

ORIGINAL ARTICLE

Identification and validation of N^7 -methylguanosine-associated gene NCBP1 as prognostic and immune-associated biomarkers in breast cancer patients

Jianrong Li¹ | Lin Zheng²  | Liying Song³ | Zhuangxia Dong⁴ | Wenqi Bai¹  | Liqiang Qi⁵

¹Department of General Surgery Sciences, Shanxi Province Cancer Hospital/Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China

²Department of Vascular Surgery, The Second Hospital of Shanxi Medical University, Taiyuan, China

³Thyroid Surgery Department, First Hospital of Shanxi Medical University, Taiyuan, China

⁴Gastroenterology, Shanxi Province Cancer Hospital/Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China

⁵Department of Breast Surgery, National Cancer Center/National Cancer Clinical Medical Research Center/Cancer Hospital, Chinese Academy of Medical Sciences, Beijing, China

Correspondence

Wenqi Bai, Department of general surgery Sciences, Shanxi Province Cancer Hospital/Shanxi Hospital Affiliated to Cancer Hospital, Chinese Academy of Medical Sciences/Cancer Hospital Affiliated to Shanxi Medical University, Taiyuan, China.
Email: baiwenqi1998@163.com

Liqiang Qi, Department of Breast Surgical Oncology, Cancer Institute and Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Chaoyang District, Panjiayuan 17#, Beijing, 100021, China.
Email: qi_liqiang@foxmail.com

Abstract

We intend to evaluate the importance of N^7 -methylguanosine (m7G) for the prognosis of breast cancer (BC). We gained 29 m7G-related genes from the published literature and among them, 16 m7G-related genes were found to have differential expression. Five differentially expressed genes (CYFIP1, EIF4E, EIF4E3, NCBP1 and WDR4) were linked to overall survival. This suggests that m7G-related genes might be prognostic or therapeutic targets for BC patients. We put the five genes to LASSO regression analysis to create a four-gene signature, including EIF4E, EIF4E3, WDR4 and NCBP1, that divides samples into two risky groups. Survival was drastically worsened in a high-risk group ($p < 0.001$). The signature's predictive capacity was demonstrated using ROC (10-year AUC 0.689; 10-year AUC 0.615; 3-year AUC 0.602). We found that immune status was significantly different between the two risk groups. In particular, NCBP1 also has a poor prognosis, with higher diagnostic value in ROC. NCBP1 also has different immune states according to its high or low expression. Meanwhile, knockdown of NCBP1 suppresses BC malignancy in vitro. Therefore, m7G RNA regulators are crucial participants in BC and four-gene mRNA levels are important predictors of prognosis. NCBP1 plays a critical target of m7G mechanism in BC.

KEYWORDS

breast cancer, differentially expressed genes, immune, N^7 -methylguanosine, NCBP1

Jianrong Li and Lin Zheng contributed equally to this work and should be considered co-first authors.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Cellular and Molecular Medicine* published by Foundation for Cellular and Molecular Medicine and John Wiley & Sons Ltd.

1 | INTRODUCTION

One of the most prevalent cancers in females, breast cancer (BC) has a devastating impact on patients' quality of life.¹ Early clinical indications of the illness are not evident, and tiny lesions are difficult to detect in clinical breast examinations, breast ultrasounds and mammograms. Also, recent studies have been searching for predictive and prognostic markers but they are not yet available for clinical use in BC patients.² Thus, most women do not receive prompt treatment until more visible clinical signs appear, missing the window of opportunity for surgery altogether.³ Especially for triple-negative BC, the incidence of brain metastasis has been estimated between 25% and 46%.⁴ Prediction of prognostics is difficult because of sophisticated detection procedures with low specificity. Furthermore, given the restricted therapy options for BC,⁵ there is a pressing need for new prognostic models to be developed.

*N*⁷-methylguanosine (m7G) exists in the internal sites of transfer RNAs and rRNA which are the most significantly edited RNA species in cells, including m7G. More and more studies have proved m7G's correlation with different diseases in humans, in particular carcinoma.^{6–8} Human malignancies are linked to epigenetic alterations in RNA that aren't properly controlled. m7G has been identified to take a significant role in BC in previous research, and several genes, such as NSUN2⁹ and the METTL1 gene,¹⁰ are known to adversely regulate m7G. METTL1 methyltransferase mediates m7G methylation within let-7e miRNA, activating the let-7e miRNA/HMGA2 axis which was linked with a poor prognosis. Meanwhile, other m7G-related genes, such as DCP2¹⁰ and DCPS, may cause illness by committing the transcript to destruction. However, it is unclear if these m7G-regulated genes are associated with the prognosis of BC samples.

To carry out this research, we used TCGA databases to gather gene expression profiles and information from BC patients. Next, using m7G-regulated differentially expressed genes (DEGs), we created a predictive multigene signature that was confirmed by qRT-PCR. Moreover, we identified potential key regulators and experimentally validated the underlying mechanism. Finally, we used functional enrichment analysis to study the main processes. The overview of the design and main findings of the current study is displayed in Data S1.

2 | METHOD

2.1 | Data collection

The TCGA database was used to retrieve the RNA sequence data (RNA-seq) as well as patient medical records of BC. The 'limma' R package's scaling method was used to standardize gene expression profiles.¹¹ The read count numbers were normalized. The TCGA data are accessible to the public. Consequently, clearance from regional ethics committees was not required. In this study, we adhere to the TCGA's criteria for data access and dissemination. Then, 29

m7G-related genes were found in prior publications¹² and the gene list of m7G modification in GSEA public datasets (<http://www.gsea-msigdb.org/gsea/login.jsp>), which are included in Data S2.

2.2 | Establishing and confirming a prognosis m7G-regulated gene signature

DEGs from tumour organs and nearby healthy tissue (FDR < 0.05) were identified using the 'limma' R program. The univariate Cox regression analysis of overall survival (OS) was performed to determine whether the genes or the model are significantly predictive in m7G. *p*-value was adjusted using the Benjamini and Hochberg method (BH). Networks of interacting DEGs (version 11.0) were constructed in the STRING database. LASSO method was used to build a prognosis model and reduce the possibility of overfitting via the processes of factor selection and shrink with the function of the 'glmnet' R package.¹³ Dependent variables included OS and clinical status, and the standardized expression matrix of predictive DEGs was the independent variables in the regression. Using minimum criterion, tenfold cross-validations were used to determine the model's penalty parameter. Patient's risk scores were calculated using the normalized amounts of mRNA, as well as the related regression coefficients. The following formula was devised:

$$\text{Risk score (patients)} = \sum_{i=1}^n \text{coefficient (gene}_i\text{)} \times \text{expression value (gene}_i\text{)}.$$

Based on the median level of the risk score, the samples were classified into low- and high-risk groups. Based on the expression matrix of genes in the signature, PCA was done utilizing the 'prcomp' technique of the 'stats' package. The finest expression cut-off value for each gene's survival analysis was established using the 'survival' R program. To verify the gene signature's predictive value, time-dependent ROC curve analysis was performed using the 'timeROC' R package.

2.3 | Functional enrichment analysis

DEGs were conducted with functional enrichment analysis of GO and KEGG analysis ($|\log_2\text{FC}| > 1$, FDR < 0.05) using 'clusterProfiler' package in R.¹⁴ The *p*-values were modified using the BH method. The activity of 13 immune-related pathways and the grade of invasion of 16 immune cells were evaluated using ssGSEA via the 'gsva' package. The file for the annotation gene set can be found in Data S3. To identify the biological functional differences between the two risk groups, gene set variation analysis (GSVA) was conducted.

2.4 | Tumour mutation burden analysis

The total number of somatic mutations (excluding silent mutations) for each tumour sample was referred to as the tumour mutation

TABLE 1 Primers for qRT-PCR.

Gene name	Sequence (5'-3')
GAPDH	F: 5'-TGA CTTCAACAGCGACACCCA-3' R: 5'-CAC CCTGTTGCTGTAGCCAAA-3'
EIF4E	F: 5'-CTGCGGCTGATCTCCAAG-3' R: 5'-TTCCACATAGGCTCAATACC-3'
EIF4E3	F: 5'-AAGGGTGGCGTATGGAAGATGAAAAG-3' R: 5'-CGGTCCCGAACACTGACACTAAC-3'
NCBP1	F: 5'-GCTAAAGAGAAACTTGCTAGGC-3' R: 5'-GGTCAAGATCATGATAAACCCGC-3'
WDR4	F: 5'-AGAGTTTGTGAGCCGTATCTC-3' R: 5'-GAAGATGTAGACCACAGGAGTG-3'

burden (TMB). The 'tmb' function of the maftools R package calculated the TMB value for each sample.

