has a dual effect on breast cancer: immune effector cells (such as CD8 + CTL, CD4 + Th1, NK, and NKT cells) and molecules (such as IFN-γ, TNF-α, IL-2) inhibit tumor cell growth and proliferation through different pathways, while immunosuppressive cells (such as regulatory T cells, tumor-associated macrophages M2, myeloid-derived suppressor cells), factors (such as IL-10, TGF-β, etc.) and inhibitors expressed or secreted by tumor cells inhibit immune responses through different pathways and participate in tumor escape and promote tumor occurrence, development and metastasis[10]. Tumor-infiltrating lymphocytes (TILs) mainly include T cells, B cells and NK cells infiltrating tumor tissue. Interestingly, high levels of CD3 + TILs are associated with HER2-negative, PR-negative, and high-grade tumor grades; high CD3 expression is associated with better prognosis and is predictive of chemosensitivity in ER-negative (ER-) cancer, including triple-negative breast cancer (TNBC) [11]. A study showed that CD4 + Th2 cells promote the occurrence and lung metastasis of advanced breast cancer by regulating the biological activities of macrophages, suggesting that CD4 + Th2 is associated with poor prognosis in breast cancer[12]. Myeloid-derived suppressor cells (MDSCs) are a group of naive myeloid cells that exhibit an inhibitory effect on immune cells (such as T/DC/NK cells). Current evidence suggests that the number of MDSCs increases with cancer burden in mouse models or in vitro experiments[13]. An increasing body of literature suggests that MDSCs are associated with poor breast cancer prognoses [14, 15]. The above findings indicate that the immune status of the tumor microenvironment has an important impact on the prognosis of breast cancer patients.

Epigenetics is a phenomenon that does not involve altering the gene sequence but heritable changes in gene expression and modification level. Among them, epigenetic methylation by 7-Methylguanosine (m7G) has attracted widespread attention. During transcription initiation, m7G is co-transcribed onto the 5' cap [16]. This cap modification stabilizes transcripts, prevents exonucleolytic degradation, and regulates nearly every stage of the mRNA life cycle, including transcription elongation, pre-mRNA splicing, polyadenylation, nuclear export, and translation. In addition to being part of the cap structure, m7G is also present inside tRNA and rRNA[17], while internal m7G modifications affect RNA processing and function and are thought to be involved in human diseases. Studies have shown that epigenetic silencing affects almost all antigen processing and presentation processes [18]. The important role of epigenetics in tumor immune escape has laid a solid theoretical foundation for using epigenetic modifiers to improve the immune targeting of tumor cells. However, it remains unclear whether m7G methylation modification affects tumor immunity.

To investigate the function of m7G modification in breast cancer, we explored the expression, copy number variation incidence and prognostic value of the 24 m7GRGs in breast cancer. Then, we performed consensus clustering and divided tumor samples into 4 subtypes with significantly different immune cell infiltration and stromal scores based on the expression of 24 m7G-related genes in the TCGA-BRCA dataset. Differentially expressed genes (DEGs) between subtypes were mainly enriched in immune-related pathways such as Ribosome, TNF signaling pathway and Salmonella infection. Multivariate cox regression analysis was performed based on these 21 prognostic-related m7G genes, and 4 genes (AGO2, EIF4E3, DPCS and EIF4E) for constructing the prediction model were identified. The samples were divided into high and low-risk groups according to the median risk score, and survival analysis showed that the
survival difference between the high and low-risk groups was significant. The 4 m7G-related genes could independently predict the prognosis of BRCA patients. Based on the risk scores of the prognostic model and clinical factors, we constructed a nomogram model to predict the prognosis of breast cancer patients. Finally, we explored the association between m7GRGs-related prognostic models and TMB and drug sensitivity. Therefore, this study reports a hitherto undocumented m7G gene signature that can assist during clinical practice in improving the prognosis of breast cancer patients.

