Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Construction and Validation of a Prognostic Risk Model for Triple Negative Breast Cancer Based on Autophagy-Related Genes

Cheng Yan
Xinxiang University
Qingling Liu
Xinxiang University
Ruoling Jia (jiaruolinggg@163.com)
Xinxiang University

Research Article

Keywords: TNBC, Autophagy, Prognostic model, eIF4EBP1, Nomogram, ROC

Posted Date: December 9th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1141994/v1

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

Version of Record: A version of this preprint was published at Frontiers in Oncology on February 4th, 2022. See the published version at https://doi.org/10.3389/fonc.2022.829045.

Abstract

Background: Autophagy plays an important role in triple negative breast cancer (TNBC). However, the prognostic value of autophagy-related genes (ARGs) in TNBC remains unknown. In this study, we established a survival model to evaluate the prognosis of TNBC patients using ARGs signature.

Methods: A total of 222 autophagy-related genes were downloaded from The Human Autophagy Database. The RNA-sequencing data and corresponding clinical data of TNBC were obtained from the TCGA database. Differential gene expression of ARGs (DE-ARGs) between normal samples and TNBC samples was determined by the EdgeR software package. Then, univariate Cox, Lasso, and multivariate Cox regression analyses were performed. According to the Lasso regression results based on univariate Cox, we identified a prognostic signature for overall-survival (OS), which was further validated by using GEO cohort. We also found an independent prognostic marker that can predict the clinicopathological features of TNBC. Furthermore, a nomogram was drawn to predict the survival probability of TNBC patients, which could help in clinical decision for TNBC treatment. Finally, we validated the requirement of a ARG in our model for TNBC cell survival and metastasis.

Results: There are 43 differentially expressed ARGs (DE-ARGs) were identified between normal and tumor samples. A risk model for OS using CDKN1A, CTSD, CTSL, EIF4EBP1, TMEM74 and VAMP3 by Lasso regression analysis was established based on univariate Cox regression analysis. Overall survival of TNBC patients was significantly shorter in the high-risk group than in the low-risk group for both the training and validation cohorts. Using the Kaplan-Meier curves and ROC curves, we demonstrated the accuracy of the prognostic model. Multivariate Cox regression analysis was used to verify risk score as independent predictor. Then a nomogram was proposed to predict 1-, 3-, and 5-year survival for TNBC patients. The calibration curves showed great accuracy of the model for survival prediction. Finally, we found that depletion of EIF4EBP1, one of ARGs in our model, significantly reduced cell proliferation and metastasis of TNBC cells.

Conclusion: An autophagy-related prognosis model in TNBCs was constructed using ARGs signature containing CDKN1A, CTSD, CTSL, EIF4EBP1, TMEM74 and VAMP3. It could serve as an independent prognostic biomarker in TNBC.

Introduction

Breast cancer is the most leading diagnosed cancer among women with the fifth-highest cancer mortality worldwide in 2020[1]. Triple-negative breast cancer (TNBC) is a subtype of breast cancer, which is defined by lacking of estrogen (ER), progesterone (PR) and Her2 receptors. TNBC takes up approximately15-20% of total breast cancers and is the second leading cause of cancer death among women worldwide[2]. TNBC is characterized by high heterogeneity, early diagnosis difficulty, rapid metastasis, poor survival and high recurrence rate [3]. Statistics shows that the five years survival rate of TNBC patients is less than 40% after diagnosis[4]. Although various therapeutic approaches have been

introduced for TNBC, the incidences and recurrence ratios of TNBC still remain unsatisfactory, especially for developed countries [5]. TNM stage and molecular subtypes have been widely used in the routine diagnosis and treatment of TNBC. However, traditional markers have limited sensitivity and specificity to precisely predict prognosis and design individualized treatment in TNBC patients. Therefore, it is imperative to establish new molecular biomarkers and prognostic models to further improve effectiveness of treatment strategies for TNBC patients.