2.5 | Correlation of NCBP1 with immune features

The prognostic values of the four genes utilized to build the risk model were evaluated using Kaplan–Meier and ROC analyses. NCBP1 was found to have excellent predictive efficiency. GSEA was performed to explore the potential mechanism underlying the function of NCBP1 in BC. For GSEA analysis, the R packages 'enrichplot', 'clusterProfiler', 'limma' and 'org.Hs.eg.db' were utilized. To investigate the correlation between the level of NCBP1 and the enrichment scores of several types of immunity cells, we used the CIBERSORT algorithm. We further analysed the relationships between the expression levels of NCBP1 and immune checkpoint genes and visualized the results in a heatmap.

2.6 | Statistical analysis

To assess gene expression between cancer samples versus neighbouring normal tissues, a Student's *t*-test was conducted. To compare proportional differences, the chi-squared test was performed. The Mann–Whitney test with *p* values adjusted by the BH approach was performed to evaluate the ssGSEA scores of immune cells or pathways between the different levels of risk groups. Using Kaplan–Meier analysis, the log-rank test was utilized to compare the OS of different groups. Both uni- and multivariate Cox regression analyses were used to identify independent factors of OS. All statistical calculations were performed using R software (Version 3.5.3). Unless otherwise noted, a *p*-value of 0.05 or less was considered statistically significant.

2.7 | Cell lines and cell culture

Hunan Fenghui Biotechnology Co., Ltd. provided the normal mammary epithelial cell line MCF10A and the BC cell lines MDA-MB-231. Beijing Biosynthesis Biotechnology Co., Ltd. provided the BC

cell lines MCF-7. MDA-MB-231 and MCF-7 cells were grown in Dulbecco's modified Eagle's medium with 10% (v/v) fetal bovine serum (Gibco) and 1% penicillin–streptomycin. MCF10A cells were cultured in DMEM/F12 media with 5% (v/v) FBS, 20 ng/mL EGF, 10 g/mL insulin and 500 ng/mL hydrocortisone. These cell lines were maintained at 37°C in a 5% CO₂ atmosphere.

2.8 | Analyses using the quantitative real-time polymerase chain reaction

Total RNA was isolated from BC cells using the TRIzol reagent. A NanoDrop 2000 spectrophotometer was used to determine the concentration and purity of total RNA (Thermo Fisher Scientific, Wilmington, DE, USA). The MonScript™ RTIII Super Mix with dsDNase kit created complementary DNA (Monad). The following conditions were used to run quantitative real-time PCR using MonAmp™ ChemoHS qPCR Master Mix: 95°C for 10 min, followed by 40 cycles at 95°C for 10 s, 55°C for 30 s. Table 1 contains the primer sequences for detection. GAPDH expression was employed as an internal standard control. The $2^{-\Delta\Delta Ct}$ technique was used to gather and quantify all gene expression levels. Three separate experiments yielded the findings.

2.9 | Western blotting

The protein was extracted with western blot (WB) lysis buffer (BMP2020, Abbkine) containing PMSF and quantified with a BCA kit (SW101-02, Severn Biotech). An equal amount (20 µg) of protein extracts was separated on 10% SDS-PAGE gels then transferred to PVDF membranes (Immobilon-P, Millipore). The cell membrane was treated with the appropriate initial antibodies at 4°C overnight after being blocked in 5% non-fat milk for 1.5 h. After washing three times with TBST for 10 min, the membranes were incubated with the secondary antibody (1: 5000 dilution; Biosharp, China) for 2 h at room temperature. After washing with TBST for half an hour, the signals were visualized using ECL reagent (BMU102-CN, Abbkine, China) and quantified using Image Lab Software (Bio-Rad). Primary antibodies were listed as follows: EIF4E (AB33766, Abcam), NCBP1

(10349-1-AP, Proteintech), WDR4 (PAB40170, Bioswamp), GAPDH (BM1623, Boster Biological Technology Co. Ltd).

2.10 | siRNA interference assay

NCBP1 siRNA and scrambled siRNA were purchased from IBSBIO (Shanghai, China). A scrambled siRNA served as a contrast. The siRNA sequences were as follows (Table 2): siNCBP1-1: GGUAU GGACUGCUGAUAAACC, siNCBP1-2: GGAUGAUGACGACGAUG AAGG, siNCBP1-3: GGAAGAAGCUAAAGAGAAACU and siRNA-NC: UUCUCCGAACGUGUCACGUTT. According to the manufacturer's instructions, the siRNAs were transfected into BC cells using Lipofectamine 2000 (Invitrogen). Total RNA was extracted 48 h after transfection, and protein was extracted 72 h after transfection. RNA and protein were estimated, respectively by quantitative real-time polymerase chain reaction (RT-PCR) and WB.

2.11 | CCK-8 assay

After MDA-MB-231 cells were transfected with NC and siNCBP1 for 24 h, they were respectively seeded into 96-well culture plates (5×10^3 /well). Each well need to add DMEM medium (100 μ L). The cell proliferation was measured using a Cell Counting Kit-8 at 0, 24, 48 and 72 h. 10 μ L of CCK8 solution was added to each well, and cells were incubated for 2 h at 37°C at these points. Finally, a microplate reader was used to measure the absorbance of each well.

2.12 | Cell migration and invasion assays

FBS-free cell suspensions were seeded in a transwell chamber (Corning, Kennebunk, ME, USA) ($1-2 \times 10^4$ /well) coated with Matrigel (invasion) or without Matrigel (migration). The lower well was supplemented with a medium containing 20% FBS. After incubation for 48 h, cells across the membrane were fixed with 4% paraformaldehyde for 20 min and then treated with crystal violet dye for 10 min. The cells in the upper chamber were gently removed with a cotton swab. Cells were photographed and counted under a microscope.

TABLE 2 siRNA sequences for knockdown NCBP1.

siRNA	siRNA sequences
NCBP1-643	Sense: 5'-GGUAUGGACUGCUGAUAAACC-3'
	Anti-sense: 5'-UUUAUCAGCAGUCCAUAACUG-3'
NCBP1-1585	Sense: 5'-GGAUGAUGACGACGAUGAAGG-3'
	Anti-sense: 5'-UUCAUCGUCGUAUCAUCCUG-3'
NCBP1-1975	Sense: 5'-GGAAGAAGCUAAAGAGAAACU-3'
	Anti-sense: 5'-UUUCUCUUUAGCUUCUCCAG-3'
siRNA-NC	Sense: 5'-UUCUCCGAACGUGUCACGUTT-3'
	Anti-sense: 5'-UUUCUCUUUAGCUUCUCCAG-3'

3 | RESULTS

3.1 | Identify prognostic m7G-related DEGs

The majority of m7G-related genes (16/29, 55.2%) were found to vary between tumour and nontumorous tissues (Table 3). In the univariate cox regression analysis, six genes have been shown to be linked with OS among 29 m7G-related genes (Figure 1A). Finally, five m7G-related genes in total are not only expressed differently but they were also linked to OS (FDR < 0.05, Figure 1B-D). The interaction network among these genes revealed that EIF4E, EIF4E3, NCBP1 and CYFIP1 are all strongly related (Figure 1E). Figure 1F depicts the relationship between these genes.

3.2 | Establishing and validating a prognostic model

The expression patterns of the five genes mentioned above were used in LASSO analysis to create a prognostic model. Based on the optimum value, a four-gene signature was discovered (Data S4). The risk score was calculated using the formula shown below: score = 1.2215 \times EIF4E expression level + 0.9397 \times EIF4E3 expression level + 1.0390 \times NCBP1 expression level + 1.0333 \times WDR4 expression level. The patients were split into two groups based on their median threshold values: high- and low-risk groups (each group number = 545) (Figure 2A). According to PCA, patients in distinct risk categories were dispersed in two directions (Figure 2B). Individuals in the high-risk group had a significantly lower life expectancy than their low-risk counterparts, according to the Kaplan-Meier curve (Figure 2C, $p < 0.001$). The AUC for the risk score for OS was 0.602 after 3 years, 0.615 after 5 years, and 0.689 after 10 years, according to time-dependent ROC curves (Figure 2D).

3.3 | The four-gene signature's important predictive value

To confirm if the risk score was an unbiased predictive factor of OS, researchers ran cox analysis (univariate and multivariate) on the relevant factors. The risk score was strongly linked with OS in univariate Cox regression models in the TCGA cohorts (HR = 1.9472, 95% CI = 1.3408-2.8279, $p = 0.0005$) (Data S5). When additional confounding factors were excluded from the Cox regression (multivariate) analyses, the risk score maintained an independent prognostic factor of OS (HR = 2.1272, 95% CI = 1.4529-3.1143, $p < 0.001$).