Materials And Methods

Datasets

Gene expression profiling data of breast cancer patients were obtained from 2 independent patient cohorts, including The Cancer Genome Atlas (TCGA) dataset TCGA-BRCA and the Gene Expression Omnibus (GEO) dataset GSE1456. The TCGA-BRCA cohort included 113 normal samples and 1113 breast cancer samples. Somatic data for breast cancer were also obtained from TCGA. TCGA-BRCA data were annotated with gene names using GENCODE22 annotation files, and TCGA-BRCA patient survival data and clinical data were obtained from UCSC Xena, including survival time, survival status, age, tumor stage, and gender. CNV data for breast cancer were also obtained from UCSC Xena. In our study, patient samples with clinical information and survival time greater than 30 days were retained. Finally, 1023 patients samples were included in TCGA-BRCA, and 159 patients samples were included in GSE1456, among which TCGA-BRCA was used for model construction, and GSE1456 was used for model validation.

Mutation analysis and prognostic analysis of m7G-related genes

M7G-related genes (m7GRGs) (DCP2, IFIT5, EIF3D, EIF4G3, NSUN2, GEMIN5, AGO2, NUDT10, EIF4E, EIF4E2, NCBP2, NUDT11, NUDT3, NCBP1, METTL1, LARP1, NUDT4, EIF4E3, SNUPN, WDR4, LSM1, NUDT16, DCPS, and CYFIP1) were obtained from the existing literature [19], and related gene sets GOMF_M7G_5_PPPN_DIPHOSPHATASE_ACTIVITY, GOMF_RNA_CAP_BINDING, and OMF_RNA_7_METHYLGUANOSINE_CAP_BINDING. We used “limma” packages to explore the expression of 24 m7GRGs in breast cancer and normal tissues. Next, we explored the CNV incidence of 24 m7GRGs and mapped their altered locations on 23 chromosomes using the “RCircos” package. To elucidate the correlation of 24 m7GRGs with prognosis of breast cancer patients, we used the "igraph", "psych", "reshape2" and "RColorBrewer" packages to draw the prognosis-related network of 24 m7GRGs.

Consensus clustering of m7GRGs and functional enrichment analysis

Based on the expression profile data of these 24 genes, the K-means clustering algorithm of the “ConsensusClusterPlus” package was used to perform consensus clustering on TCGA-BRCA patients to
obtain 4 subtypes. The clustering was repeated 1000 times to ensure the accuracy and stability of the results.

We used the “limma” package to identify differentially expressed genes (DEGs) across the 4 subtypes, with thresholds set at \( p < 0.05 \) and \(|\log\text{Fold Change}| > 1\). To identify the enriched pathways between different subtypes, we conducted GO and KEGG enrichment analyses for DEGs between subtypes and visualized the top 5 pathways with the most significant enrichment results. Gene set enrichment analysis was performed (c2.cp.kegg.v7.5.1.symbols.gmt of the reference gene set MSigDB database), and the top 10 pathways were visualized. The above enrichment analysis was conducted using the “clusterProfiler” package.

**Comparison of immune cell infiltration among m7G patterns**

To explore the extent of immune cell infiltration among different subtypes, we used the “IOBR” package to assess immune cell infiltration with the ESTIMATE algorithm and the CIBERSORT algorithm on the TCGA-BRCA dataset and obtained the immune cell infiltration of each sample in both algorithms. The indicators evaluated by the ESTIMATE algorithm include immune score, stromal score and tumor purity. The CIBERSORT algorithm enables quantification of the relative abundance of 22 types of immune cells. Differences in immune cell infiltration between different subtypes were considered significantly different, with \( p \)-values less than 0.05.

**Support vector machines and multivariate Cox regression analysis**

We used the support vector machine (SVM) algorithm of the “e1071” package to select the prognosis-related features on the TCGA-BRCA expression profiling data of 24 m7G-related genes. 21 out of the 24 prognosis-related genes were screened out and further used for survival analysis. Multivariate cox analysis was used to construct predictive models, and survival analysis was performed using the “survival” package. Subsequently, we validated the model in the GSE1456 dataset, and a \( p \)-value less than 0.05 was statistically significant. Kaplan-Meier (KM) analysis were used to evaluate the the difference in survival between high- and low-risk groups. In addition, we integrated the risk scores and clinical features for univariate and multivariate cox analyses to verify that risk scores are predictive markers independent of other clinical features and constructed nomograms based on the clinical features associated with prognosis using the rms package to more accurately predict the prognosis of patients. The results of nomogram prediction were verified by calibration curve and ROC curve analysis to ensure the accuracy of the nomogram for predicting the 1-year, 3-year and 5-year survival rates of patients.

