RESEARCH

Construction and validation of a prognostic risk model for breast cancer based on protein expression

Bo Huang[1†], Xujun Zhang[2†], Qingyi Cao[2], Jianing Chen[2], Chenhong Lin[3], Tianxin Xiang[4] and Ping Zeng[4*]

Abstract

Breast cancer (BRCA) is the primary cause of mortality among females globally. The combination of advanced genomic analysis with proteomics characterization to construct a protein prognostic model will help to screen effective biomarkers and find new therapeutic directions. This study obtained proteomics data from The Cancer Proteome Atlas (TCPA) dataset and clinical data from The Cancer Genome Atlas (TCGA) dataset. Kaplan–Meier and Cox regression analyses were used to construct a prognostic risk model, which was consisted of 6 proteins (CASPASE-7CLEAVEDD198, NFKBP65-pS536, PCADHERIN, P27, X4EBP1-pT70, and EIF4G). Based on risk curves, survival curves, receiver operating characteristic curves, and independent prognostic analysis, the protein prognostic model could be viewed as an independent factor to accurately predict the survival time of BRCA patients. We further validated that this prognostic model had good predictive performance in the GSE88770 dataset. The expression of 6 proteins was significantly associated with the overall survival of BRCA patients. The 6 proteins and encoding genes were differentially expressed in normal and primary tumor tissues and in different BRCA stages. In addition, we verified the expression of 3 differential proteins by immunohistochemistry and found that CDH3 and EIF4G1 were significantly higher in breast cancer tissues. Functional enrichment analysis indicated that the 6 genes were mainly related to the HIF-1 signaling pathway and the PI3K-AKT signaling pathway. This study suggested that the prognosis-related proteins might serve as new biomarkers for BRCA diagnosis, and that the risk model could be used to predict the prognosis of BRCA patients.

Keywords: Breast cancer, Proteomics, TCPA, TCGA, Prognostic risk model

---

[†]Bo Huang and Xujun Zhang have contributed equally to this work and share authorship

*Correspondence: ndyfy08004@ncu.edu.cn

[4] Department of Hospital Infection Control, The First Affiliated Hospital of Nanchang University, 17 Yongwai Road, Donghu District, Nanchang, China
Full list of author information is available at the end of the article

Huang *et al. BMC Medical Genomics*   (2022) 15:148

Page 2 of 14

Introduction

Breast cancer is the most common malignant tumor in women and ranks as the leading cause of cancer-related death in women, accounting for 15.3% of all cancer deaths among females. In 2020, according to the latest data released by the International Agency for Research on Cancer (IARC) of the World Health Organization (WHO) (http://www.irac.fr), there were 19.29 million new cancer patients, of which 2.26 million were breast cancer, replacing lung cancer as the leading cause of cancer [1]. One of the underlying causes of the increased incidence of breast cancer is the changing risk factors, such as delayed and reduced childbearing, overweight and obesity, lifestyle, and heredity. The prognosis of patients with breast cancer depends on the tumor stage. The five-year survival rate for early breast cancer is close to 100%, while the survival rate for advanced-stage patients remains poor [2]. Therefore, the early screening and diagnosis of breast cancer are critical. As the gold standard for the early detection of breast cancer, the screening modality of mammography is sensitively for fatty breast tissue and contributes to significantly reduced mortality [3]. However, mammography is not particularly useful for dense breast tissue. Studies indicate that women with dense tissue have a 4- to 6-fold greater risk of developing breast cancer than those with no dense tissue in the breast [4, 5]. The automated breast ultrasound (ABUS) screening system designed for dense breast tissue perfectly addressed the limitations of mammography [6]. Although breast cancer screening can assist diagnosis and treatment, yet it has several limitations. A frequent limitation is a false-positive screening outcome, leading to overdiagnosis, which may cause distress and anxiety in women [7]. In addition, screening cannot detect all breast cancers. Therefore, it is very important to search for more accurate and sensitive diagnostic and prognostic biomarkers.

To date, some proteins and RNAs have been proved to be prognostic factors for cancer. Common serum protein tumor markers, such as CEA, AFP, CA-125, and CA-199, have been widely used for the diagnosis and treatment of various cancer types. Carcinoembryonic antigen (CEA) originates from the cavitary organs and has relatively high sensitivity for adenocarcinomas [8]. CEA is the most extensively used tumor marker, including in lung adenocarcinoma, colorectal cancer, and gastric cancer [9–11]. Serum α-fetoprotein (AFP) is the most classic diagnostic marker in hepatocellular carcinoma (HCC) [12]. Cancer antigen 125 (CA125) is a traditional marker for ovarian cancer screening [13] and is insufficient to diagnose ovarian cancer due to a lack of specificity [14]. Carbohydrate antigen 199 (CA199) is the most sensitive tumor marker for pancreatic cancer [15]. Squamous cell carcinoma antigen (SCC-Ag) is the most common early tumor marker for cervical cancer [16]. In addition, a growing number of studies have reported the potential of different RNAs as diagnostic biomarkers in cancer. The long noncoding RNA CCEPR (cervical carcinoma expressed PCNA regulatory) is a potential prognostic biomarker and therapeutic target [17]. Plasma miR-21 and miR-222 were increased in patients with gastric cancer and could be potential diagnostic biomarkers of gastric cancer [18]. miR1246 is a biomarker in esophageal and gastric cancers [19].

In human breast cancer, cathepsin D (CTSD), IL4 receptor (IL4R), mucin-1 (MUC1, CD227), and serine protease 3 (PRSS3) may serve as valuable biomarkers for the diagnosis and treatment of breast cancer [20–23]. Breast cancer is divided into different types based on pathology features. A single protein or RNA is unlikely to fulfill the criteria for identifying all types of breast cancer, resulting in missed diagnosis and overdiagnosis. Therefore, prognostic models based on multiple protein biomarkers have great potential for the diagnosis and prognosis of patients with breast cancer.

In this study, we downloaded the protein expression profile of breast invasive carcinoma (BRCA) from the Cancer Proteome Atlas (TCPA) database, and based on the Kaplan–Meier method and Cox regression analysis, we constructed a prognostic risk model for BRCA.

Materials and methods

Protein data acquisition and processing

The Cancer Proteome Atlas (TCPA) database provides the protein expression profiles for a variety of human cancers by integrating reversed-phase protein array (RPPA) chip data from The Cancer Genome Atlas (TCGA) database and several independent tumor research projects [24]. The TCPA database contains two separate web applications, one of which is the RPPA data for the patients with tumors, containing approximately 8000 cancer samples from 32 cancer types in the TCGA database, and another approximately 500 samples from an independent patient cohort. We downloaded the level 4 dataset of the BRCA protein expression profile from the TCPA database (http://www.tcpaportal.org/tcpa/) [25]. Matching clinical information of BRCA patients was downloaded from the TCGA database (http://portal.gdc.cancer.gov/), which is the currently largest cancer genetic information database. The missing data of the protein expression profiles were filled in with the "impute" package in the R software. Survival information was extracted from the clinical data and merged with the protein expression profile using Perl software.

Huang *et al. BMC Medical Genomics*   (2022) 15:148

Page 3 of 14

Screening the prognosis-related proteins

Kaplan−Meier (KM) survival analysis and univariate Cox analysis were used to screen the prognosis-related protein by the "survival" package in R software [26]. The patients were divided into two groups according to the protein expression and the KM method was used to analyze the difference of the two groups (p value < 0.05). The COX method compared protein expression as a continuous variable with survival time to observe whether there was a correlation (p value < 0.05). R packages "ggplot2" and "ggrepel" were used to draw the volcano map of differently expressed proteins [27]. Proteins with Hazard ratio (HR) < 1 were considered as low-risk proteins, while proteins with HR > 1 were considered high-risk proteins.

Constructing the protein prognostic risk model

The proteins with statistical difference were analyzed by multivariate Cox analysis and to build a prognostic risk model with the "survival" package in R software [28]. The coefficients of each protein in the model and the risk values of all samples were obtained. The risk score formula was defined as follows: risk score = (Coefficient Protein1 × expression of Protein1) + (Coefficient Protein2 × expression of Protein2) + ⋯ + (Coefficient Protein$n$ × expression of Protein$n$) [29]. Patients were divided into high risk group and low risk group according to the median of the risk value.

Assessing the protein prognostic risk model

The survival analysis of protein expression and risk values was conducted by the "survminer" and "survival" packages in R software. Risk curves, including the risk score, survival time, and protein expression, were performed using the "pheatmap" package. An independent prognostic analysis was using Perl software and the "survival" package based on the risk score and clinical data of each sample. The age, pathology stage, tumor (T), metastasis (M), node (N) states, and riskScore of the sample were considered by univariate and multivariate Cox analyses. The receiver operating characteristic (ROC) curve was analyzed by the "survivalROC" package in R software.