3.4 | Functional analysis

DEGs of low and high groups were applied in GO and KEGG pathway analysis to discover associated mechanisms and pathways associated with risk scores. Surprisingly, several biological processes

TABLE 3 The significant results of differently expressed analysis among m7G-related genes.

Number	Gene	logFC	FDR
1	CYFIP1	-0.43883829	9.37E-33
2	DCPS	0.563991753	7.42E-15
3	EIF4E	0.345914628	7.31E-13
4	EIF4E1B	2.740602283	1.85E-06
5	EIF4E2	0.192549034	1.23E-06
6	EIF4E3	-1.115893709	4.45E-43
7	EIF4G3	-0.177080175	1.02E-07
8	GEMIN5	-0.211412734	2.37E-09
9	LSM1	0.604553678	2.84E-09
10	METTL1	0.722795603	1.96E-26
11	NCBP1	0.24329601	2.91E-08
12	NCBP2	0.168526178	3.27E-05
13	NCBP3	-0.269352253	2.43E-17
14	NUDT16	-0.657372405	4.58E-40
15	NUDT3	0.335225767	4.22E-12
16	WDR4	0.792631846	9.58E-26

and KEGG pathways associated with neuroactive ligand–receptor interaction were clearly enhanced in the DEGs ($p < 0.001$, Data S6).

We employed ssGSEA to calculate the enrichment scores of various immune cell subpopulations, associated functions, and pathways in order to examine the correlation between the risk score and the immunological state. Significant differences were observed in the contents of the antigen presentation process, including neutrophil, B cells, DCs, mast cells, Th1 cells and T helper cell scores, among the groups ($p < 0.001$, Figure 3A). Within the low-risk individuals, higher scores were observed for the Type II IFN response and HLA ($p < 0.001$, Figure 3B). We then utilized the GSVA package to investigate the KEGG pathways associated with the risk score (Figure 3C). These pathways, which are involved in nucleic acid metabolism and protein metabolism, include the pentose phosphate pathway, citrate cycle, terpenoid backbone biosynthesis, and steroid biosynthesis.

3.5 | Examining the frequency of gene mutations and the chromosomal copy number variation in BC patients

Upon completing the process of filtering away samples with missing information taken from the TCGA dataset, we obtained mutation information of BC patients, which we used to investigate the effects of mutations for these patients. First, we tried to analyse the mutation profile of BC patients, finding that missense mutations accounted for the vast majority, that SNPs were the most common type of variant, and that C>T was the most common type of SNP in patients with BC (Figure 4A–C). We obtained the number of base changes for each patient and distinguished the mutation types with different colours in the box plot (Figure 4D,E). PIK3CA (34%), TP53 (34%), TTN (17%),

CDH1 (13%), GATA3 (13%), MUC16 (10%), MAP3K1 (9%), KMT2C (9%), HMCN1 (6%) and RYR2 (6%) were the top five genes with the highest mutation frequencies in BC (Figure 4F). The top 30 mutant genes for each patient are specifically listed in Figure 4G. Figure 4H reveals the relationship between these genes.

3.6 | Relationship between four-gene signature and drug sensitivity

We investigated the IC_{50} values of 20 chemotherapeutics between the groups at high and low risk. The result showed that IC_{50} values of imatinib, lenalidomide, masitinib, mitomycin C, navitoclax, pyrimethamine, thapsigargin, tivozanib, veliparib, AICAR, epothilone B and gefitinib of the high-risk group were less than those of low-risk group, indicating that patients in the high-risk group were more likely to benefit from these 20 drugs (Data S7).

3.7 | Survival analysis of the four genes in breast cancer

OS analyses of the four genes (EIF4E, EIF4E3, NCBP1, WDR4) used to construct the model were further conducted by the Kaplan–Meier plotter. As demonstrated in Figure 5A–D, the high expression level of NCBP1 in BC patients was associated with poor OS ($p = 0.004$), while the high expression level of EIF4E3 in BC patients was associated with better OS ($p = 0.016$). Moreover, according to the ROC result of other genes, NCBP1 had better 3-, 5- and 10-year prognosis in BC patients (Figure 5E, Data S8). To investigate the potential function of NCBP1, we performed a GSEA analysis (Figure 5F). The result shows that the function of NCBP1 in BC is mainly enriched in the neuroactive interaction between ligand and receptor and transduction.

3.8 | Relationship between NCBP1 and tumour immune microenvironment

To investigate the relationship between NCBP1 expression and cells infiltrating the immune system in BC, we employed the ssGSEA method and the 'limma' R package. The findings demonstrated a correlation between the expression of NCBP1 and the infiltration of immune cells in the tumour (Figure 6A), including naive B cells, CD8 T cells, CD4 memory resting T cells, CD4 memory activated T cells, regulatory T cells, NK cells resting and activated NK cells, macrophages M1 and resting mast cells. NCBP1 has a positive correlation with the expression levels of macrophages M1 ($r = 0.12$, $p = 0.00049$), NK cells resting ($r = 0.11$, $p = 0.0022$), T cells CD4 memory activated ($r = 0.23$, $p = 1.5e-11$) and T cells CD4 resting ($r = 0.14$, $p = 2.9e-05$), while negatively correlated with the expression levels of B cell memory ($r = -0.11$, $p = 0.0015$), mast cells resting ($r = -0.12$, $p = 0.00049$), NK cells activated ($r = -0.2$, $p = 3.4e-09$), T cells CD8

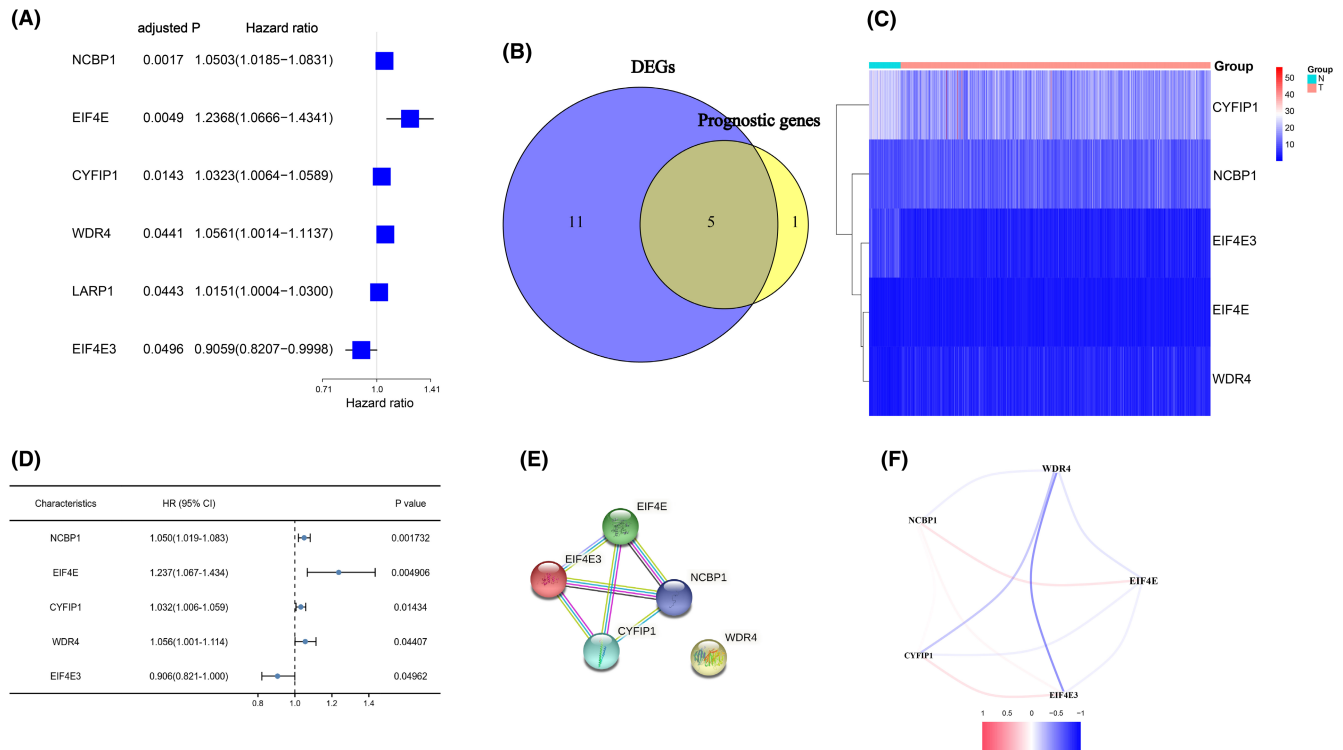


FIGURE 1 The candidate m7G-related genes have been identified. (A) Forest plots displaying the findings of the univariate Cox regression analysis of 6 genes with $p < 0.05$. (B) A Venn diagram was used to identify genes that were not only differently expressed between the tumour and neighbouring normal tissue but also were associated with overall survival (OS). (C) In tumour tissue, all five overlapping genes were elevated. (D) Forest plots displaying the findings of the univariate Cox regression analysis of gene expression versus OS. (E) The interactions between the candidate genes were highlighted by the PPI network obtained from the STRING database. (F) The candidate gene correlation network. Different hues show the correlation coefficients.