**Tumor mutation burden analysis and drug sensitivity prediction**
The "maftools" package was used to draw waterfall plots related to tumor mutational burden (TMB). KM analysis was used to evaluate the effect of TMB on the prognosis of breast cancer patients. Moreover, we used t-tests to determine differences in TMB levels between high- and low-risk groups. Spearman was used to calculate the correlation between TMB and risk score. Then, we estimated each patient's sensitivity to chemotherapeutic drugs using the Genomics of Cancer Drug Sensitivity (GDSC) database. The IC50 values for high and low-risk groups were quantified by the "pRRophetic" package. The CellMiner database was used to mine sensitive drugs to build a predictive model. These drugs can affect the changes in the expression value of the corresponding genes, thereby affecting the patients' risk score. Therefore, it is necessary to build a predictive model based on genes associated with drug sensitivity. The overall flow chart of this study was shown in Fig. 1.

Cell culture and qRT-PCR

MDA-MB-231, MDA-MB-468, SKBR3 cells and MCF-10A were purchased from Procell Life Science&Technology Co.,Ltd. (Wuhan, China). MDA-MB-231 was incubated in the DMEM culture medium (Gibco, 11965092) with 10% fetal bovine serum. MDA-MB-468 was incubated in the RPMI 1640 culture medium (Gibco, 11875093) with 10% fetal bovine serum. SKBR3 was cultured in special culture medium (Procell, CM-0211) containing McCoy's 5A, 10% FBS and 1% P/S. MCF-10A was cultured in special culture medium (Procell, CM-0525) containing DMEM/F12, 5% HS, 20ng/mL EGF, 0.5µg/mL Hydrocortisone, 10µg/mL Insulin, 1% NEAA and 1% P/S. Real-time qPCR was performed by PowerUp SYBR Green Master Mix (Thermo Fisher, A25742) after RNA extraction and reversed transcription from all these four cell lines. The primers sequences are listed in Supplementary Table S1.

Results

Landscape of m7GRGs expression, genetic variation, and prognostic relevance in breast cancer

Firstly, we explored the expression of 24 m7GRGs in breast cancer tissues and normal tissues in the TCGA-BRCA cohort. The expression of 19 m7GRGs was differentially expressed in breast cancer tissues and normal tissues. The expressions of NSUN2, EIF4E, EIF4E2, NCBP2, NUDT3, NCBP1, LARP1, WDR4, and LSM1 were up-regulated (Fig. 2A). Subsequently, we identified the CNV incidence of 24 m7GRGs. In the TCGA-BRCA cohort, CNV alterations were found in all 24 m7GRGs. And more than half of m7GRGs have copy number amplification, while CYFIP1, NUDT4, DCPS, NCBP1, EIF3D, DCP2, EIF4E3, IFIT5, and EIF4G3 have CNV deletions (Fig. 2B). Figure 2C shows the location of CNV alterations of 24 m7GRGs on chromosomes. Finally, to explore the association of 24m7GRGs with the prognosis of breast cancer, we mapped a network of interactions among 24 m7GRGs and the effect of their expression on prognosis in breast cancer patients. The results showed that 17 m7GRGs were risk factors and significantly impacted breast cancer prognosis, while NUDT10, NUDT3, EIF4E3, SNUPN, NUDT16, IFIT5, and EIF3D were favorable factors for the prognosis of breast cancer patients (Fig. 2D).
Identification of breast cancer classification patterns mediated by 24 m7GRGs