Autophagy is an important cellular catabolic process which maintain homeostasis by eliminating aggregated proteins and damaged organelles in eukaryotic cell[6]. Recent studies showed that autophagy plays a paradoxical tumor-suppressive or tumor-promoting role in different contexts and stages of progression: it prevents tumorigenesis in the early stage, but supports various tumor growth in late stage[7]. Increasing evidences indicated that autophagy has a high vital function in tumorigenesis and metastasis of TNBC. Mingyang Li et al. found that the expression of autophagy regulator ATG7 is significant lower in TNBC samples compared with normal tissue [8]. Overexpression of ATG7 inhibited cell proliferation and migration in MDA-MB-231 and BT-549 cell lines. However, autophagy may also promote TNBC migration and invasion by regulating YAP localization and ANKRD1 expression[9]. These findings have confirmed the importance of autophagy in TNBC and suggest that ARGs may serve as prognostic markers for TNBC. To our knowledge, there is no prognosis model of ARGs in TNBC has been constructed to predict the prognosis of TNBC patients. Therefore, a novel prognostic model with ARGs for predicting survival in TNBC is highly needed.

In this study, we analyzed the transcriptome and clinic data of TNBC obtained from TCGA database detailly and built the ARGs based prognosis model using Lasso regression. This model could provide an effective multi-dimensional biomarker strategy to perform a new individualized survival prediction and treatment plan selection of TNBC patients.

Methods

Data Acquisition

The Human Autophagy Database (HADb, http://www.autophagy.lu/index.html) can provide the entire set of human genes associated with autophagy[10]. We collected 222 ARGs from HADB. In addition, the RNA-sequencing and corresponding clinical data of triple negative breast cancer in TCGA were downloaded from the UCSC XENA database (https://xena.ucsc.edu/). The microarray and corresponding clinical data of GSE58812 were downloaded from Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/). TNBC samples were selected using negative for “breast carcinoma estrogen receptor status”, “breast carcinoma progesterone receptor status” and “lab procedure her2 neu in situ hybrid outcome type” as screening criteria. TNBC was screened according to the status of ER, PR, and HER2 based on immunohistochemical (IHC) classification. We obtained the expression profiles contain 162 normal samples and 103 TNBC samples from TCGA. Simultaneously, we also obtained 107 TNBC samples from GEO.

Identification of Differentially Expression ARGs

Differential gene expression of ARGs (DE-ARGs) in 162 normal samples and 103 TNBC samples was carried out by the DESeq2 package. We set false discovery rate (FDR) < 0.05 and |log2 fold change (FC) | > 1 as cut-off criteria to obtain DE-ARGs. Volcano plots of DE-ARGs were constructed with the OmicStudio tools (https://www.omicstudio.cn/tool), boxplots were plotted using the ggplot2 R package, Heatmap were obtained using Morpheus (https://software.broadinstitute.org/morpheu). Protein–protein interaction (PPI) networks were generated using the String website (https://string-db.org).

Functional Enrichment Analysis

We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) biological process enrichment of the DE-ARGs by R statistical software including packages of “clusterProfiler”, “org.Hs.eg.db”, “enrichplot”, “ggplot2”, and “GOplot”. An adjusted p-value of <0.05 was considered statistically significant. Moreover, Gene Set Enrichment Analysis (GSEA) of TCGA and GEO was conducted to reveal the signaling pathways and biological processes between high- and low-risk groups in TNBC patients (http://software.broadinstitute.org/gsea/).

Identification of Prognostic Gene Signatures

We used the 103 TNBC samples from TCGA cohort as a training set. Univariate Cox regression analysis was performed on ARGs of train set to identify the association between the expression levels of the genes and patients’ overall survival (OS) using the ‘survival’ package (http://bioconductor.org/packages/survival/) in R 3.6.1. The hazard ratio (HR) and p-value of each gene were calculated. Genes with p < 0.05 were screened for further analysis. Later, we further used Lasso Cox regression to reduce the number of genes and eliminate collinearity between genes. Finally, we performed multivariate Cox regression analysis based on univariate Cox regression.