($r = -0.11$, $p = 0.00098$) and T cells regulatory ($r = -0.18$, $p = 1.7e-07$) (Figure 6B–J). Figure 6K,L shows the relationship between NCBP1 and immunity checkpoints. We found that the expression levels of TNFSF15, NRP1, PDCD1LG2, TNFSF4, TNFRSF9, CD160, CD80, IDO1, ICOS and CD274 correlated strongly to NCBP1, while TNFRSF25, TNFRSF4, CD40, TNFRSF14, TNFRSF18, LGALS9 and TNFSF9 were negatively associated with the expression level of NCBP1. In conclusion, NCBP1 expression might be closely associated with the reprogramming of the immunosuppressive microenvironment in BC.

3.9 | Identification of NCBP1 critical role in BC

To better confirm the importance of NCBP1 in BC, researchers also confirmed the expression of four hub genes in normal mammary epithelial cell line MCF10A, the BC cell line MDA-MB-231, and the BC cell line MCF-7. The findings revealed that the expression of NCBP1 was increased in MDA-MB-231 and MCF-7, with MCF-7 showing a statistically significant difference ($p < 0.05$) (Figure 7A). The findings of WB also demonstrated that MB-231 and MCF-7 expressed themselves greatly more than did typical endothelial cells (Figure 7B,C). EIF4E is also highly expressed significantly in the MDA-MB-231 cell line, yet not in MCF-7 (Figure 7D).

3.10 | NCBP1 promotes proliferation, migration and invasion of BC

NCBP1 dysregulation in BC suggested that NCBP1 might influence the progression of BC. To test this hypothesis, we successfully constructed 231 cell lines with knocked-down NCBP1 by transfecting siRNA (Figure 7E–G). CCK-8 assays were used to assess the proliferation abilities of BC cells. As shown in Figure 7H, the proliferation rate of 231 cells was significantly repressed when knocking down NCBP1 levels. Transwell assays were performed to assess cell migration and invasion abilities. It was obvious that migration and invasion capabilities were suppressed in NCBP1 knocked-down cells (Figure 7I–K).

4 | DISCUSSION

Although some researchers^{15–17} have suggested that various genes may influence m7G in BC, their relationship with BC patients' OS remained uncertain. The current work explored the expression level of 29 m7G-regulated genes in BC tissue and their relationships with OS. Surprisingly, the majority (55.2%) of m7G-related genes were differentially expressed between tumour and normal samples, and five genes were associated with OS in Cox regression analysis. These

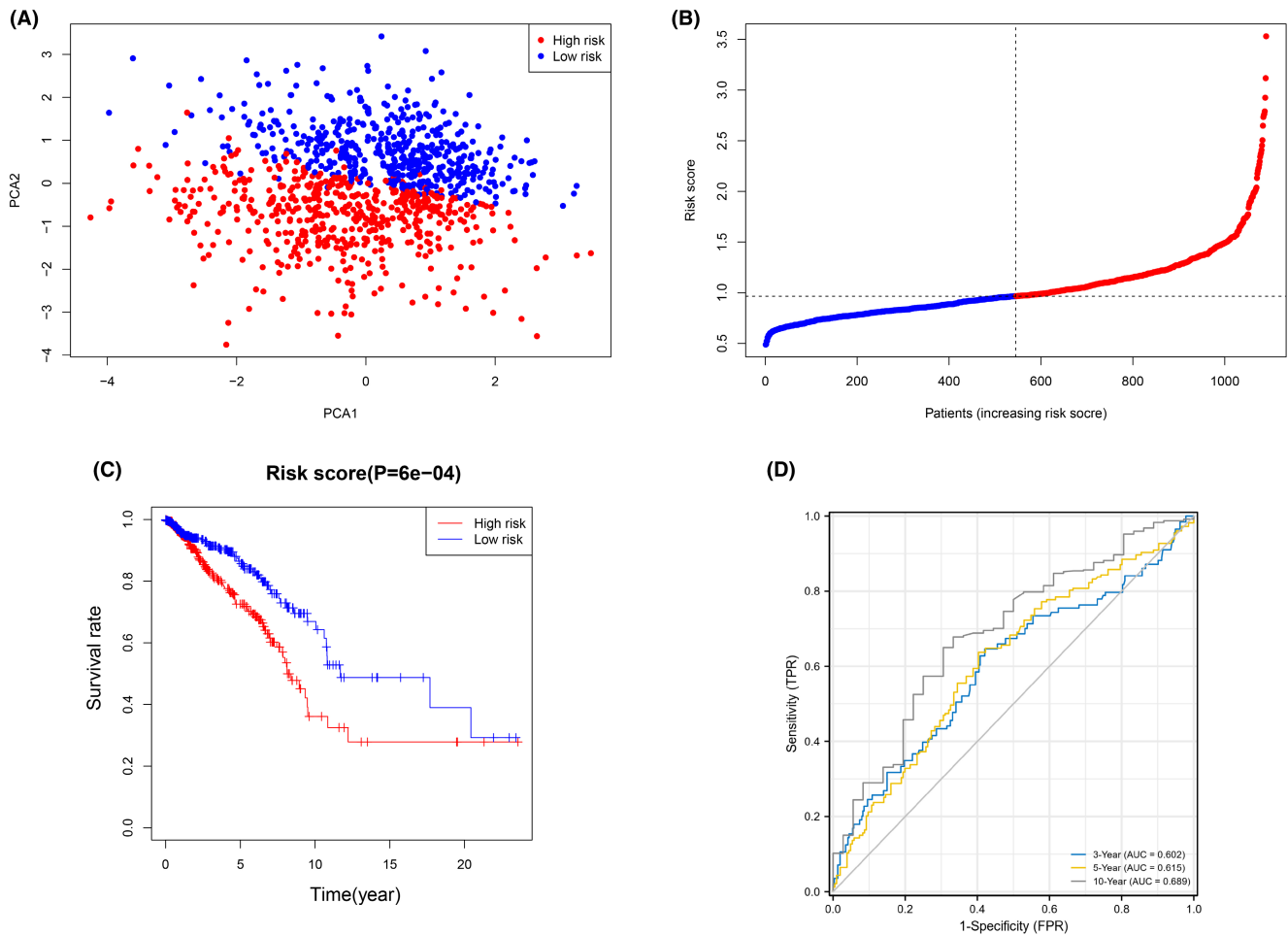


FIGURE 2 Prognostic analysis of the four-gene signature model. (A) The distribution and median value of the risk scores. (B) PCA plot of the breast cancer expression matrix. (C) The distributions of overall survival status, and risk score. (D) The risk score's prognostic ability was confirmed using the AUC of time-dependent ROC curves.

findings strongly suggested that m7G may have a role in BC, as well as the ability to develop a prognostic model using these m7G-regulated genes. Then one new predictive model including four m7G-regulated genes was developed and identified in a separate population. Analysis of enrichment function revealed that the mechanism of hub genes is associated with immune pathways. Moreover, we discovered a strong correlation between nuclear cap binding protein 1 (NCBP1) and the prognosis of BC patients. Further explorations on the relationship among NCBP1 and the tumour's microenvironment revealed that NCBP1 was significantly associated with several kinds of specific immunity cells like T cells or non-specific immunity like macrophages, and many immune checkpoints, providing a potential direction for the treatment of BC.