Correlation analysis of 24 m7G-related genes in the TCGA-BRCA dataset showed a close correlation between these genes (Fig. 3A) (p < 0.05). Based on the expression of 24 m7GRGs, we used consensus clustering analysis to classify 1023 breast cancer samples into 4 subtypes (Fig. 3B-D). Principal component analysis (PCA) results showed that the 4 subtypes were clearly separated (Fig. 3E), indicating there were significant differences in the expression profiles among different subtypes. At the same time, the expression levels of 24 genes among the 4 subtypes were significantly different (Fig. 3F), indicating that our typing results have good stability and accuracy. To explore differences in immune infiltration levels among the 4 breast subtypes, we performed immune infiltration analysis on the 4 breast cancer subtypes. The CIBERSORT results showed that the infiltration degree of 22 types of immune cells was different between different samples (Fig. 4A), and a significant correlation was present between T cells and natural killer cells (Fig. 4B), indicating a synergistic effect between these cells in BRCA patients. Moreover, the expression of most immune cells differed significantly among different subtypes. Specifically, Plasma_cells, T_cells_CD8, T_cells_follicular_helper, T_cells_regulatory_(Tregs), NK_cells_resting, NK_cells_activated, Macrophages_M0, Macrophages_M1, Dendritic_cells_resting, Dendritic_cells_activated, Mast_cells_resting, Mast_cells_activated, and Neutrophils had significant differences in immune infiltration abundance among the 4 breast cancer subtypes (Fig. 4C). The ESTIMATE results showed that the stromal score was significantly different among different subtypes (p < 0.05) (Fig. 5A-D), indicating a significant difference in stromal cell composition among subtypes.

Functional enrichment analysis among breast cancer subtypes

To identify the biological functions and pathways involved in the differential genes among the four breast cancer subtypes in breast cancer, we performed enrichment analysis of DEGs between different breast cancer subtypes (Subtype 1 vs. Subtype 2, Subtype 1 vs. Subtype 3, Subtype 1 vs. Subtype 4, Subtype 2 vs. Subtype 3, Subtype 2 vs. Subtype 4, Subtype 3 vs. Subtype 4). GO enrichment analysis (Fig. 6A-F) showed significant enrichment in ribosome biogenesis, RNA splicing and macrophage pathways. KEGG enrichment analysis results (Fig. 7A-F) showed that the DEGs were mainly enriched in Ribosome, TNF signaling pathway and Salmonella infection pathways. The GSEA results (Fig. 8A-F) showed significant enrichment in KEGG_RIBOSOME, KEGG_VIBRIO_CHOLERAE_INFECTION and KEGG_PROTEASOME and other pathways. These results suggest that m7g-related genes may be involved in numerous biological processes and pathways related to immunity.

Construction of a prognostic model related to m7GRGs

To screen for m7GRGs associated with breast cancer patient survival time, we performed SVM analysis on 24 m7GRGs. 21 m7G-related genes associated with prognosis were obtained using the SVM (Fig. 9A-B) and used to perform a multivariate cox regression analysis. Four genes (AGO2, EIF4E3, DPCS and
EIF4E) were identified and used to construct a prediction model (Fig. 9C), which could provide the risk score of each sample in TCGA-BRCA. We found that the scores of the high- and low-risk groups in the TCGA cohort were positively correlated with the risk degree of the sample (Fig. 9D). Finally, the samples were divided into high and low-risk groups according to the median risk score. The survival analysis showed a significant difference between the high and low-risk groups (Fig. 9E). Similarly, we found that samples with higher risk values in the TCGA cohort had a worse prognosis (Fig. 9F). The difference in survival between high- and low-risk group samples in the validation set GSE1456 was also significant (Fig. 9G). The expression of the four genes used to construct the predictive model was also significantly different in the high- and low-risk groups (Fig. 9H). Single gene survival analysis was performed for the four genes AGO2 (Fig. 10A, 10B), EIF4E3 (Fig. 10C, 10D), DPCS (Fig. 10E, 10F) and EIF4E (Fig. 10G, 10H) in TCGA cohort and GEO cohort, respectively. The results showed significant differences in survival between the high and low-expression groups, further indicating that these four genes were significantly correlated with the prognosis of BRCA.