Construction and Validation of a Prognostic Model

According to the results of Lasso Cox regression, the risk scores of all samples were calculated according to the equation:

$$\text{risk score} = \sum_{j=1}^{n} (Coef_j \times Xj)$$

Coef refers to the regression coefficient of ARGs in Lasso Cox regression analysis, “*X*” is the expression value of the gene, and “*n*” is the number of prognostic ARGs. Using the median risk score as threshold, patients were divided into the high-risk group and low-risk group. We used the R package “survival” to assess differences in OS and obtain Kaplan-Meier survival plots. The ROC curve was generated by the timeROC package to evaluate the prognostic ability of the model. Simultaneously, we used samples from GEO database as the validation set. We calculated risk scores for patients in the GEO cohort using the

same formula of the train set. Then, we performed univariate and multivariate Cox regression analyses to investigate whether risk score was an independent prognostic factor for OS in TNBC patients in train set. N, T, stage and risk score were used as covariates. T tests were used to test the correlation between risk score and clinicopathological factors. P value under 0.05 ($P < 0.05$) was considered statistically significant.

The Construction of Nomogram and Calibration Curves

A nomogram and calibration plots were generated by using the “rms” package in R software. The nomogram was used to investigate the level of consistency between the actual and predicted probabilities. Calibration plot was used to predict 1-, 3-, 5-year survival rates.

TNBC cell culture, proliferation, colony formation and migration assays

MDA-MB-231 and BT549 cells were cultured in DMEM medium. Two siRNAs were employed to knockdown eIF4EBP1 and the sequences were: siRNA1: 5'-GGGAGGTACCAGGATCATCTA-3', siRNA2: 5'-GGAGGTACCAGGATCATCTAT-3'. Cell proliferation of control and eIF4EBP1 knockdown cells was determined by CCK8 and Edu. Cells transfected with control or eIF4EBP1 siRNAs were seed in 6-well plates and colonies were measured by crystal violet staining after 15 days culture. Transwell and wound healing assays were performed and quantified using control and eIF4EBP1 knockdown cells to determine cell migration.

Results

Identification of DE-ARGs

Autophagy has been reported to contribute to TNBC progression. In this study, we want to construct a prognosis model using ARGs signature for TNBCs. The overall experimental design in this study was indicated as diagram (Figure 1). We first obtained the expression profiles contain 162 normal samples and 103 TNBC samples from TCGA database. A total of 43 differentially expressed ARGs (DE-ARGs) were identified by comparing normal and tumor samples with the cut off criteria of FDR< 0.05 and |log2 FC| >1. The volcano map (Figure 2.A), box plots (Figure 2.C), and heatmap (Figure 2.D) demonstrated that 21 ARGs were significantly down-regulated, while 22 ARGs were up-regulated in TNBC patients. These DE-ARGs interacted with each other forming an autophagy network as measured by string (Figure 2.B). Moreover, we observed many mutations occur on these DE-ARGs in TNBCs (Supplementary Figure 1).

Enrichment Analysis of DE-ARGs

To determine the functional enrichment of DE-ARGs, we performed GO and KEGG enrichment analysis. We found that these 43 DE-ARGs were highly correlated to autophagy, process utilizing autophagic mechanism and peptidyl serine modification in the term of biological process (BP). In the aspect of cellular components (CC), these genes were enriched in nuclear envelope, mitochondrial outer membrane and organelle outer membrane. For the molecular functions (MF), these genes were mainly concentrated

in ubiquitin protein ligase binding, ubiquitin-like protein ligase binding and protein phosphatase (Supplementary Figure 2.A). Moreover, KEGG enrichment analysis indicated that the DE-ARGs were involved in the signaling pathways such as autophagy-animal, EGFR tyrosine kinase inhibitor resistance and apoptosis (Supplementary Figure 2.B). Overall, these data suggested that these ARGs play a role in other biological process in addition to autophagy.