Cell viability and poly (A) RNA output require NCBP1 and nuclear cap binding protein 2 to form a nuclear cap binding complex (CBC), which opens the flow of genetic information from DNA to protein.^{18,19} In higher eukaryotes, NCBP1 and another nuclear cap binding protein NCBP3 together form another CBC, which binds to the components of the mRNA processing mechanism and promotes the output of poly (A) RNA.²⁰ It binds to the m7G'-cap structure of

freshly transcribed mRNA and coordinates downstream RNA biosynthesis processes such as nucleoplasm transport and recruitment of translation factors in the cytoplasm.^{21,22} The abnormal expression of NCBP1 was observed in lung adenocarcinoma and diffuse large B-cell lymphoma.^{23,24}

Meanwhile, EIF4E3 and EIF4E also exert an important effect on the development of BC based on our survival analysis. EIF4E3 (The translation initiation factor 4E), a member of the EIF4E family of translation initiation factors, interacts with the 5'-cap structures of mRNA.²⁵ In the eukaryotic protein translation initiation machinery, EIF4E (eukaryotic translation initiation factor 4) is a critical part of the protein known as the translation initiation factor complex. EIF4E recognizes 7-methylguanosine-containing caps on mRNA and binds to them, releasing the mRNA's secondary structure and so facilitating ribosome interaction at the initiation of protein synthesis. eIF4E expression and activation are linked to cell transformation and the tumour starts as a proto-oncogene. Specific RNA binding proteins and microRNAs, as well as influencing the 5-cap binding activity of EIF4E, may selectively modulate mRNA transcript translation.^{26,27} Previous research has revealed

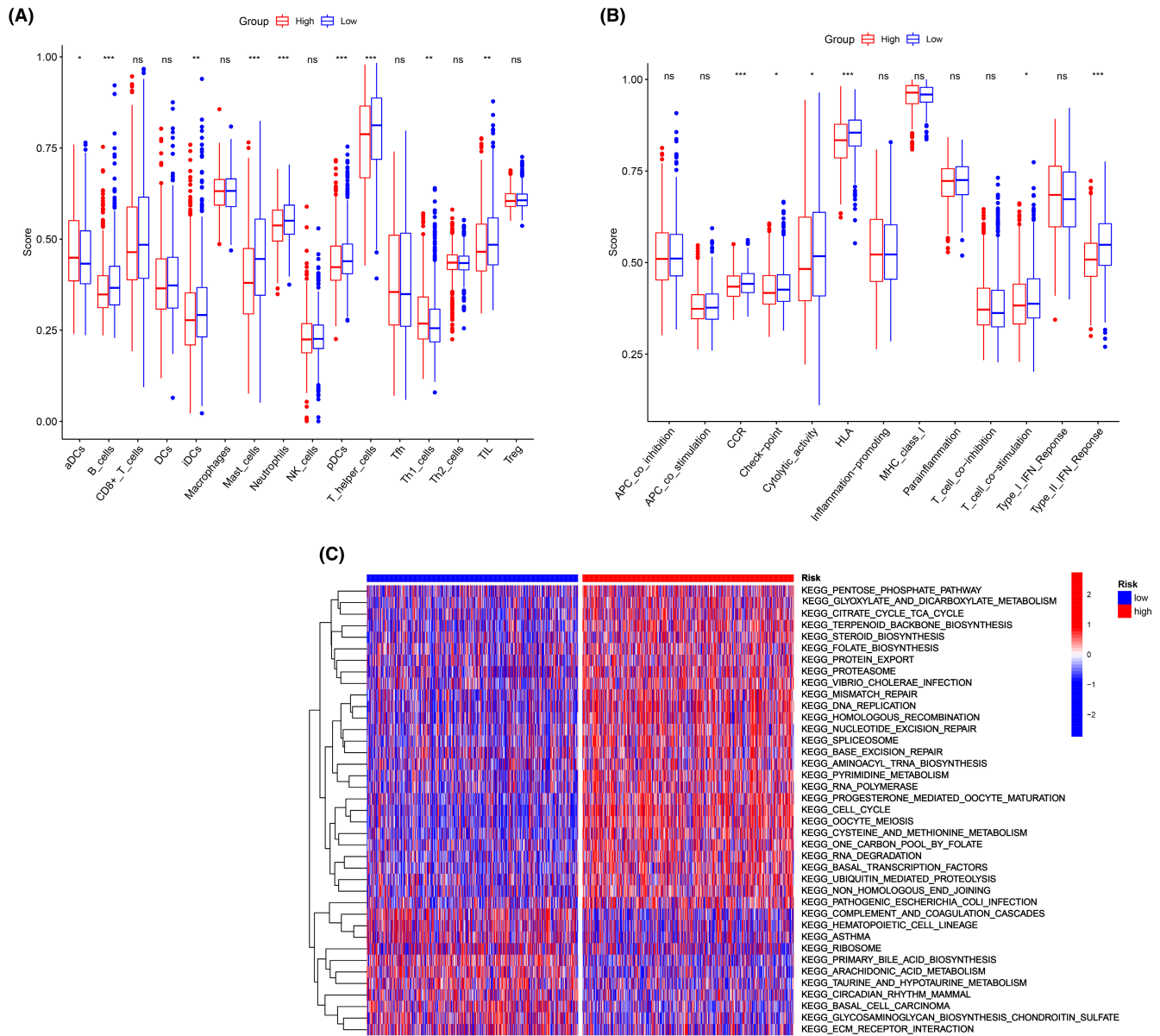


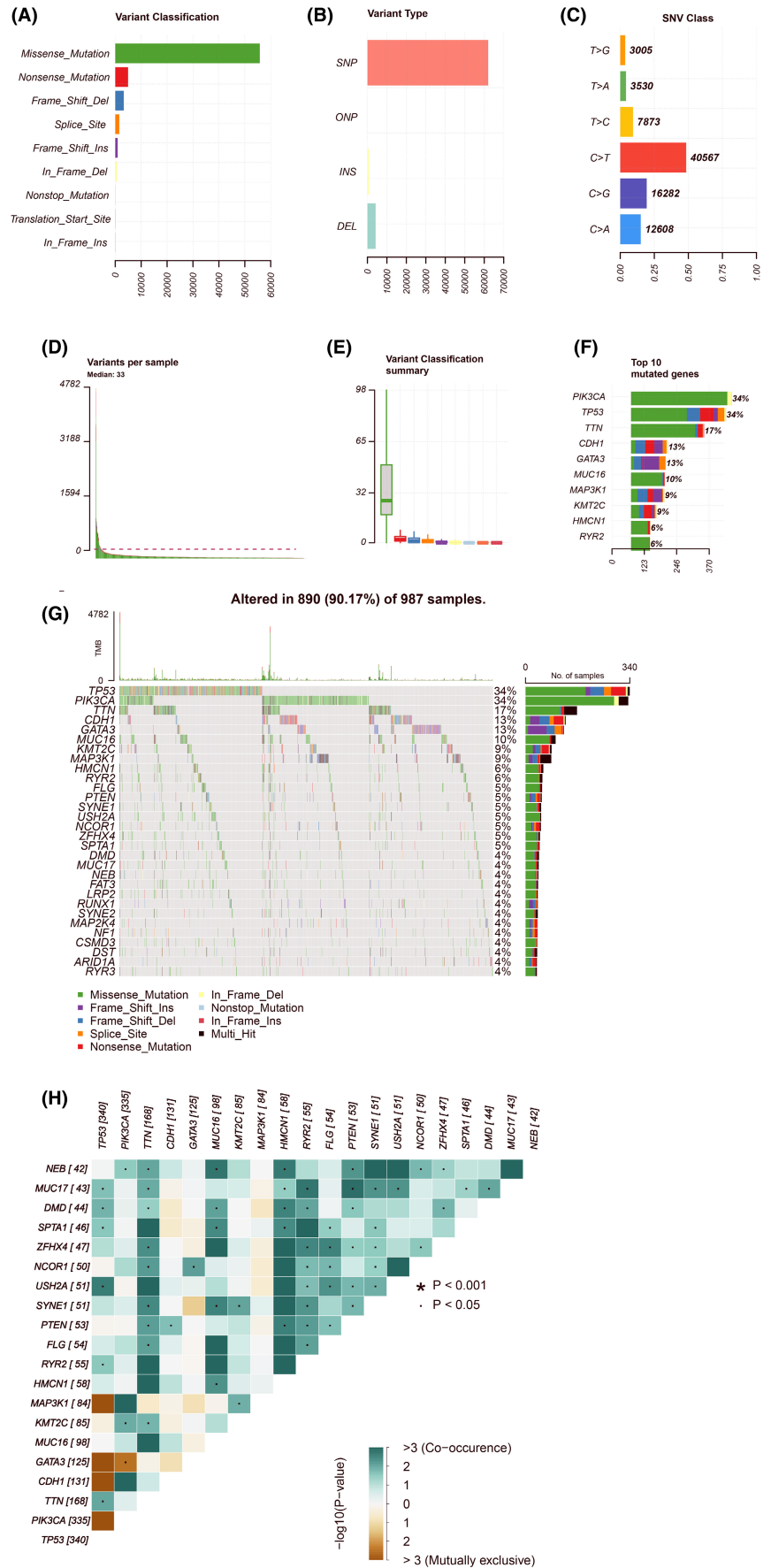
FIGURE 3 Analysis of ssGSEA scores over risk groups. Boxplots depict the scores for 16 immune cells (A) and 13 immunological-related activities (B). (C) GSVA of KEGG pathways between high- and low-risk groups. ns, not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