**Assessment of 4 m7G-related genes as Independent BRCA Prognostic Factors**

To assess the prognostic value of the risk score, we performed a prognostic analysis of risk scores and other clinical characteristics in the TCGA-BRCA cohort, and univariate cox regression analysis (Fig. 11A) showed that age, tumor stage, and risk score were significantly associated with prognosis (p < 0.05). Further multivariate cox regression analysis confirmed (Fig. 11B) that age, tumor stage, and risk score were independent predictors of prognosis (p < 0.05). The calibration curve results showed that the predicted values at 1, 3 and 5 years slightly deviated from the diagonal line (Fig. 11C), indicating that the nomogram can predict the prognosis of breast cancer patients well compared to the ideal model. We constructed a nomogram (Fig. 11D) based on these three clinical characteristics to predict the 1-, 3-, and 5-year survival rates of patients. Calibration and ROC curves were used to validate the accuracy of the nomogram in predicting survival time in breast cancer patients. The ROC curve showed that the AUC values at 1 year (Fig. 11E, AUC = 0.778), 3 years (Fig. 11F, AUC = 0.678) and 5 years (Fig. 11G, AUC = 0.631) were higher. Overall, both methods demonstrated that the nomogram has better accuracy in predicting patient prognosis.

**TMB and drug-sensitivity analysis of m7GRGs**

Since genetic mutations are an essential cause of BRCA development, we explored differences in the distribution of somatic mutations between high- and low-risk populations. The top 20 most frequently mutated genes for these two groups are shown in Fig. 12A and Fig. 12B, respectively. There was no significant difference in the mutation frequency of the top 20 genes between the high-risk and low-risk groups. The expression levels of risk scores were statistically different between the low TMB group and the high TMB group (Fig. 12C). The correlation between the mutational burden of BRCA and the population risk score was weak but statistically significant (Fig. 12D). The Kaplan-Meier curve of OS indicated that the OS of the patients in the high TMB group was significantly lower than that of the patients in the low TMB group (p = 0.03) (Fig. 12E). In addition, the OS of patients with high/low
mutational burden in the high-risk group was statistically different from that in the low-risk group with high/low mutational burden (p = 0.03) (Fig. 12F).

We used the pRRophetic package to compare the differences in the estimated half-maximal inhibitory concentration (IC50) levels of six chemotherapeutic agents, including erlotinib (Fig. 13A), gemcitabine (Fig. 13B), cytarabine (Fig. 13C), gefitinib (Fig. 13D), Akt1/2/3 inhibitor MK.2206 (Fig. 13E), PPM1D (WIP1) inhibitor CCT007093 (Fig. 13F). Our data showed that the high-risk score group was more sensitive to gemcitabine and cytarabine than the low-risk group. In contrast, the low-risk score group was more sensitive to gefitinib and CCT007093. Drug sensitivity analysis of the four-gene signature showed that the DCPS, EIF4E, EIF4E3 and AGO2 genes were significantly associated with Dasatinib, Chelerythrine, E7820, and Imexon, respectively (Fig. 14A-P).

**Expression of the Signature m7GRGs**

Breast cancer cell lines (MDA-MB-231, MDA-MB-468 and SKBR3) and normal breast cell lines (MCF-10A) were used to validate the expression levels of the signature m7GRGs (AGO2, EIF4E3, DPCS and EIF4E). The results showed that the expressions of AGO2 (Fig. 15A) and EIF4E3 (Fig. 15B) in breast cancer cell lines were significantly lower than those in normal breast cell lines, while the expressions of DCPS (Fig. 15C) and EIF4E (Fig. 15D) in breast cancer cell lines were significantly higher than those in normal breast cell lines. The above results were consistent with the results of the bioinformatics analysis. Therefore, further exploration of the exact mechanism of these 4 m7Gs in breast cancer is of great significance for improving the prognosis and treatment of breast cancer patients.