Construction of a Prognostic ARG Signature of TNBC in the Train Set

To build a ARG prognostic model, we first analyze the risk score of all ARGs in TNBC by performing univariate Cox regression analysis. Eight ARGs were screened out including seven potential risky genes and one potential protective gene (Figure 3.A). Subsequently, we performed Lasso regression analysis on the basis of univariate Cox regression analysis (Figure 3.C, D). Then, we constructed the prognostic ARG signature for OS using CDKN1A, CTSD, CTSL, EIF4EBP1, TMEM74 and VAMP3 by Lasso regression. Finally, we performed multivariate Cox regression analysis and screened out four ARGs including three potential risk genes and one potential protective gene (Figure 3.B).

Next, we tested if the expression of these six ARGs were correlated with prognosis of TNBCs. We found that high expression of EIF4EBP1 (p=0.046), CTSL (p=0.009) and CTSD (p=0.07) might indicate a worse prognosis. There were no statistical differences in the survival analyses of CDKN1A (p=0.362), TMEM74 (p=0.017) and VAMP3 (p=0.0189) (Supplementary Figure 3). Then we want to validate whether this ARG signature can predict OS of TNBC. We first divided TNBC patients into “high risk” (n=50) and “low risk” (n=51) group according to the threshold of the median risk score (Figure 4.A). The risk score for each patient was calculated based on the formulate : risk score=(0.246026×CDKN1A)+(0.359130×CTSD)+ (0.234375×CTSL)+(0.590736×EIF4EBP1)+(-0.281261×TMEM74)+(0.338378×VAMP3).Patients were assigned to high-risk (n=50) and low-risk groups (n=51) according to the threshold of the median risk score. Patients with higher scores were more likely to poorer prognosis (Figure 4.C). A heatmap was used to visualize differences in expression levels of the six ARGs between groups (Figure 4.D). Survival curves further indicated that patients in high risk group showed a significantly lower probability of survival compared to low risk group ($p < 0.05$) (Figure 4.B). ROC analysis showed that the AUCs for 1-year, 3-years and 5-years OS were 0.925、0.866 and 0.784, respectively (Figure 4.E). These data suggested that ARGs signature in our model could benefit the prognosis prediction of TNBCs.

Validation of the Risk Score of ARG signature in a GEO Test Set

To further validate the prognostic and predictive role of ARGs signature, we employed another GEO cohort as a test set to calculate risk scores using the same formula used in the train set. The patients from of the test set were divided into the high-risk group (n=34) and low-risk group (n=73) by the median value of the train set (Figure 5.A), and higher risk score predicted poorer prognosis in the patients (Figure 5.C). A heatmap was presented to visualize the difference expression levels of the six ARGs between test groups (Figure 5.D). Similar to the train set, patients in the high-risk score group showed a poorer prognosis compared to low-risk group in the test set ($p < 0.05$) (Figure 5.B). Time-dependent ROC analysis showed that the prognostic accuracy of OS was 0.798 at 1 year, 0.564at 3 years and 0.696 at 5 years (Figure 5.E).

Independent Prognostic Indicator of the Prognostic Risk Model

To confirm whether risk scores can be used as an independent predictor for TNBC patients' survival, we further performed univariate analysis in training set. Univariate Cox regression analysis revealed that N, T, stage and risk score were meaningful for predicting OS (Figure 6.A). Subsequently, we performed a multivariate Cox regression analysis to verify risk score as independent predictor ($p<0.001$) (Figure 6.B). Moreover, we identified that the expression of CTSD was significantly associated with stages ($p=0.025$) (Supplementary Figure 4.A) and T ($p=0.031$) (Supplementary Figure 4.B). These data demonstrated that our model could be a reliable prognostic predictor and biomarker in addition to known clinical classification.

Construction of the Nomogram and Performance Validation

To provide the clinician with a better quantitative method to predict prognosis of TNBC patient, we established a nomogram with multiple factors including N, T, stage and risk score (Figure 7.A). The nomogram was used to evaluate the survival probability of 1-, 3-, and 5-year. Nomograms showed a good performance with a high C-index of 0.764, suggesting that it could be served as an effective tool for prognostic evaluation of patients with TNBC. In addition, we constructed calibration curves which showed that the predicted and actual survival rates were in agreement with 1, 3, and 5 years (Figure 7.B-D). Finally, we compared the predictive accuracy for TNBC between the nomogram and clinicopathological risk factors by the values of AUC. Our model's AUC value (AUC of 1-year, 3-year, 5-year OS) was higher than the traditional prognostic scoring systems (Figure 7.E-G). These findings revealed that the nomogram with our risk scores can improve predicting OS.