that individuals who had increased EIF4E expression are more likely to recur and die compared to those who had reduced eIF4E expression in triple-negative BC patients.²⁸ EIF4E3 is a member of the EIF4E family of proteins that bind to the 5'-cap structure of messenger RNA.^{27,29} According to one research, EIF4E3 depends on cap-binding activity to operate as a tumour suppressor and compete with EIF4E's growth-promoting capabilities. In fact, tumours that have increased EIF4E and low EIF4E3 expression demonstrate that EIF4E3 is of essential inhibitory mechanism that is lost in certain cancers.³⁰ Other investigations have shown that EIF4E3 may prevent oncogenic transformation.^{31–33} WDR4, a member of the WD repeat protein family, plays a crucial role in various aspects of cell development including cell cycle progression, signal transduction, gene regulation, apoptosis and

more.^{8,34,35} Studies have shown that WDR4 plays a crucial role in the development of malignant tumours, the aberrant expression of WDR4 is closely associated with tumour immunity and tumour mutation load.³⁶

While the mechanisms behind tumour sensitivity to m7G have been extensively studied, the potential impact of m7G on tumour immunity remains elusive. We used the DEGs to do a GO enrichment analysis, which showed significant enrichment for a wide variety of immune-related biological functions and pathways across all of the different risk categories. Especially, antigen presentation pathway components varied significantly from low-risk groups to high-risk groups in this study. Previous research has shown that Ribavirin would compete with EIF4E to join with m7G, hence limiting mRNA transcription and having a function in tumour therapy.³⁷

FIGURE 4 Panoramic view of breast cancer mutations in TCGA dataset. (A–F) Mutation panorama analysis; (G) top 30 gene mutation waterfall map; (H) The coincident and exclusive relationship among mutated genes.



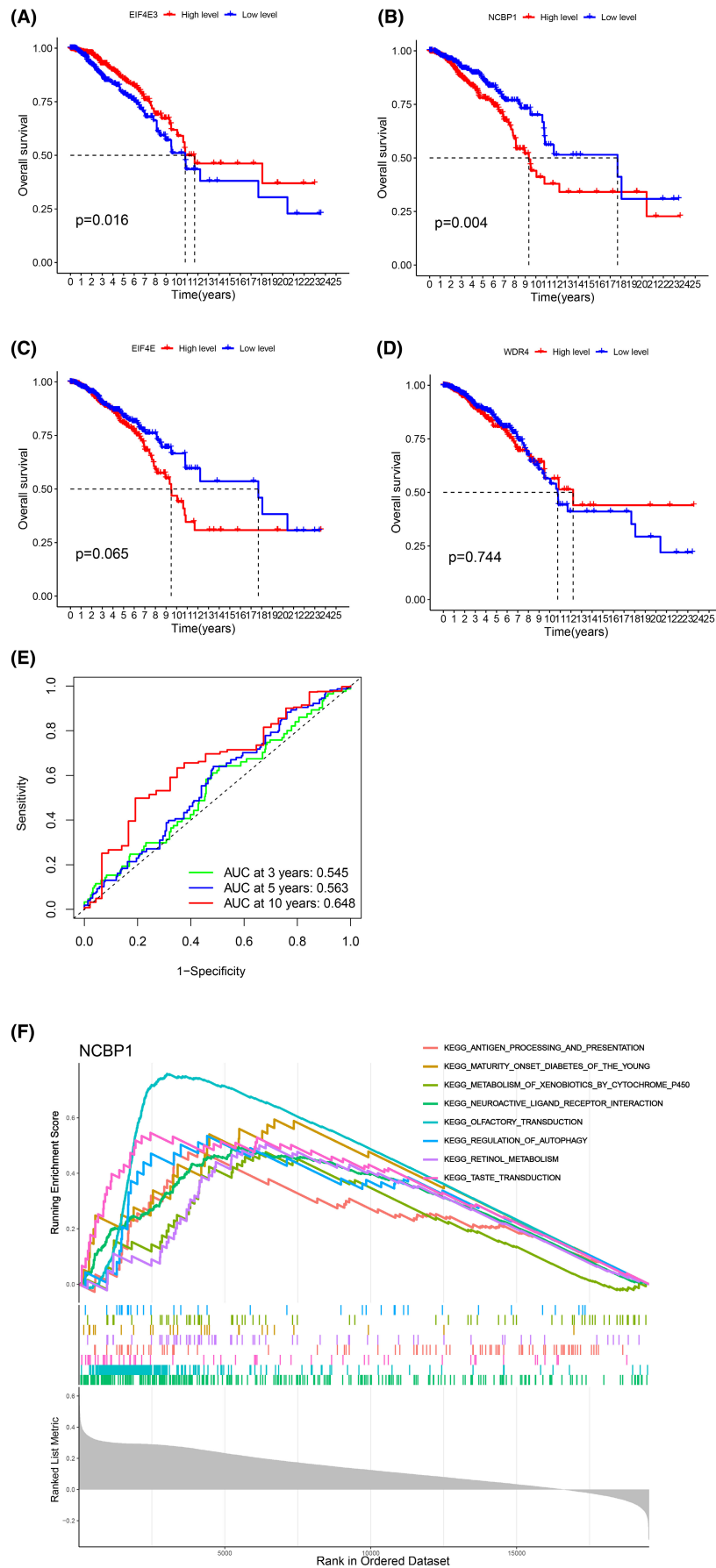


FIGURE 5 (A–D) Overall survival (OS) of the four genes in patients with breast cancer was analysed by Kaplan–Meier plotter. EIF4E3, $p=0.016$; NCBP1, $p=0.004$; EIF4E, $p=0.065$; WDR4, $p=0.744$. (E) ROC plots of NCBP1. (F) GSEA analysis of NCBP1.