**Discussion**

Breast cancer is a solid tumor with high tumor heterogeneity. According to the latest survey results released in 2020, 2.26 million people were newly diagnosed with breast cancer worldwide, surpassing lung cancer to become the cancer with the highest incidence globally [20]. In recent years, much emphasis has been placed on achieving early diagnosis and treatment of breast cancer. The tumor immune microenvironment is an emerging research field, with an ample number of studies reporting that the immune microenvironment of breast cancer is closely related to the patient's treatment response, tumor development and recurrence. It has been found that both the body's innate and acquired immunity play an important role in breast cancer[21]. For example, regarding innate immunity, an in vitro experiment in breast cancer mice found that the immunosuppression of neutrophils can be achieved by inhibiting T cell proliferation[22]. Moreover, tumor-associated macrophages related to innate immunity can be divided into classically activated M1 and selectively activated M2 types. M1 type can secrete TNFα and nitric oxide synthase 2 (NOS2) and play an anti-tumor immune response[23], while the M2 type can secrete IL-4, IL-10, etc., and finally form a microenvironment that promotes tumor growth and metastasis [24]. For acquired immunity, relevant studies have shown that T cells are the main tumor-infiltrating lymphocytes in breast cancer, and B cells play an important role [25]. Moreover, these tumor-infiltrating lymphocytes can be used to evaluate the prognosis of patients with triple-negative breast
cancer. In recent years, it has been established that tumor-infiltrating lymphocytes are associated with a better prognosis[26].

To investigate the function of m7G modification in breast cancer, we first identified the expression, CNV incidence and prognostic value of the 24 m7GRGs in breast cancer. We found that the expression of NSUN2, EIF4E, EIF4E2, NCBP2, NUDT3, NCBP1, LARP1, WDR4, and LSM1 was increased in breast cancer tissues. Moreover, the results of prognostic analysis showed that more than half of m7GRGs were risk factors for the prognosis of breast cancer patients. Therefore, it is necessary to further explore the prognostic impact of m7GRGs in breast cancer. Then, we performed consensus clustering and divided tumor samples into 4 subtypes with significantly different immune cell infiltration and stromal scores between subtypes based on the expression of 24 m7G-related genes in the TCGA-BRCA dataset. DEGs between subtypes were mainly enriched in immune-related pathways such as Ribosome, TNF signaling pathway and Salmonella infection. Subsequently, based on the 24 m7GRGs, we screened 21 prognosis-related m7GRGs using the SVM algorithm. Multivariate cox regression analysis was performed based on these 21 prognostic-related m7G genes, and 4 genes (AGO2, EIF4E3, DPCS and EIF4E) for constructing the prediction model were identified.

The m7g-related genes AGO2, EIF4E3, DPCS and EIF4E have been extensively studied over the years, and some of them have been associated with tumor progression. AGO2 is the only catalytically active member of the Argonaute family and is involved in small RNA-guided post-transcriptional gene silencing (including mRNA degradation and translational repression)[27]. In addition, AGO2 plays multiple roles in nuclear gene regulation, such as chromatin remodeling, double-strand break repair, alternative splicing transcriptional repression and activation[28]. It has been reported that AGO2 may play a role in double-strand break repair in tumor cells. eIF4E is a eukaryotic translation initiation factor and an oncogene with elevated expression in approximately 30% of human cancers [29, 30]. Its elevation in mouse models is associated with tumorigenesis, and tissue culture experiments have shown that the expression of eIF4E is associated with oncogenic transformation. eIF4E functions in mRNA export and translation of specific transcripts by binding to the methyl 7-guanosine cap found at the 5' end of mRNAs. These transcripts often encode proteins involved in proliferation, survival, invasion and metastasis. Unlike eIF4E1, eIF4E3 functions as a tissue-specific tumor suppressor[31]. In this respect, it has been shown that eIF4E3 inhibits the expression of both the mRNA export and translation targets of eIF4E1. The protein encoded by the DPCS gene is an mRNA decapping enzyme scavenger and is thought to be a key gene for AML cell survival. Mass spectrometry analysis revealed that dcp5 enzymes interact with components of the pre-mRNA metabolic pathway, including the spliceosome, to function [32].