Validating the protein prognostic risk model

Microarray datasets, including gene expression profiles and corresponding clinical information data of GSE88770, were downloaded from the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/). GSE88770 was conducted by GPL570 (Affymetrix Human Genome U133 Plus 2.0 Array), including 117 breast cancer samples that were enrolled in our testing dataset. The expression profiles of mRNAs from GEO are

shown as raw data and each mRNA was normalized by log2 transformation for further analysis.

Expression of prognosis-related proteins and their encoding genes

The UALCAN portal (http://ualcan.path.uab.edu/analysis-prot.html) analyzed the expression of prognosis-related proteins and their encoding genes using the Clinical Proteomic Tumor Analysis Consortium (CPTAC) dataset (http://proteomics.cancer.gov/programs/cptac) and TCGA database. The immunohistochemical images of these proteins were obtained from the Human Protein Atlas (https://www.proteinatlas.org/).

Functional enrichment analysis

The functional enrichment analysis was constructed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/home.jsp) including Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis [30–34]. We used the bioinformatics online tool (http://www.bioinformatics.com.cn) to display the result.

Immunohistochemistry

Immunohistochemistry was performed as standard protocols. Tissue sections of tumor and para-tumor tissues underwent dewaxing, hydration, antigen repair, and blocking, the sections were incubated overnight at 4 °C in primary antibody (anti-CDH3, ab242060, Abcam; anti-CASP7, ab255818, Abcam; anti-EIF4G1, 15704-1-AP, Proteintech.). After rinsing in phosphate-buffered saline (PBS) three times, the slides were incubated with a secondary antibody and the streptavidin−horseradish peroxidase in turn. At last, the slides were stained with DAB solution. Images were obtained using a direct optical microscope.

Statistical analysis

R software (version 3.5.1) or Perl software (Strawberry Perl 5.30.0.1 64-bit) was performed analyses and chart visualize in this study. Statistical analysis of GSE88770 dataset was performed by using GraphPad Prism version 8.0 or SPSS version 19.0 software package. A two-tailed $p < 0.05$ was considered statistically significant.

Results

Screening prognostic-related proteins and constructing a prognostic risk model

The major work of this study is shown in Fig. 1. The protein expression profile of BRCA from the TCPA database and clinical data from the TCGA database were processed
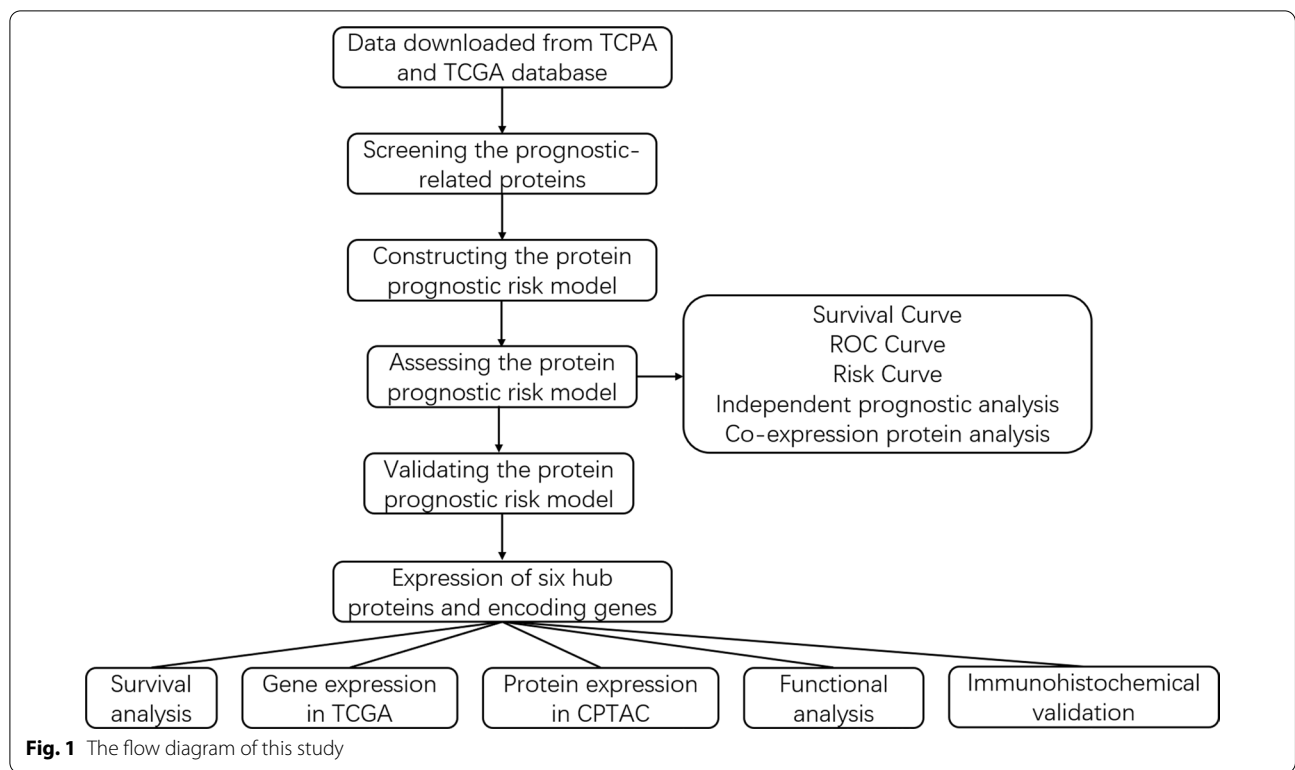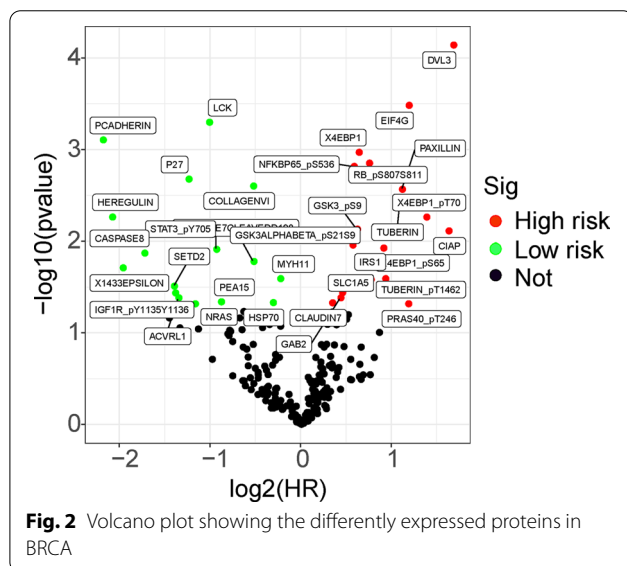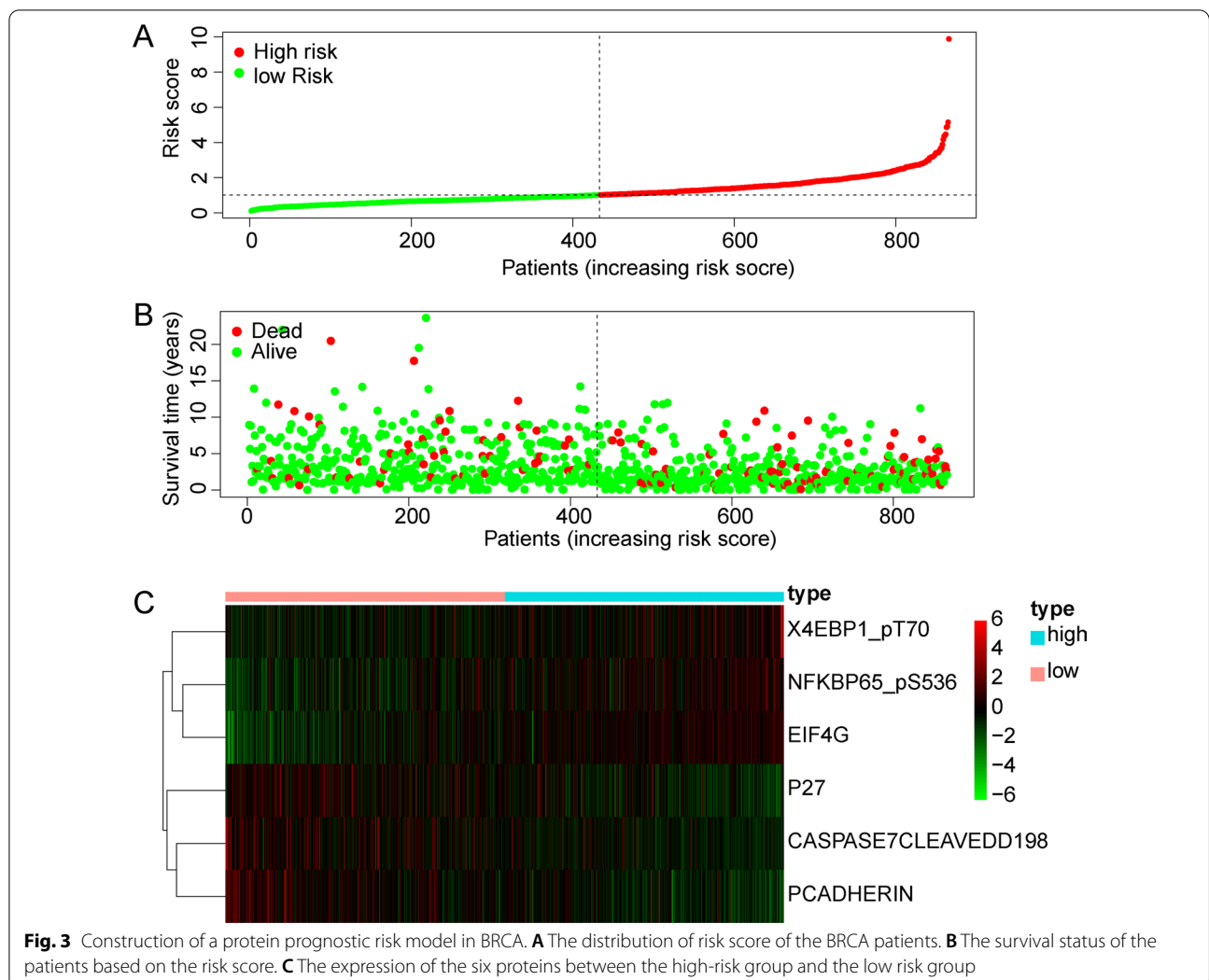
Fig. 1 The flow diagram of this study