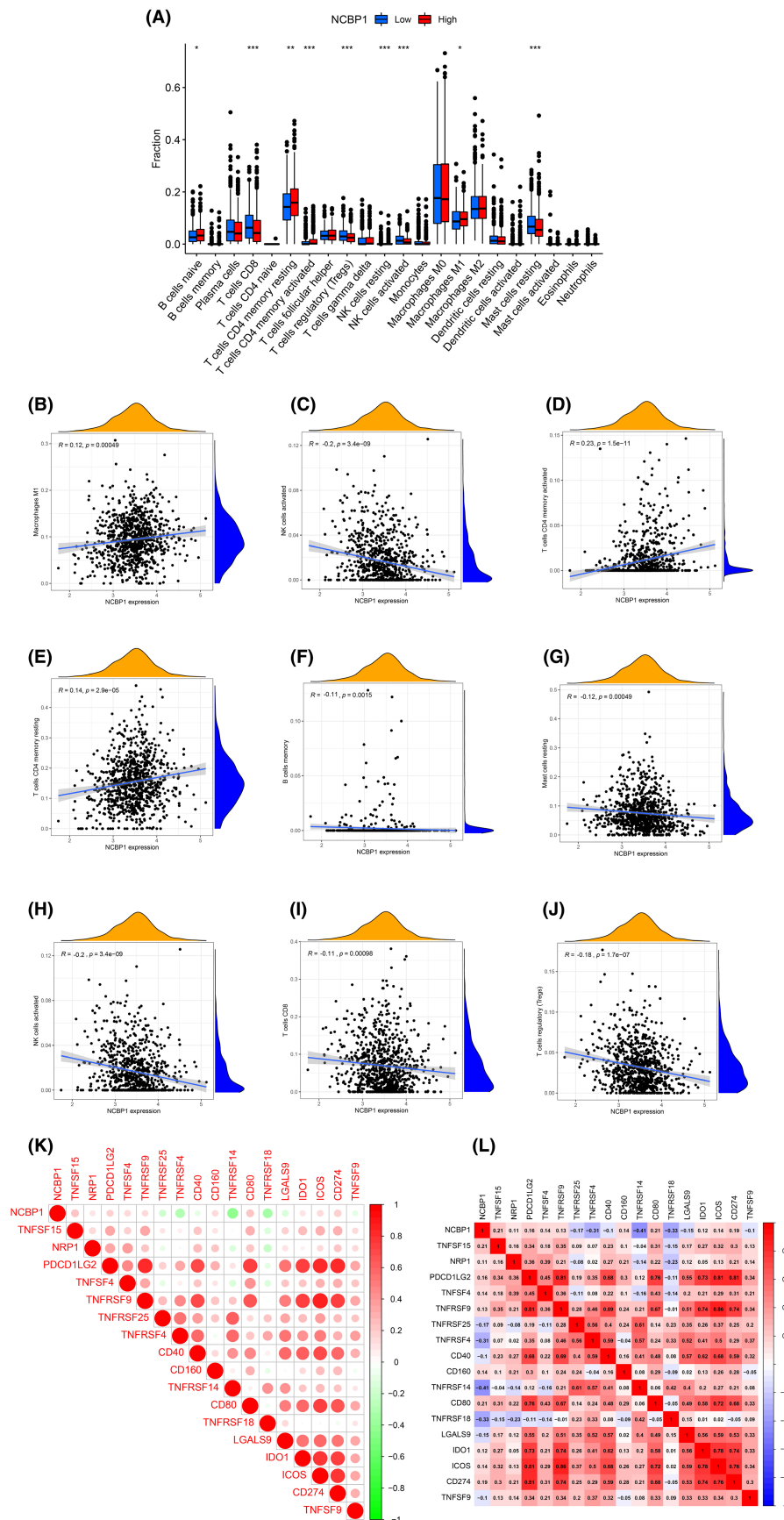


FIGURE 6 (A) Changes of 22 immune cell subtypes between high and low NCBP1 expression groups in breast cancer (BC) samples. (B–J) Correlation between the expression of NCBP1 and immune infiltrating cells in BC. (K, L) Correlations between NCBP1 and immune checkpoints.

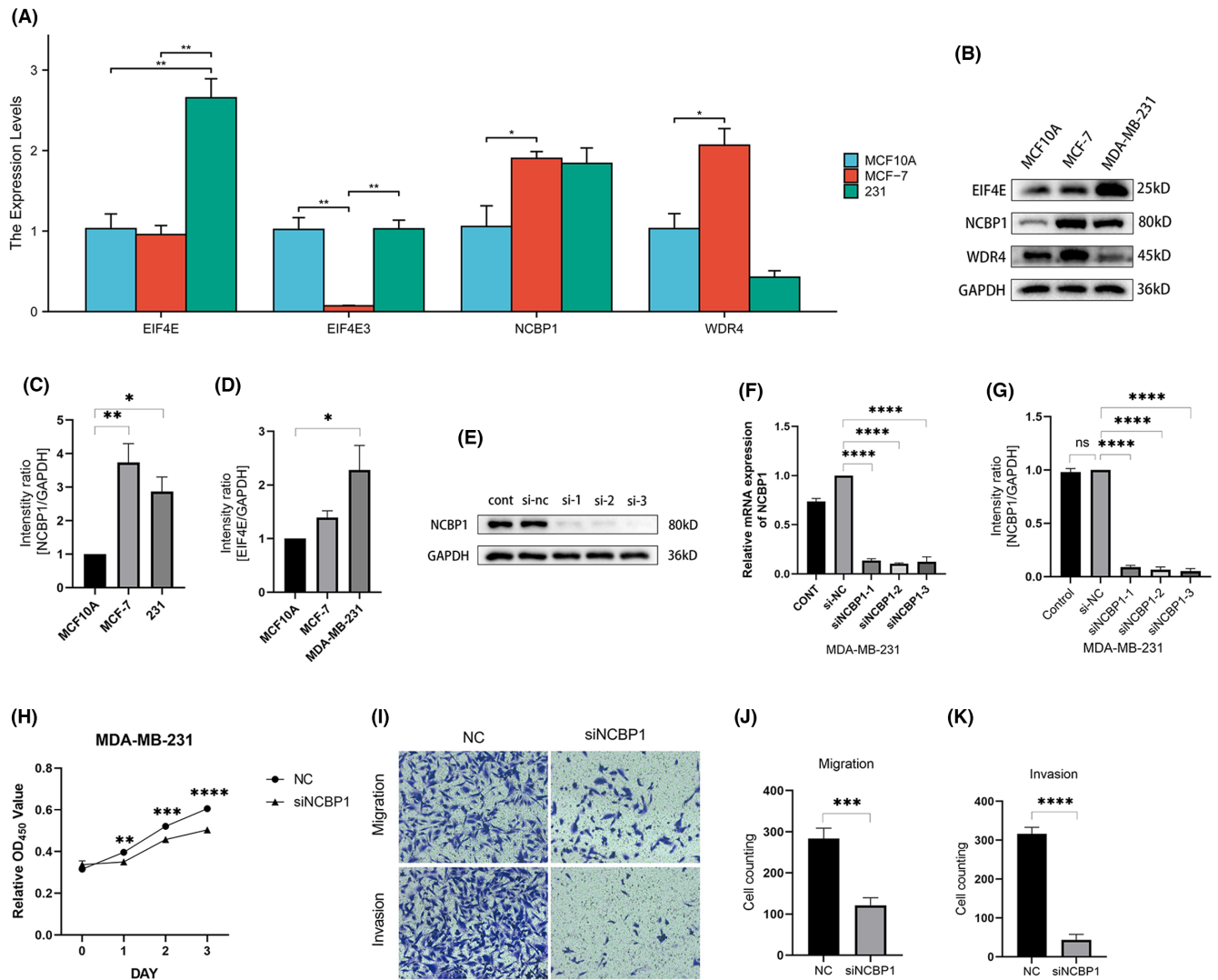


FIGURE 7 Hub genes expression level in breast cancer (BC) cell lines and the epithelial cell line. (A) Expression level of EIF4E3, EIF4E, NCBP1 and WDR4 in tumour and normal cells. Green represents 231 BC cells. Red presents MCF-7 BC cells. Blue represents MCF10A normal endothelial cells. (B–D) Western blot analysis for the hub genes in 231 BC cells, MCF-7 BC cells and MCF10A normal endothelial cells. (E–G) NCBP1 expression was verified after transfection in NCBP1. (H) Proliferation assays revealed attenuated capabilities of 231 after knocking down NNCBP1. (I–K) Migration and invasion abilities of 231 were weakened after knocking down NCBP1. * $p < 0.05$; ** $p < 0.001$.

It can directly improve tumour immunotherapy by targeting these methylation-regulating molecules. Using a bioinformatics tool to evaluate the database, it was shown that several modulators had a substantial link with tumour immune responses and immunotherapy in nasopharyngeal cancer, BC, and pancreatic cancer. Furthermore, the greater risk score was related to poorer antitumor immunity, especially type II IFN response activity. As a result, reduced immunity against tumours in high-risk individuals may explain their poor prognosis.

Through analysing the co-expression association between immune checkpoint genes and NCBP1 in TCGA datasets, our work elucidated that NCBP1 and various immune checkpoint genes have a positive correlation in BC. Furthermore, there exists a correlation between the expression of NCBP1 and the infiltration of immune cells in adrenal cancer and osteosarcoma. Hence, despite NCBP1's function as a cap-binding

complex facilitating molecular into the cell matrix, our hypothesis assumes that NCBP1 might potentially have a significant influence on immunological dysfunction in BC. In this work, an investigation was conducted to analyse the biological function of NCBP1 using GO analysis. The results revealed that NCBP1 plays a crucial role in inflammatory activities and immunological responses in BC. Furthermore, the utilisation of both the CIBERSORT algorithm and ssGSEA algorithm revealed that the overexpression of NCBP1 was associated with increased infiltration of naive B cells, NK cells, Macrophages M1, mast cells, regulatory T cells, and various subtypes of activated memory CD4⁺ and CD8 T cells. The increased infiltration level of immune cells in BC induces tumour cells to upregulate many immunological checkpoints as a defence mechanism against the immune system. Hence, a strong upregulation of NCBP1 was shown to have a significant association with the infiltration of immune cells and was indicative of a reduced OS in patients diagnosed with BC.

This study possesses several limitations. First, the established prognostic model was developed and validated using historical datasets, necessitating stronger evidence to ascertain its clinical value. Additionally, the inherent limitation of relying on a single predictor marker to construct the prognostic model was unavoidable, potentially resulting in the exclusion of crucial prognostic genes in BC. Furthermore, it is important to highlight that the relationship between the risk score and immunological activity remains unexplored experimentally. In conclusion, the research identified a new predictive model based on four m7G-related genes. This model was shown to be independently linked with OS, offering insight into BC prognosis prediction. The fundamental processes underlying the relationship between m7G-related genes and tumour immunity in BC are yet unknown and deserve additional exploration.