Enrichment Analysis between high-risk group and low-risk group

Finally, we performed Gene Set Enrichment Analysis (GSEA) between the high-risk group and the low-risk group in TCGA and GSE58812 cohort respectively to further provide biological insight. We found that the enriched KEGG pathways of high-risk group in TCGA cohort included Apoptosis, Fc epsilon RI signaling pathway, Glycosylphosphatidylinositol GPI anchor biosynthesis, Lysosome and Olfactory transduction. Meanwhile, enriched KEGG pathways of low risk group included Protein export, RIG-I like receptor signaling pathway, RNA polymerase, Taste transduction, Toll-like receptor signaling pathway (Supplementary Figure 5.A). In addition, KEGG enrichment pathway analysis of high-risk group in GSE58812 cohort indicated that the genes were enriched in ABC transporters, Arginine and proline metabolism, Lysosome, Pathogenic Escherichia coli infection and Pentose phosphate pathway. KEGG enrichment pathways analysis of the low-risk group in GSE58812 cohort were mainly concentrated in Regulation of autophagy, RIG-I like receptor signaling pathway, RNA degradation, Spliceosome and Vibrio cholerae infection (Supplementary Figure 5.B).

Knockdown of eIF4EBP1 inhibited TNBC cell proliferation and migration

We next want to test the biological function of these ARGs in our model in TNBCs. Among these six genes, the function of eIF4EBP1 in TNBCs remains unknown. We knockdown eIF4EBP1 using two independent siRNAs in two TNBC cell lines: MDA-MB-231 and BT549 (Figure. 8A). Depletion of eIF4EBP1resulted in dramatic decrease of cell growth and colony formation (Figure 8.B, C, D, F). Edu staining showed that knockdown of eIF4EBP induced a dramatically decrease in proliferation (Figure. 8E, H). In addition, eIF4EBP1 knockdown significantly impaired cell metastasis as measured by trans-well and wound healing assay (Figure 8.G, I, J, K). Furthermore, we observed increased eIF4EBP1 expression in primary TNBC samples compared to adjacent normal tissues in collected 3 TNBC patients (Figure 8.L). Based on the Human Protein Atlas database, the protein expression levels of eIF4EBP1 were evaluated by CAB005032 antibody. Among 12 TNBC tissues examined, 6 cases had medium to high staining (4 medium and 8 high), while no cases had low staining. Representative IHC image showed that eIF4EBP1 staining was higher in TNBC than in normal tissues (Figure 8.M). Overall, these findings suggest a potential oncogenic role of eIF4EBP1 in TNBCs supporting the importance of our prognosis model in TNBCs.

Discussion

TNBC is one of the most serve malignant tumors among women in the world. Although less than 20% of all diagnosed breast cancer patients are triple-negative breast cancer, there are still 25 to 40% of patients of the total breast cancer population with metastases, accounting for a disproportionately number of deaths from breast cancer [11]. Due to lack of targetable receptors, TNBC represents a clinically challenging endeavor. Currently, treating TNBC mainly includes adjuvant chemotherapy plus surgical resection for an early-stage and adjuvant chemotherapy for an advanced stage. However, surgical resection may provide an unsatisfactory effect because of its highly invasive growth pattern and developed metastasis. Additionally, chemotherapy effects are diminished due to tumor heterogeneity. Even worse, TNBC is insensitive to the usual hormone therapies because of lack of hormone receptors expression. Therefore, it is essential to establish a novel biomarker to predict the prognosis and provide reliable treatment targets of TNBC.