Fig. 2 Volcano plot showing the differently expressed proteins in BRCA

and analyzed. Univariate Cox regression analysis and the KM method identified 34 proteins that were significantly associated with the survival of BRCA patients, including 18 high risk proteins (hazard ratio > 1) and 16 low-risk proteins (hazard ratio < 1). All the significantly expressed proteins are displayed in a volcano plot (Fig. 2, Additional file 1: Table S1). Diahevelled-3 (DVL3, HR = 3.206, 95%

CI = 1.803–5.702) was the highest-risk protein, while placental-cadherin (PCADHERIN, HR = 0.220, 95% CI = 0.091–0.533) was the lowest-risk protein. Multivariate Cox regression analysis demonstrated that six proteins were used to construct the prognostic model, including CASPASE7CLEAVEDD198, NFKBP65-pS536, PCADHERIN, P27, X4EBP1-pT70, and EIF4G. The formula for the risk model was as follows: Risk Score = CASPASE7CLEAVEDD198 * (− 0.406) + NFKBP65-pS536 * 0.242 + PCADHERIN * (− 1.157) + P27 * (− 0.640) + X4EBP1-pT70 * 0.710 + EIF4G * 0.534. These proteins were expressed in all samples and risk values for all patients were calculated based on this formula.
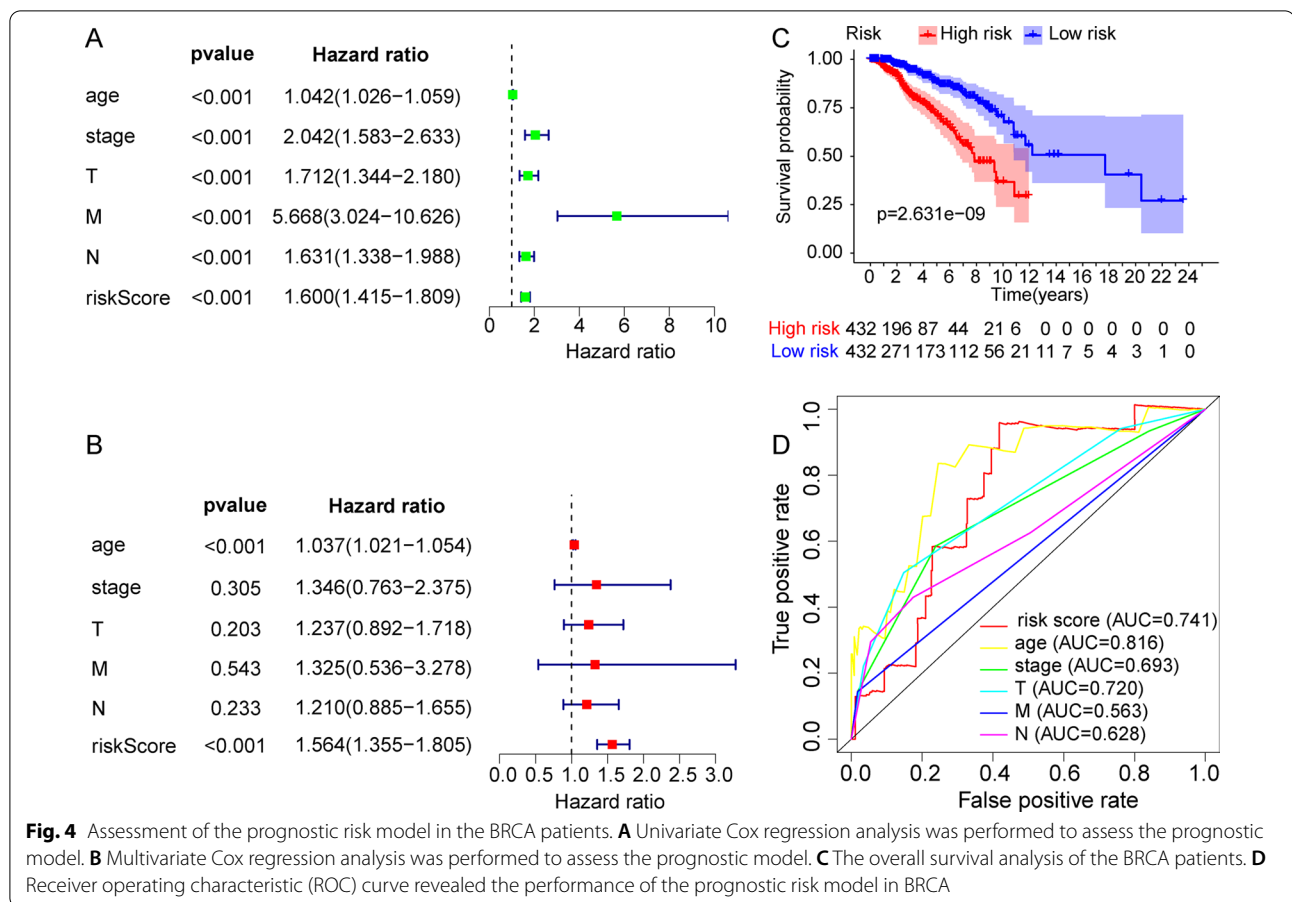
Assessing of the protein prognostic risk model
To further develop a comprehensive prognostic risk model, we built a risk curve, including risk score, survival time, and protein expression. The risk score distribution of the BRCA patients is shown in Fig. 3A. The higher the risk score, the worse the survival state (Fig. 3B). The heatmap displays the expression profiles of the 6 proteins in the high-risk and low-risk groups. EIF4G (EIF4G1), NFKBP65-pS536 (RELA), and X4EBP1-pT70 (EIF4EBP1) were highly expressed in the high-risk score group, while CASPASE7CLEAVEDD198 (CASP7), P27 (CDKN1B),

Huang *et al. BMC Medical Genomics* (2022) 15:148

Page 5 of 14

Fig. 3 Construction of a protein prognostic risk model in BRCA. **A** The distribution of risk score of the BRCA patients. **B** The survival status of the patients based on the risk score. **C** The expression of the six proteins between the high-risk group and the low risk group

and PCADHERIN (CDH3) were highly expressed in the low-risk score group (Fig. 3C). To evaluate whether the prognostic model was independent of other clinical factors, we performed univariate and multivariate Cox regression analyses. We found that clinical traits were associated with survival time and survival state, including age, pathology, T stage, M stage, N stage, and risk score, by univariate Cox regression analysis. The clinical characteristics of BRCA patients were shown in Additional file 1: Table S2. Univariate Cox analysis demonstrated that the risk score (HR 1.600 (1.415–1.809), $p < 0.001$), age (HR 1.042 (1.026–1.059), $p < 0.001$), stage (HR 2.042 (1.583–2.633), $p < 0.001$), T (HR 1.712 (1.344–2.180), $p < 0.001$), M (HR 5.668 (3.024–10.626), $p < 0.001$), N (HR 1.631(1.338–1.988), $p < 0.001$). Multivariate Cox regression analysis again suggested that risk score (HR 1.564 (1.355–1.805), $p < 0.001$) and age (HR 1.037 (1.021–1.054), $p < 0.001$) could be viewed as independent

prognostic factors for BRCA patients (Fig. 4A, B). In addition, we divided all samples into high- and low-risk groups according to the median risk value of each sample. Overall survival analysis showed that survival probability and survival time were significantly decreased in the high-risk group compared with the low risk group (Fig. 4C). The ROC curve could evaluate the accuracy of the prognostic model in predicting the survival time of the BRCA patients. The area under the curve (AUC) of the risk score was 0.741, suggesting that the predictive effectiveness was sensitive and significant for the prognostic risk model (Fig. 4D). These results suggested that this prognostic model could be viewed as an independent factor to accurately predict the survival time of BRCA patients.