In the present study, the samples were divided into high and low-risk groups according to the median risk score, and survival analysis showed that the survival difference between both groups was significant. We found that the 4 m7G-related genes were independent BRCA prognostic factors. Our study found that the high-risk group had higher TMB levels than the low-risk group and that patients with high TMB levels had higher survival rates than those with low TMB levels. The above results suggest that high TMB may be associated with a poorer prognosis in high-risk patients. Moreover, the 4-gene signature could predict the
response to chemotherapy. In addition, PCR results showed that the expression of 4 m7Gs was significantly different between breast cancer cell lines and normal breast cells (p < 0.05). Overall, we provided compelling evidence that the m7G gene signature was associated with the prognosis of breast cancer patients. Our work provides an M7G-related prognostic model for breast cancer patients that may be useful clinically.

Taken together, the present study found that the m7G gene signature in the immune microenvironment of breast cancer affects the prognosis of breast cancer patients. Importantly, our study provides the theoretical basis for applying the m7G gene signature for personalized treatment of breast cancer patients and novel insights into breast cancer progression mechanism and treatment.

Declarations

ETHICAL APPROVAL AND CONSENT

Not applicable.

Consent for publication

All the authors have consent of publication.

AVAILABILITY OF DATA AND MATERIALS


COMPETING INTERESTS

The authors declare no potential conflicts of interest.

FUNDING

The work was supported by National Natural Science Foundation of China grants (82002779, 81760478,81860464,81960477); the Seedling Plan of Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region (GXWCH-YMJH-2018005); and Guangxi Medical University Training Program for Distinguished Young Scholars(2019GXNSFAA245083).

AUTHORS’ CONTRIBUTIONS

You Pan contributed to the conception of the study; Qinghua Huang, Jianlan Mo, Huawei Yang performed the experiment; Yinan Ji contributed significantly to analysis and manuscript preparation; Rong Huang performed the data analyses and wrote the manuscript; Yan Liu helped perform the analysis with constructive discussions. All authors have read and approved the manuscript.
ACKNOWLEDGMENTS

Not applicable.

References


Figures
Figure 1

Flowchart of this study
Figure 2

Characteristics and differences of 24 m7GRGs in breast cancer. (A) The expression of 24 m7GRGs in breast cancer and normal tissues. *p<0.05, **p<0.01, ***p<0.001. (B) and (C) The mutation frequency and classification of 24 m7GRGs in breast cancer. (D) Circos graph for univariate cox regression analysis, which represents the association of 24 m7GRGs expression and prognosis of breast cancer in the TCGA-BRCA cohort.
Figure 3

Subgroups of breast cancer related by 24 m7GRGs. (A) Correlations among 24 m7GRGs in breast cancer. (B) The consensus score matrix of breast cancer samples when $k = 4$ in TCGA-BRCA cohorts. (C) Census CDF curves for the TCGA-BRCA cohort. (D) Delta area under CDF curve shows the change of accumulative risk along with the increased consensus clustering matrixes and demonstrated that three
clusters were optimal (k=4). (E) PCA plots for 4 clusters in the TCGA-BRCA cohort. (F) Expression of 24 m7GRGs among 4 breast cancer subtypes. *p<0.05, **p<0.01, ***p<0.001.

Figure 4

Landscape of immune infiltration among 4 breast cancer subtypes. (A) Distribution of 22 immune cells in 1023 breast cancer samples. (B) Correlations among 22 immune cells in the TCGA-BRCA cohort. (C) Differences in infiltration levels of 22 immune cells among 4 breast cancer subtypes. *p<0.05, **p<0.01, ***p<0.001, ns means no significance.
Figure 5

Results of ESTIMATE analysis among 4 breast cancer subtypes. (A) ESTIMATEScore for 4 breast cancer subtypes. (B) ImmuneScore for 4 breast cancer subtypes. (C) StromalScore for 4 breast cancer subtypes. (D) TumorPurity for 4 breast cancer subtypes. p-values less than 0.05 were considered statistically significant.
Figure 6