In recent year, several prognostic factors are established. Yiduo Liu et al screened four heterogeneous-related genes (FAM83B, KITLG, RBM24 and S100B) from 105 genes to construct a prognostic signatures in the disease-free interval (DFI) of TNBC[12]. Chao Li et al identified a prognosis related signature associated with energy metabolism including eight energy metabolism-associated genes (IL1RL2, FBLN7, CA3, PDE1B, SLURP1, CILP, AQP7, TPSB) in triple negative breast cancer[13]. Ji Yeon Kim et al obtained thirteen immune related genes to predict distant recurrence of early TNBC [14]. Huan-Ming Hsu et al unveiled six immunoglobulin genes as biomarkers in TNBC patients and explored a potential biomarkers of recurrence for TNBC [15]. Fei Chen et al confirmed nine steroidhormone related genes as independent prognostic markers based on RNA-seq analysis in TNBC [16]. To some degree, these models all showed better predicting ability than other clinicopathological factors, and added prognostic value to the TNM staging system. Many studies have shown that autophagy plays an important role in prognosis of multiple cancers. However, to our knowledge, autophagy related prognostic risk models haven’t been

established for TNBC yet. It is of great significance to develop an autophagy associated biomarker for TNBC prognosis prediction.

In this study, we mined forty-three DE-ARGs by comparing TNBC samples to normal samples. Subsequently, GO and KEGG pathway enrichment of these DE-ARGs revealed that some cancer-related signaling pathway were significantly enriched, such as autophagy, apoptosis and HIF-1 signaling pathway. Through further univariate Cox regression and Lasso regression analysis, six ARGs (CDKN1A, CTSD, CTSL, EIF4EBP1, TMEM74 and VAMP3) were obtained. Finally, we established a prognostic signature based on the ARGs to effectively predict the prognosis of TNBC patients. These six ARGs have been reported in various cancer types. A wide array of studies documented that CTSD plays an important role in tumor growth, invasion, and metastatic dissemination [17-20]. Wei Zhang et al found that the CTSL expression levels in malignant ovarian tumors were significantly higher than in normal or benign tissues [21]. Luosheng Zhang et al also found that CTSL involved in proliferation and invasion of ovarian cancer cells [22]. Some studies indicated EIF4EBP1 is involved in affecting the progression of various cancer types (including renal cell carcinoma, breast cancer) based on regulating gene transcription[23, 24]. Kevin Luftman et al confirmed that silencing of VAMP3 could inhibit cancer metastasis[25]. These results were in consistence with our findings. Of note, data on the prognostic relevance of CDKN1A expression showed that increased expression of CDKN1A were associated with poor prognosis in esophageal, ovarian, prostate cancers and gliomas[26-32]. In contrast, some studies also indicated that low expression level of CDKN1A was correlated with better survival in cervical, gastric, cholangiocarcinoma and ovarian cancers [33-35]. These findings suggested the dual role of CDKN1A in cancer, that need to be further explored. Interestingly, TMEM74 had been regarded as an oncogene in cancers including liver cancer, lung cancer, breast cancer, colon cancer, cervical cancer and hepatic carcinoma. Higher expression level of TMEM74 was associated with poorer survival, which was not consistent with our study[36]. This could be due to the variation of genetic context in different cancer types. High level expression of TMEM74 played protective role in TNBC not in others.

This study need to be expanded in the future as the sample number of each cohort used is relatively small. Additionally, further studies are required to understand the role of autophagy in TNBC and its potential molecular mechanisms.

Conclusions

Based on six ARGs (CDKN1A, CTSD, CTSL, EIF4EBP1, TMEM74 and VAMP3) through mining the TCGA database, we developed a risk prediction model that can help us effectively predict the survival status of TNBC patients. These findings emphasized the vital role of the autophagy-related genes in TNBC and may provide new therapeutic targets for TNBC.

Declarations

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Key Scientific and Technological Project of Henan Province (Grant No.202102310416) and the Key Scientific Research Projects of Henan Colleges and Universities (Grant No. 21A310008).

Author contribution

Ruoling Jia and Cheng Yan designed the research and wrote the paper. Cheng Yan and Qingling Liu downloaded and analyzed the data. Cheng Yan conducted the cellculture-related experiment. All authors have agreed to the manuscript.