Huang *et al. BMC Medical Genomics*   (2022) 15:148

Page 6 of 14

Fig. 4 Assessment of the prognostic risk model in the BRCA patients. **A** Univariate Cox regression analysis was performed to assess the prognostic model. **B** Multivariate Cox regression analysis was performed to assess the prognostic model. **C** The overall survival analysis of the BRCA patients. **D** Receiver operating characteristic (ROC) curve revealed the performance of the prognostic risk model in BRCA

Validation of the protein prognostic risk model

To demonstrate the prognostic performance and a possible future application of this model, we downloaded the GSE88770 dataset and validated the prognostic model in the microarray datasets. The clinical characteristics of BRCA patients were shown in Additional file 1: Table S3. The results showed that higher risk scores indicated a shorter survival time (Fig. 5A, B). Similarly, we found significantly higher survival rates in the low-risk group than in the high-risk group by overall survival analysis (Fig. 5C). The AUC of the risk score in the GSE88770 dataset was 0.712 indicating good predictive ability of the prognostic risk model (Fig. 5D). The above results validated that this prognostic model had good predictive performance and could be used to predict the risk of BRCA patients.
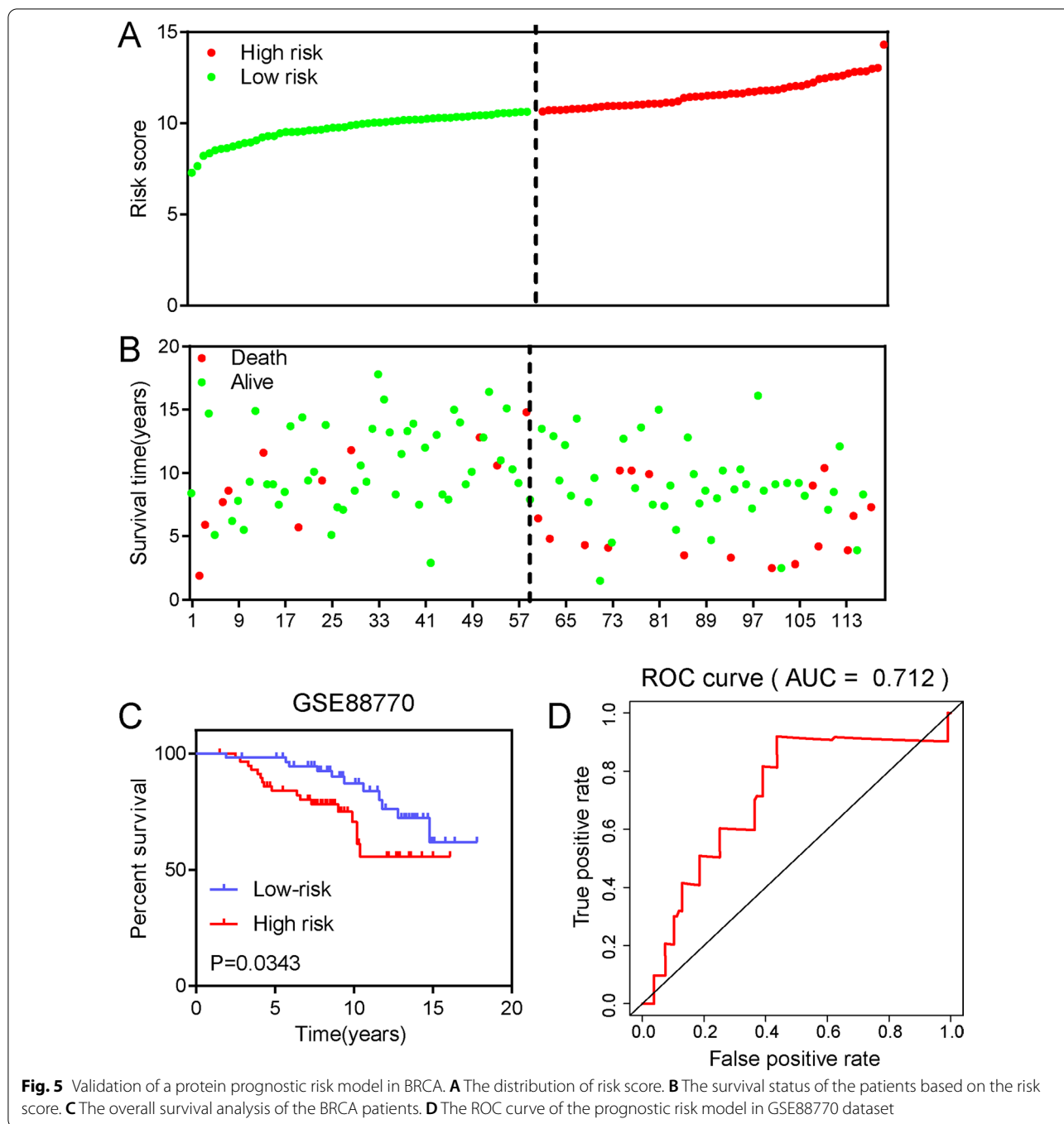
Survival analysis of the 6 proteins

To investigate the relationship between the expression of the 6 proteins and overall survival, we used the KM method to carry out survival analysis. According to the median value of protein expression, we divided all samples into high-expression and low-expression groups.

The higher the protein expression levels were for CASPASE7CLEAVEDD198, P27, and PCADHERIN, the better the overall survival (Fig. 6A–C). The higher the protein expression levels were for EIF4G, NFKBP65-pS536, and X4EBP1-pT70, the poorer the overall survival (Fig. 6D–F).

Differential expression analysis of the 6 proteins and their encoding genes
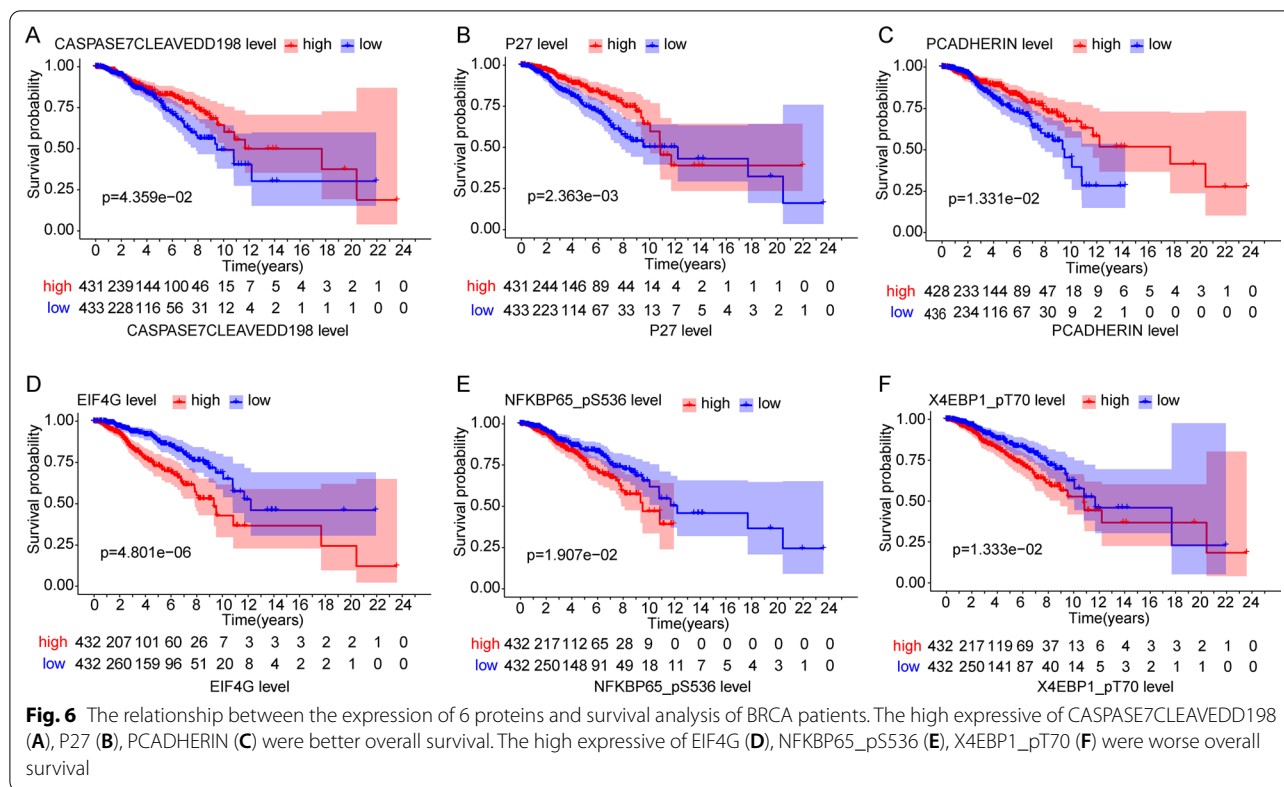
To better understand the 6 prognosis-related proteins, we explored the expression of these proteins using the CPTAC dataset. The CASPASE7CLVEADD198 (CASP7), PCADHERIN (CDH3), and EIF4G (EIF4G1) were expressed at significantly higher levels in tumor tissue than in normal tissue, while P27 (CDKN1B), NFKBP65-pS536 (RELA), and X4EBP1-pT70 (EIF4EBP1) showed no difference (Fig. 7A). Except for CDKN1B and RELA, the expression levels of CASP7, CDH3, EIF4G1, and EIF4EBP1 were increased significantly in most individual cancer stages compared with normal samples (Fig. 7B). Moreover, we evaluated immunohistochemical images of these proteins from the Human Protein Atlas database (Fig. 7C) and the information of patients