AUTHOR CONTRIBUTIONS

Wenqi Bai: Conceptualization (equal). **Jianrong Li:** Methodology (equal). **Lin Zheng:** Data curation (equal). **Liying Song:** Formal analysis (equal). **Zhuanxia Dong:** Project administration (equal). **Liqiang Qi:** Funding acquisition (equal).

ACKNOWLEDGEMENTS

The authors would like to thank Mr Guo Qiang for his help, understanding and support of our research. Moreover, this work was supported by a grant from the Shanxi Provincial Health Commission (2023069).

CONFLICT OF INTEREST STATEMENT

The authors confirm that there are no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available in article supplementary material.

ORCID

Lin Zheng  <https://orcid.org/0000-0001-7771-5483>

Wenqi Bai  <https://orcid.org/0009-0003-9421-2513>

REFERENCES

- Kanematsu M, Futamura M, Takata M, et al. Clinical significance of glycoprotein nonmetastatic B and its association with HER2 in breast cancer. *Cancer Med*. 2015;4(9):1344-1355.
- Lin HY, Liang YK, Dou XW, et al. Notch3 inhibits epithelial-mesenchymal transition in breast cancer via a novel mechanism, upregulation of GATA-3 expression. *Oncogene*. 2018;7(8):59.
- Choong GM, Cullen GD, O'Sullivan CC. Evolving standards of care and new challenges in the management of HER2-positive breast cancer. *CA Cancer J Clin*. 2020;70(5):355-374.
- Costa R, Carneiro BA, Wainwright DA, et al. Developmental therapeutics for patients with breast cancer and central nervous system metastasis: current landscape and future perspectives. *Ann Oncol*. 2017;28(1):44-56.
- Liang Y, Zhou X, Xie Q, et al. CD146 interaction with integrin β 1 activates LATS1-YAP signaling and induces radiation-resistance in breast cancer cells. *Cancer Lett*. 2022;546:215856.
- Chen B, Jiang W, Huang Y, et al. N7-methylguanosine tRNA modification promotes tumorigenesis and chemoresistance through WNT/ β -catenin pathway in nasopharyngeal carcinoma. *Oncogene*. 2022;41(15):2239-2253.
- Ying X, Liu B, Yuan Z, et al. METTL1-m7G-EGFR/EFEMP1 axis promotes the bladder cancer development. *Clin Transl Med*. 2021;11(12):e675.
- Ma J, Han H, Huang Y, et al. METTL1/WDR4-mediated m7G tRNA modifications and m7G codon usage promote mRNA translation and lung cancer progression. *Mol Ther*. 2021;29(12):3422-3435.
- Rong D, Sun G, Wu F, et al. Epigenetics: roles and therapeutic implications of non-coding RNA modifications in human cancers. *Mol Ther Nucleic Acids*. 2021;25:67-82.
- Kozlova NV, Pichon C, Rahmouni AR. mRNA with mammalian codon bias accumulates in yeast mutants with constitutive stress granules. *Int J Mol Sci*. 2020;21(4):1234.
- Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
- Tomikawa C. 7-Methylguanosine modifications in transfer RNA (tRNA). *Int J Mol Sci*. 2018;19(12):4080.
- Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-287.
- Ouyang W, Jiang Y, Bu S, et al. A prognostic risk score based on hypoxia-, immunity-, and epithelial-to-mesenchymal transition-related genes for the prognosis and immunotherapy response of lung adenocarcinoma. *Front Cell Dev Biol*. 2021;9:758777.
- Zhang L, Liu C, Zhang X, Wang C, Liu D. Breast cancer prognosis and immunological characteristics are predicted using the m6A/m5C/m1A/m7G-related long noncoding RNA signature. *Funct Integr Genomics*. 2023;23(2):117.
- Zhang W, Zhang S, Wang Z. Prognostic value of 12 m7G methylation-related miRNA markers and their correlation with immune infiltration in breast cancer. *Front Oncol*. 2022;12:929363.
- Gebhardt A, Habjan M, Benda C, et al. mRNA export through an additional cap-binding complex consisting of NCBP1 and NCBP3. *Nat Commun*. 2015;6:8192.
- Goldson TM, Vielhauer G, Staub E, Miller S, Shim H, Hagedorn CH. Eukaryotic initiation factor 4E variants alter the morphology, proliferation, and colony-formation properties of MDA-MB-435 cancer cells. *Mol Carcinog*. 2007;46(1):71-84.
- Dou Y, Barbosa I, Jiang H, et al. NCBP3 positively impacts mRNA biogenesis. *Nucleic Acids Res*. 2020;48(18):10413-10427.
- Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj JW. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*. 1994;78(4):657-668.
- Izaurralde E, Lewis J, Gamberi C, Jarmolowski A, McGuigan C, Mattaj JW. A cap-binding protein complex mediating U snRNA export. *Nature*. 1995;376(6542):709-712.
- Meng S, Xia Y, Li M, et al. NCBP1 enhanced proliferation of DLBCL cells via METTL3-mediated m6A modification of c-Myc. *Sci Rep*. 2023;13(1):8606.
- Zhang H, Wang A, Tan Y, et al. NCBP1 promotes the development of lung adenocarcinoma through up-regulation of CUL4B. *J Cell Mol Med*. 2019;23(10):6965-6977.
- Weiss B, Allen GE, Kloehn J, Abid K, Jaquier-Gubler P, Curran JA. eIF4E3 forms an active eIF4F complex during stresses (eIF4FS) targeting mTOR and re-programs the translome. *Nucleic Acids Res*. 2021;49(9):5159-5176.
- Piccirillo CA, Bjur E, Topisirovic I, Sonenberg N, Larsson O. Translational control of immune responses: from transcripts to translomes. *Nat Immunol*. 2014;15(6):503-511.
- Osborne MJ, Volpon L, Kornblatt JA, Culjkovic-Kraljacic B, Baguet A, Borden KLB. eIF4E3 acts as a tumor suppressor by utilizing an

- atypical mode of methyl-7-guanosine cap recognition. *Proc Natl Acad Sci U S A*. 2013;110(10):3877-3882.
28. Wang K, Li L, Fu L, et al. Integrated bioinformatics analysis the function of RNA binding proteins (RBPs) and their prognostic value in breast cancer. *Front Pharmacol*. 2019;10:140.
 29. Li F, Sun H, Li Y, et al. High expression of eIF4E is associated with tumor macrophage infiltration and leads to poor prognosis in breast cancer. *BMC Cancer*. 2021;21(1):1305.
 30. Xu X, Zhao Y, Ying Y, et al. m7G-related genes-NCBP2 and EIF4E3 determine immune contexture in head and neck squamous cell carcinoma by regulating CCL4/CCL5 expression. *Mol Carcinog*. 2023;62(8):1091-1106.
 31. Zhu WL, Zeng H, Huang DP, Ouyang WJ, Wei CS, Tong GD. The potential impact of EIF4E3 and LARP1 on tumor immunity in hepatocellular carcinoma. *Biomed Environ Sci*. 2023;36(5):469-475.
 32. Zhang Y, Gan W, Ru N, et al. Comprehensive multi-omics analysis reveals m7G-related signature for evaluating prognosis and immunotherapy efficacy in osteosarcoma. *J Bone Oncol*. 2023;40:100481.
 33. Deng J, Lin J, Liu C, et al. N7-methylguanosine methylation-related regulator genes as biological markers in predicting prognosis for melanoma. *Sci Rep*. 2022;12(1):21082.
 34. Michaud J, Kudoh J, Berry A, et al. Isolation and characterization of a human chromosome 21q22.3 gene (WDR4) and its mouse homologue that code for a WD-repeat protein. *Genomics*. 2000;68(1):71-79.
 35. Rastegari E, Kajaal K, Tan BS, et al. WD40 protein Wuho controls germline homeostasis via TRIM-NHL tumor suppressor Mei-p26 in *Drosophila*. *Development*. 2020;147(2):dev182063.
 36. Zeng H, Xu S, Xia E, Hirachan S, Bhandari A, Shen Y. Aberrant expression of WDR4 affects the clinical significance of cancer immunity in pan-cancer. *Aging*. 2021;13(14):18360-18375.
 37. Kentsis A, Volpon L, Topisirovic I, et al. Further evidence that ribavirin interacts with eIF4E. *RNA*. 2005;11(12):1762-1766.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Li J, Zheng L, Song L, Dong Z, Bai W, Qi L. Identification and validation of *N*⁷-methylguanosine-associated gene NCBP1 as prognostic and immune-associated biomarkers in breast cancer patients. *J Cell Mol Med*. 2024;28:e18067. doi:[10.1111/jcmm.18067](https://doi.org/10.1111/jcmm.18067)