Acknowledgements

The authors thank participants and staff of Xinxiang University for their contributions.

Availability of data and materials

All data and R script in this study are available from the corresponding author upon reasonable request. All authors read and approved the final manuscript.

References

1. **Erratum: Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA Cancer J Clin* 2020, **70**(4):313.
2. DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, Jemal A, Siegel RL: **Breast cancer statistics, 2019**. *CA Cancer J Clin* 2019, **69**(6):438-451.
3. Scott LC, Mobley LR, Kuo TM, Il'yasova D: **Update on triple-negative breast cancer disparities for the United States: A population-based study from the United States Cancer Statistics database, 2010 through 2014**. *Cancer* 2019, **125**(19):3412-3417.
4. Harbeck N, Gnant M: **Breast cancer**. *Lancet* 2017, **389**(10074):1134-1150.
5. Peng F, Tang H, Liu P, Shen J, Guan X, Xie X, Gao J, Xiong L, Jia L, Chen J *et al*: **Isoliquiritigenin modulates miR-374a/PTEN/Akt axis to suppress breast cancer tumorigenesis and metastasis**. *Sci Rep* 2017, **7**(1):9022.
6. Rabinowitz JD, White E: **Autophagy and metabolism**. *Science* 2010, **330**(6009):1344-1348.

7. Xia H, Green DR, Zou W: **Autophagy in tumour immunity and therapy**. *Nat Rev Cancer* 2021, **21**(5):281-297.
8. Li M, Liu J, Li S, Feng Y, Yi F, Wang L, Wei S, Cao L: **Autophagy-related 7 modulates tumor progression in triple-negative breast cancer**. *Lab Invest* 2019, **99**(9):1266-1274.
9. Chen W, Bai Y, Patel C, Geng F: **Autophagy promotes triple negative breast cancer metastasis via YAP nuclear localization**. *Biochem Biophys Res Commun* 2019, **520**(2):263-268.
10. Singh SS, Vats S, Chia AY, Tan TZ, Deng S, Ong MS, Arfuso F, Yap CT, Goh BC, Sethi G *et al*: **Dual role of autophagy in hallmarks of cancer**. *Oncogene* 2018, **37**(9):1142-1158.
11. Yao H, He G, Yan S, Chen C, Song L, Rosol TJ, Deng X: **Triple-negative breast cancer: is there a treatment on the horizon?** *Oncotarget* 2017, **8**(1):1913-1924.
12. Liu Y, Teng L, Fu S, Wang G, Li Z, Ding C, Wang H, Bi L: **Highly heterogeneous-related genes of triple-negative breast cancer: potential diagnostic and prognostic biomarkers**. *BMC Cancer* 2021, **21**(1):644.
13. Li C, Li X, Li G, Sun L, Zhang W, Jiang J, Ge Q: **Identification of a prognosisassociated signature associated with energy metabolism in triplenegative breast cancer**. *Oncol Rep* 2020, **44**(3):819-837.
14. Kim JY, Jung HH, Sohn I, Woo SY, Cho H, Cho EY, Lee JE, Kim SW, Nam SJ, Park YH *et al*: **Prognostication of a 13-immune-related-gene signature in patients with early triple-negative breast cancer**. *Breast Cancer Res Treat* 2020, **184**(2):325-334.
15. Hsu HM, Chu CM, Chang YJ, Yu JC, Chen CT, Jian CE, Lee CY, Chiang YT, Chang CW, Chang YT: **Six novel immunoglobulin genes as biomarkers for better prognosis in triple-negative breast cancer by gene co-expression network analysis**. *Sci Rep* 2019, **9**(1):4484.
16. Chen F, Li Y, Qin N, Wang F, Du J, Wang C, Du F, Jiang T, Jiang Y, Dai J *et al*: **RNA-seq analysis identified hormone-related genes associated with prognosis of triple negative breast cancer**. *J Biomed Res* 2020, **34**(2):129-138.
17. Gemoll T, Epping F, Heinrich L, Fritzsche B, Roblick UJ, Szymczak S, Hartwig S, Depping R, Bruch HP, Thorns C *et al*: **Increased cathepsin D protein expression is a biomarker for osteosarcomas, pulmonary metastases and other bone malignancies**. *Oncotarget* 2015, **6**(18):16517-16526.
18. Park YJ, Kim EK, Bae JY, Moon S, Kim J: **Human telomerase reverse transcriptase (hTERT) promotes cancer invasion by modulating cathepsin D via early growth response (EGR)-1**. *Cancer Lett* 2016, **370**(2):222-231.
19. Maynadier M, Farnoud R, Lamy PJ, Laurent-Matha V, Garcia M, Rochefort H: **Cathepsin D stimulates the activities of secreted plasminogen activators in the breast cancer acidic environment**. *Int J Oncol* 2013, **43**(5):1683-1690.
20. Zhang M, Wu JS, Yang X, Pang X, Li L, Wang SS, Wu JB, Tang YJ, Liang XH, Zheng M *et al*: **Overexpression Cathepsin D Contributes to Perineural Invasion of Salivary Adenoid Cystic Carcinoma**. *Front Oncol* 2018, **8**:492.
21. Zhang W, Wang S, Wang Q, Yang Z, Pan Z, Li L: **Overexpression of cysteine cathepsin L is a marker of invasion and metastasis in ovarian cancer**. *Oncol Rep* 2014, **31**(3):1334-1342.