Fig. 5 Validation of a protein prognostic risk model in BRCA. **A** The distribution of risk score. **B** The survival status of the patients based on the risk score. **C** The overall survival analysis of the BRCA patients. **D** The ROC curve of the prognostic risk model in GSE88770 dataset

were shown in Additional file 1: Table S4. In addition, at the mRNA expression levels, there were differences in the expression of these encoding genes except RELA. Compared with normal samples, the expression level of CASP7, CDKN1B, EIG4G1, and EIF4EBP1 were significantly increased in the primary tumor samples, while that of CDH3 was significantly decreased (Fig. 8A). Similarly, except for RELA, the other encoding genes were

expressed differently in the most individual cancer stages (Fig. 8B).

To further confirm our analytical results, we used breast cancer tissues and paracancerous tissues to verify the expression of prognosis-related proteins with immunohistochemistry experiments. Pathology sections information was shown in Additional file 1: Tables S5 and S6. CDH3 and EIF4G1 were expressed at significantly higher

Huang *et al. BMC Medical Genomics*   (2022) 15:148

Page 8 of 14

Fig. 6 The relationship between the expression of 6 proteins and survival analysis of BRCA patients. The high expressive of CASPASE7CLEAVEDD198 (**A**), P27 (**B**), PCADHERIN (**C**) were better overall survival. The high expressive of EIF4G (**D**), NFKBP65_pS536 (**E**), X4EBP1_pT70 (**F**) were worse overall survival

levels in breast cancer tissues than in paracancerous tissues, but CASP7 was not (Fig. 9).

Functional enrichment analysis

To determine the functions of these encoding genes, we performed the functional enrichment analysis by DAVID. KEGG pathway analysis indicated that these genes were mainly related to the HIF-1 and PI3K-AKT signaling pathways (Fig. 10A). GO analysis suggested that these genes were mainly located in the cytoplasm, cytosol, and nucleoplasm, and participated in certain biological processes, such as response to drug and positive regulation of cell proliferation (Fig. 10B).
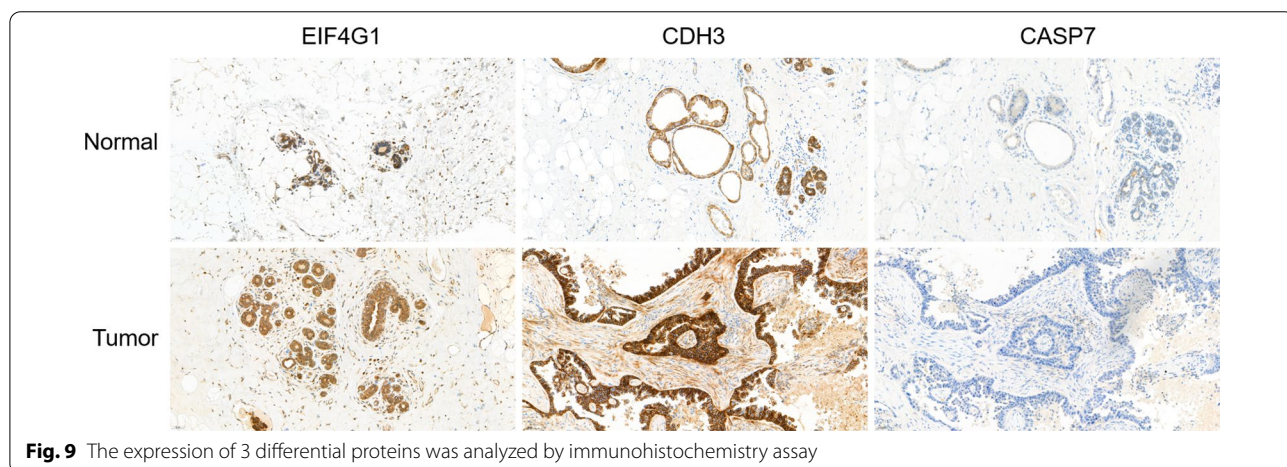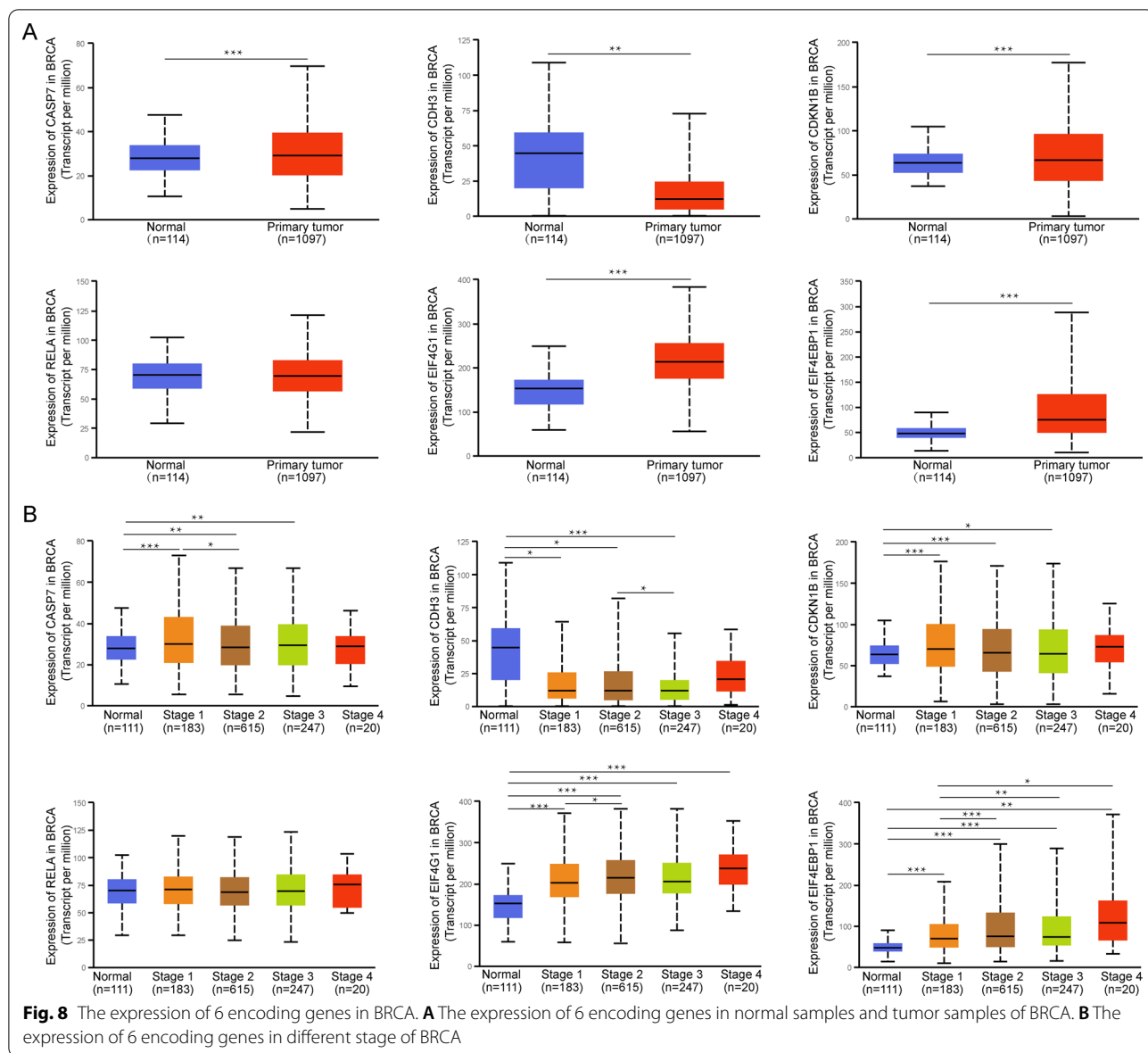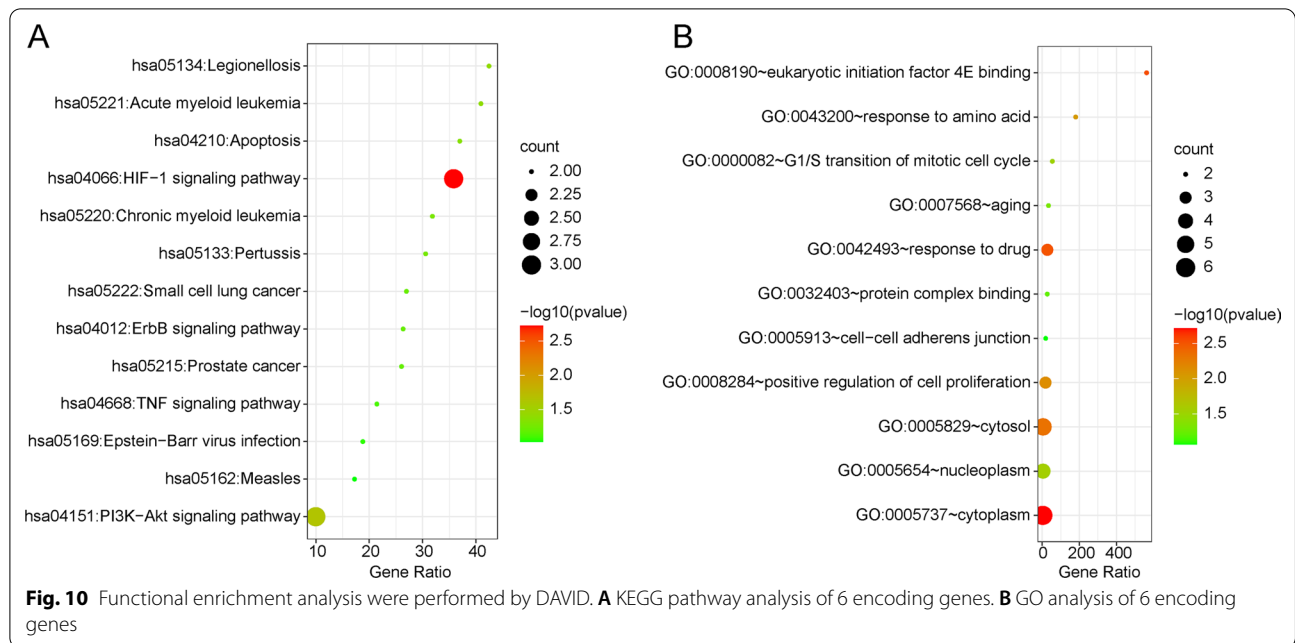
Discussion