Figure 7

KEGG enrichment analysis among four breast cancer subtypes. (A) Subtype 1 vs. Subtype 2. (B) Subtype 1 vs. Subtype 3. (C) Subtype 1 vs. Subtype 4. (D) Subtype 2 vs. Subtype 3. (E) Subtype 2 vs. Subtype 4. (F) Subtype 3 vs. Subtype 4. p-values less than 0.05 were considered statistically significant.
Figure 8

GSEA enrichment analysis among four breast cancer subtypes. (A) Subtype 1 vs. Subtype 2. (B) Subtype 1 vs. Subtype 3. (C) Subtype 1 vs. Subtype 4. (D) Subtype 2 vs. Subtype 3. (E) Subtype 2 vs. Subtype 4. (F) Subtype 3 vs. Subtype 4. p-values less than 0.05 were considered statistically significant.
Figure 9

Construction of a prognostic 4 m7GRGs-related model. (A) and (B) Accuracy and error of 5-fold cross-validation (CV) in SVM analysis, respectively. (C) Forest plot for hazard ratios of 4 m7GRGs. (D) Distribution of patients in the TCGA-BRCA based on the risk score. (E) KM curves for breast cancer patients in the high-/low-risk group in TCGA-BRCA. (F) The survival status for each patient in the TCGA-BRCA. (G) KM curves for breast cancer patients in the high-/low-risk group in GSE1456. (H) Heatmap for
the connections between expression of 4 m7GRGs and the risk groups in the TCGA-BRCA. p-values less than 0.05 were considered statistically significant.

Figure 10

The prognostic value of 4 m7GRGs (AGO2, EIF4E3, DPCS and EIF4E) in breast cancer. The overall survival curve of AGO2 (A), EIF4E3 (C), DPCS (E) and EIF4E (G) in breast cancer patients in the high-/low-
expression group (GSE1456). The overall survival curve of AGO2 (B), EIF4E3 (D), DPCS (F) and EIF4E (H) in breast cancer patients in the high-/low-expression group (TCGA-BRCA). p-values less than 0.05 were considered statistically significant.

Figure 11

Construction of the nomogram. (A) and (B) Forest plot for hazard ratios of clinical the parameters in breast cancer. (C) 1-, 3-, and 5-year calibration curves for nomograms. A dashed diagonal line represents
the ideal nomogram. (D) Nomogram to predict the 1-, 3-, and 5-year overall survival (OS) rate of breast cancer patients. 1- (E), 3- (F), and 5-year (G) ROC curves for nomograms.

Figure 12
Tumor somatic mutation analysis between high and low-risk scores. (A) and (B) are waterfall charts for high and low-risk groups, respectively. (C) Boxplot of TMB scores in the high and low-risk groups. (D)
Scatter plot of the correlation between TMB score and risk score. (E) The OS survival analysis of BRCA patients in the high and low TMB groups. (F) The OS survival analysis of patients with high/low mutational burden in the high-risk group and patients with high/low mutational burden in the low-risk group.

Figure 13

Differences in drug sensitivity between high- and low-risk groups. (A) Erlotinib. (B) Gemcitabine. (C) Cytarabine. (D) Gefitinib. (E) MK.2206. (F) CCT007093.
Figure 14

The correlation between 4 prognostic m7GRGs and drug sensitivity in breast cancer. Correlation between DCPS and Dasatinib (A), Vorinostat (B), Pipobroman (E), Chelerythrine (J), Hydroxyurea (K) and Nelarabine (N). Correlation between EIF4E and Chelerythrine (C), Nelarabine (F), Amonafide (I) and Everolimus (L). Correlation between EIF4E3 and E-7820 (D), Hydrastinine HCl (G), Buthionine sulphoximine (H) and Nelnavir (M). Correlation between AGO2 and Imexon (O) and Teniposide (P).
Figure 15

The expression of AGO2 (A), EIF4E3 (B), DPCS (C), and EIF4E (D) in breast cancer cell lines.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementary.docx