Through combined analysis of the TCGA and TPCA datasets, cancer proteomics was used to study the pathogenesis and prognosis of tumors at the protein level and to explore candidate proteins that can be used as tumor biomarkers. This approach can provide new guidance for prognosis and personalized precision medicine of tumors. In this study, we obtained BRCA samples from the TCPA and TCGA datasets with clinical information and protein expression data. We screened six prognosis-related proteins using Kaplan–Meier and Cox analysis methods and further constructed a prognostic

risk model using multivariate Cox analysis. These six prognosis-related proteins were PCADHERIN, CASPASE7CLVEADD198, P27, EIF4G, NFKBP65-pS536, and X4EBP1-pT70, which were remarkably correlated with the overall survival of patients with BRCA. We assessed the effectiveness of the prognostic risk model in the GSE88770 dataset by survival analysis and ROC curve analyses. This study indicated that the prognostic risk model could serve as a sensitive independent prognostic factor and effectively predict the prognosis of BRCA patients.

Studies have shown that some of the prognosis-related proteins are involved in the prognosis of malignancy. Caspase-7 is the main executioner of mitochondrial apoptosis. Apoptosis maintains the homeostasis balance between cell proliferation and cell death. The dysregulation of apoptosis can lead to cancer [35]. As an indicator of apoptosis impairment, high cleaved caspase-7 levels indicate a decreased risk in BRCA and may be an alternative prognostic biomarker of clinical outcome [36]. In colorectal cancer (CRC) patients, the upregulation of caspase-7 caused by downregulation of sterol regulatory element binding proteins (SREBP1) may improve chemosensitivity to gemcitabine in CRC cells, which may serve as a novel prognostic biomarker for CRC [37]. This is basically consistent with the results

Fig. 7 The expression of 6 proteins in BRCA. **A** The expression of 6 proteins in normal samples and tumor samples of BRCA. **B** The expression of 6 proteins in different stage of BRCA. **C** 6 proteins expressions in the Human Protein Atlas database

Fig. 8 The expression of 6 encoding genes in BRCA. **A** The expression of 6 encoding genes in normal samples and tumor samples of BRCA. **B** The expression of 6 encoding genes in different stage of BRCA

Fig. 9 The expression of 3 differential proteins was analyzed by immunohistochemistry assay

Huang *et al. BMC Medical Genomics*   (2022) 15:148

Page 11 of 14

Fig. 10 Functional enrichment analysis were performed by DAVID. **A** KEGG pathway analysis of 6 encoding genes. **B** GO analysis of 6 encoding genes

of our studies. High expression of caspase-7 is associated with a good prognosis. Placental-cadherin (CDH3) is a cell adhesion molecule that has different roles in different tumors. CDH3 is highly expressed and promotes tumorigenesis in pancreatic cancer, gastric cancer, and breast cancer, while it is expressed at low levels and suppresses tumorigenesis in non-small-cell lung cancer, hepatocellular carcinoma and thyroid cancer [38–41]. P27 is an atypical tumor suppressor that regulates the cell cycle, cell migration and development and plays both tumor-suppressive and oncogenic roles [42, 43]. Deletion or mutational inactivation in the p27 gene is rare in human cancers. However, excessive proteolysis of p27 results in loss of growth restraint function in human cancers [44]. In PI3K/AKT activated cancers, C-terminally phosphorylated p27 is overexpressed in the cytoplasm and nucleus, and binds proteins to drive tumor progression, which shifts p27 from a cyclin-dependent kinase inhibitor to an oncogene [45]. In PI3K/AKT-activated human breast cancers, highly stable p27 accumulates in the cytoplasm and increases tumor metastasis, which is associated with poor patient outcome [46]. p27 is a key target of miR-221/222 in triple-negative breast cancer [47]. In our study, the expression of p27 was positively correlated with prognosis. NF-κB is a family of ubiquitous transcription factors that regulate DNA transcription, cytokine production, and cell survival [48]. The NF-κB signaling pathway plays an important role in inflammation and the immune response. Inflammatory cytokines can drive NF-κB activation within the tumor microenvironment.

In solid tumors, NF-κB is activated and promotes cancer cell growth and metastasis in breast, pancreatic, and colorectal cancers [49–51]. RELA/p65 is an important subunit of the NF-κB family, and Ser536 is an important phosphorylation site in RELA/p65. The phosphorylation of p65 at Ser536 was upregulated with the maturation and apoptotic shedding of epithelial cells in normal colon mucosa but was downregulated in colon cancer. In colon, breast, and prostate cancer cells, the RELA/p65 phosphomimetic mutation at Ser536 triggered dramatic apoptosis and suppressed tumor growth by affecting the expression of genes related to cell death or survival in nude mice [52]. In our study, the expression of RELA/p65 showed no significant difference between normal breast tissue and tumor breast tissue. The phosphorylation of p65 at Ser536 is a high-risk protein in BRCA. EIF4G1 is the major isoform of the EIF4G family and is a critical component of the eukaryotic initiation factor (EIF)4F complex. EIF4G1 is required for cap-dependent mRNA translation which is a necessary process for tumor growth and survival. Studies suggest that EIF4G1 is overexpressed in several solid tumors and plays an important role in the tumorigenesis. High expression of EIF4G1 has been found in various tumors and is associated with poor prognosis, including breast, lung, hypopharyngeal, and nasopharyngeal cancers [53–55], which is consistent with our findings. Recent studies suggest that high expression of EIF4G1 is associated with poor prognosis of pancreatic ductal adenocarcinoma and prostate cancer and may serve as a novel prognostic biomarker [56, 57]. Based on previous studies, these

prognosis-related proteins play an important role in various tumors. Our study further reveals the prognostic roles of these proteins in BRCA.

Conclusions

In summary, we constructed and assessed a prognostic risk model based on these proteins for BRCA patients. There are few protein databases that can be used publicly and contain relevant clinical information at present. Therefore, we chose GSE database to verify the prognostic model at the mRNA levels. The validate results showed that this prognostic model had good predictive performance and could be used to predict the risk of BRCA patients, which could explain the role of the protein prognosis model to a certain extent. The prognostic risk model will provide new insight into the diagnosis and prognosis of BRCA. In the future, we will continue to collect clinical samples and explore new databases for a more comprehensive validation.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12920-022-01299-5.

Additional file 1: Supplementary Table 1. Low-risk and high-risk proteins associated with the survival of BRCA patients. **Supplementary Table 2.** Clinical characteristics of BRCA patients included in this study. **Supplementary Table 3.** Clinical characteristics of BRCA patients in the GSE88770 dataset. **Supplementary Table 4.** The immunohistochemical images information of these proteins in the Human Protein Atlas database. **Supplementary Table 5.** Clinicopathological features of 20 BRCA patients. **Supplementary Table 6.** The EIF4G1, CDH3, CASP7 expression level between tumor and para-tumor tissues.

Availability of data and materials
Publicly available datasets were analyzed in this article. These data can be found here: The TCPA database (http://www.tcpaportal.org/tcpa/). The TCGA database (http://portal.gdc.cancer.gov/). The Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/).

Declarations

Ethics approval and consent to participate
Ethics approval was obtained from the ethics committee of the First Affiliated Hospital, Zhejiang University (Permit Number: 2021-029). The study was in accordance with the Declaration of Helsinki. All participants gave written informed consent to participate in the study.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Gynecology and Obstetrics, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [2]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases National Clinical Research Center for Infectious Diseases, National Medical Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China. [3]Department of Gastroenterology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China. [4]Department of Hospital Infection Control, The First Affiliated Hospital of Nanchang University, 17 Yongwai Road, Donghu District, Nanchang, China.

References
1. Global Cancer Observatory: Cancer Tomorrow. Accessed 10 Nov 2020.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69(1):7–34. https://doi.org/10.3322/caac.21551.
3. Al-Mousa DS, Alakhras M, Spuur KM, Alewaidat H, Abdelrahman M, Rawashdeh M, et al. The implications of increased mammographic breast density for breast screening in Jordan. J Med Radiat Sci. 2020;67(4):277–83. https://doi.org/10.1002/jmrs.414.
4. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomark Prev. 2006;15(6):1159–69. https://doi.org/10.1158/1055-9965.EPI-06-0034.
5. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. N Engl J Med. 2007;356(3):227–36. https://doi.org/10.1056/NEJMoa062790.
6. Wilczek B, Wilczek HE, Rasouliyan L, Leifland K. Adding 3D automated breast ultrasound to mammography screening in women with heterogeneously and extremely dense breasts: Report from a hospital-based, high-volume, single-center breast cancer screening program. Eur J Radiol. 2016;85(9):1554–63. https://doi.org/10.1016/j.ejrad.2016.06.004.
7. Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, et al. Systematic review of the psychological consequences of false-positive screening mammograms. Health Technol Assess. 2013;17(13):1–170. https://doi.org/10.3310/hta17130.
8. Thomas DS, Fourkala EO, Apostolidou S, Gunu R, Ryan A, Jacobs I, et al. Evaluation of serum CEA, CYFRA21-1 and CA125 for the early detection of colorectal cancer using longitudinal preclinical samples. Br J Cancer. 2015;113(2):268–74. https://doi.org/10.1038/bjc.2015.202.
9. Enz N, Fragoso F, Gamrekeli A, Lippek F, Jungraithmayr W. Carcinoembryonic antigen-positive pleural effusion in early stage non-small cell lung cancer without pleural infiltration. J Thorac Dis. 2018;10(5):E340–3. https://doi.org/10.21037/jtd.2018.04.111.
10. Hu PJ, Chen MY, Wu MS, Lin YC, Shih PH, Lai CH, et al. Clinical evaluation of CA72–4 for screening gastric cancer in A healthy population: a multicenter retrospective study. Cancers. 2019;11(5):733. https://doi.org/10.3390/cancers11050733.
11. Wang FH, Shen L, Li J, Zhou ZW, Liang H, Zhang XT, et al. The Chinese Society of Clinical Oncology (CSCO): clinical guidelines for the diagnosis and treatment of gastric cancer. Cancer Commun (Lond). 2019;39(1):10. https://doi.org/10.1186/s40880-019-0349-9.
12. Egashira Y, Suganuma M, Kataoka Y, Higa Y, Ide N, Morishita K, et al. Establishment and characterization of a fucosylated alpha-fetoprotein-specific monoclonal antibody: a potential application for clinical research. Sci Rep. 2019;9(1):12359. https://doi.org/10.1038/s41598-019-48821-x.

Huang *et al. BMC Medical Genomics*   (2022) 15:148

Page 13 of 14

13. Cui R, Wang Y, Li Y, Li Y. Clinical value of ROMA index in diagnosis of ovarian cancer: meta-analysis. Cancer Manag Res. 2019;11:2545–51. https://doi.org/10.2147/CMAR.S199400.

14. Pierredon S, Ribaux P, Tille JC, Petignat P, Cohen M. Comparative secretome of ovarian serous carcinoma: gelsolin in the spotlight. Oncol Lett. 2017;13(6):4965–73. https://doi.org/10.3892/ol.2017.6096.

15. Li Z, Wang J, Han X, Wang F, Hu H, Yuan J, et al. Association between cancer antigen 19–9 and diabetes risk: a prospective and Mendelian randomization study. J Diabetes Investig. 2020;11(3):585–93. https://doi.org/10.1111/jdi.13166.

16. Liu Z, Shi H. Prognostic role of squamous cell carcinoma antigen in cervical cancer: a meta-analysis. Dis Markers. 2019;2019:6710352. https://doi.org/10.1155/2019/6710352.

17. Zhan Y, Li Y, Guan B, Chen X, Chen Z, He A, et al. Increased expression of long non-coding RNA CCEPR is associated with poor prognosis and promotes tumorigenesis in urothelial bladder carcinoma. Oncotarget. 2017;8(27):44326–34. https://doi.org/10.18632/oncotarget.17872.

18. Emami SS, Nekouian R, Akbari A, Faraji A, Abbasi V, Agah S. Evaluation of circulating miR-21 and miR-222 as diagnostic biomarkers for gastric cancer. J Cancer Res Ther. 2019;15(1):115–9. https://doi.org/10.4103/jcrt.JCRT_592_17.

19. Wei C, Li Y, Huang K, Li G, He M. Exosomal miR-1246 in body fluids is a potential biomarker for gastrointestinal cancer. Biomark Med. 2018;12(10):1185–96. https://doi.org/10.2217/bmm-2017-0440.

20. Li R, Zhang Y, Zhu W, Ding C, Dai W, Su X, et al. Effects of olanzapine treatment on lipid profiles in patients with schizophrenia: a systematic review and meta-analysis. Sci Rep. 2020;10:17028. https://doi.org/10.1038/s41598-020-73983-4.

21. Yoo JD, Bae SM, Seo J, Jeon IS, Vadevoo SMP, Kim SY, et al. Designed ferritin nanocages displaying trimeric TRAIL and tumor-targeting peptides confer superior anti-tumor efficacy. Sci Rep. 2020;10(1):19997. https://doi.org/10.1038/s41598-020-77095-x.

22. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol. 2007;25(33):5287–312. https://doi.org/10.1200/JCO.2007.14.2364.

23. Qian L, Gao X, Huang H, Lu S, Cai Y, Hua Y, et al. PRSS3 is a prognostic marker in invasive ductal carcinoma of the breast. Oncotarget. 2017;8(13):21444–53. https://doi.org/10.18632/oncotarget.15590.

24. Wu P, Heins ZJ, Muller JT, Katsnelson L, de Bruijn I, Abeshouse AA, et al. Integration and analysis of CPTAC proteomics data in the context of cancer genomics in the cBioPortal. Mol Cell Proteom. 2019;18(9):1893–8. https://doi.org/10.1074/mcp.TIR119.001673.

25. Li J, Akbani R, Zhao W, Lu Y, Weinstein JN, Mills GB, et al. Explore, visualize, and analyze functional cancer proteomic data using the cancer proteome atlas. Cancer Res. 2017;77(21):e51–4. https://doi.org/10.1158/0008-5472.CAN-17-0369.

26. Wu M, Xia Y, Wang Y, Fan F, Li X, Song J, et al. Development and validation of an immune-related gene prognostic model for stomach adenocarcinoma. 2020. Biosci Rep. https://doi.org/10.1042/BSR20201012.

27. Saxena V, Gao H, Arregui S, Zollman A, Kamocka MM, Xuei X, et al. Kidney intercalated cells are phagocytic and acidify internalized uropathogenic Escherichia coli. Nat Commun. 2021;12(1):2405. https://doi.org/10.1038/s41467-021-22672-5.

28. Wang P, Wang Y, Hang B, Zou X, Mao JH. A novel gene expression-based prognostic scoring system to predict survival in gastric cancer. Oncotarget. 2016;7(34):55343–51. https://doi.org/10.18632/oncotarget.10533.

29. Wang Y, Liu X, Guan G, Zhao W, Zhuang M. A risk classification system with five-gene for survival prediction of glioblastoma patients. Front Neurol. 2019;10:745. https://doi.org/10.3389/fneur.2019.00745.

30. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44–57. https://doi.org/10.1038/nprot.2008.211.

31. da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37(1):1–13. https://doi.org/10.1093/nar/gkn923.

32. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27.

33. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947–51. https://doi.org/10.1002/pro.3715.

34. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545–51. https://doi.org/10.1093/nar/gkaa970.

35. Lee JH, Park SY, Hwang W, Sung JY, Cho ML, Shim J, et al. Isoharringtonine induces apoptosis of non-small cell lung cancer cells in tumorspheroids via the intrinsic pathway. Biomolecules. 2020;10(11):1521. https://doi.org/10.3390/biom10111521.

36. Lindner AU, Lucantoni F, Vareslija D, Resler A, Murphy BM, Gallagher WM, et al. Low cleaved caspase-7 levels indicate unfavourable outcome across all breast cancers. J Mol Med (Berl). 2018;96(10):1025–37. https://doi.org/10.1007/s00109-018-1675-0.

37. Shen W, Xu T, Chen D, Tan X. Targeting SREBP1 chemosensitizes colorectal cancer cells to gemcitabine by caspase-7 upregulation. Bioengineered. 2019;10(1):459–68. https://doi.org/10.1080/21655979.2019.1676485.

38. Conger RD, Martin MJ, Masarik AS, Widaman KF, Donnellan MB. Social and economic antecedents and consequences of adolescent aggressive personality: predictions from the interactionist model. Dev Psychopathol. 2015;27(4 Pt 1):1111–27. https://doi.org/10.1017/S0954579415000711.

39. Bauer R, Valletta D, Bauer K, Thasler WE, Hartmann A, Muller M, et al. Downregulation of P-cadherin expression in hepatocellular carcinoma induces tumorigenicity. Int J Clin Exp Pathol. 2014;7(9):6125–32.

40. Zhou Y, Chi Y, Bhandari A, Xia E, Thakur PC, Qu J, et al. Downregulated CDH3 decreases proliferation, migration, and invasion in thyroid cancer. Am J Transl Res. 2020;12(6):3057–67.

41. Liu P, Weng Y, Sui Z, Wu Y, Meng X, Wu M, et al. Quantitative secretomic analysis of pancreatic cancer cells in serum-containing conditioned medium. Sci Rep. 2016;6:37606. https://doi.org/10.1038/srep37606.

42. Choi HH, Guma S, Fang L, Phan L, Ivan C, Baggerly K, et al. Regulating the stability and localization of CDK inhibitor p27(Kip1) via CSN6-COP1 axis. Cell Cycle. 2015;14(14):2265–73. https://doi.org/10.1080/15384101.2015.1046655.

43. Zhao H, Faltermeier CM, Mendelsohn L, Porter PL, Clurman BE, Roberts JM. Mislocalization of p27 to the cytoplasm of breast cancer cells confers resistance to anti-HER2 targeted therapy. Oncotarget. 2014;5(24):12704–14. https://doi.org/10.18632/oncotarget.2871.

44. Chu IM, Hengst L, Slingerland JM. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. Nat Rev Cancer. 2008;8(4):253–67. https://doi.org/10.1038/nrc2347.

45. Razavipour SF, Harikumar KB, Slingerland JM. p27 as a transcriptional regulator: new roles in development and cancer. Cancer Res. 2020;80(17):3451–8. https://doi.org/10.1158/0008-5472.CAN-19-3663.

46. Wander SA, Zhao D, Besser AH, Hong F, Wei J, Ince TA, et al. PI3K/mTOR inhibition can impair tumor invasion and metastasis in vivo despite a lack of antiproliferative action in vitro: implications for targeted therapy. Breast Cancer Res Treat. 2013;138(2):369–81. https://doi.org/10.1007/s10549-012-2389-6.

47. Nassirpour R, Mehta PP, Baxi SM, Yin MJ. miR-221 promotes tumorigenesis in human triple negative breast cancer cells. PLoS ONE. 2013;8(4): e62170. https://doi.org/10.1371/journal.pone.0062170.

48. Kolati SR, Kasala ER, Bodduluru LN, Mahareddy JR, Uppulapu SK, Gogoi R, et al. BAY 11–7082 ameliorates diabetic nephropathy by attenuating hyperglycemia-mediated oxidative stress and renal inflammation via NF-kappaB pathway. Environ Toxicol Pharmacol. 2015;39(2):690–9. https://doi.org/10.1016/j.etap.2015.01.019.

49. Sakamoto K, Maeda S. Targeting NF-kappaB for colorectal cancer. Expert Opin Ther Targets. 2010;14(6):593–601. https://doi.org/10.1517/14728221003769903.

50. Mander S, Kim DH, Thi Nguyen H, Yong HJ, Pahk K, Kim EY, et al. SP-8356, a (1S)-(-)-verbenone derivative, exerts in vitro and in vivo anti-breast cancer effects by inhibiting NF-kappaB signaling. Sci Rep. 2019;9(1):6595. https://doi.org/10.1038/s41598-019-41224-y.

51. Yu C, Chen S, Guo Y, Sun C. Oncogenic TRIM31 confers gemcitabine resistance in pancreatic cancer via activating the NF-kappaB signaling pathway. Theranostics. 2018;8(12):3224–36. https://doi.org/10.7150/thno.23259.

52. Bu Y, Li X, He Y, Huang C, Shen Y, Cao Y, et al. A phosphomimetic mutant of RelA/p65 at Ser536 induces apoptosis and senescence: an implication for tumor-suppressive role of Ser536 phosphorylation. Int J Cancer. 2016;138(5):1186–98. https://doi.org/10.1002/ijc.29852.

53. Tu L, Liu Z, He X, He Y, Yang H, Jiang Q, et al. Over-expression of eukaryotic translation initiation factor 4 gamma 1 correlates with tumor progression

and poor prognosis in nasopharyngeal carcinoma. Mol Cancer. 2010;9:78. https://doi.org/10.1186/1476-4598-9-78.

54. Silvera D, Arju R, Darvishian F, Levine PH, Zolfaghari L, Goldberg J, et al. Essential role for eIF4GI overexpression in the pathogenesis of inflammatory breast cancer. Nat Cell Biol. 2009;11(7):903–8. https://doi.org/10.1038/ncb1900.

55. Cao Y, Wei M, Li B, Liu Y, Lu Y, Tang Z, et al. Functional role of eukaryotic translation initiation factor 4 gamma 1 (EIF4G1) in NSCLC. Oncotarget. 2016;7(17):24242–51. https://doi.org/10.18632/oncotarget.8168.

56. Goh TS, Ha M, Lee JS, Jeong DC, Jung ES, Han ME, et al. Prognostic significance of EIF4G1 in patients with pancreatic ductal adenocarcinoma. Onco Targets Ther. 2019;12:2853–9. https://doi.org/10.2147/OTT.S202101.

57. Jaiswal PK, Koul S, Shanmugam PST, Koul HK. Eukaryotic Translation Initiation Factor 4 Gamma 1 (eIF4G1) is upregulated during Prostate cancer progression and modulates cell growth and metastasis. Sci Rep. 2018;8(1):7459. https://doi.org/10.1038/s41598-018-25798-7.

Publisher's Note