22. Zhang L, Wei L, Shen G, He B, Gong W, Min N, Zhang L, Duan Y, Xie J, Luo H *et al*: **Cathepsin L is involved in proliferation and invasion of ovarian cancer cells**. *Mol Med Rep* 2015, **11**(1):468-474.
23. Wan P, Chen Z, Zhong W, Jiang H, Huang Z, Peng D, He Q, Chen N: **BRDT is a novel regulator of eIF4EBP1 in renal cell carcinoma**. *Oncol Rep* 2020, **44**(6):2475-2486.
24. Rutkovsky AC, Yeh ES, Guest ST, Findlay VJ, Muise-Helmericks RC, Armeson K, Ethier SP: **Eukaryotic initiation factor 4E-binding protein as an oncogene in breast cancer**. *BMC Cancer* 2019, **19**(1):491.
25. Luftman K, Hasan N, Day P, Hardee D, Hu C: **Silencing of VAMP3 inhibits cell migration and integrin-mediated adhesion**. *Biochem Biophys Res Commun* 2009, **380**(1):65-70.
26. Sarbia M, Stahl M, zur Hausen A, Zimmermann K, Wang L, Fink U, Heep H, Dutkowski P, Willers R, Muller W *et al*: **Expression of p21WAF1 predicts outcome of esophageal cancer patients treated by surgery alone or by combined therapy modalities**. *Clin Cancer Res* 1998, **4**(11):2615-2623.
27. Ohashi R, Angori S, Batavia AA, Rupp NJ, Ajioka Y, Schraml P, Moch H: **Loss of CDKN1A mRNA and Protein Expression Are Independent Predictors of Poor Outcome in Chromophobe Renal Cell Carcinoma Patients**. *Cancers (Basel)* 2020, **12**(2).
28. Lin Y, Shen LY, Fu H, Dong B, Yang HL, Yan WP, Kang XZ, Dai L, Zhou HT, Yang YB *et al*: **P21, COX-2, and E-cadherin are potential prognostic factors for esophageal squamous cell carcinoma**. *Dis Esophagus* 2017, **30**(2):1-10.
29. Ferrandina G, Stoler A, Fagotti A, Fanfani F, Sacco R, De Pasqua A, Mancuso S, Scambia G: **p21WAF1/CIP1 protein expression in primary ovarian cancer**. *Int J Oncol* 2000, **17**(6):1231-1235.
30. Baretton GB, Klenk U, Diebold J, Schmeller N, Lohrs U: **Proliferation- and apoptosis-associated factors in advanced prostatic carcinomas before and after androgen deprivation therapy: prognostic significance of p21/WAF1/CIP1 expression**. *Br J Cancer* 1999, **80**(3-4):546-555.
31. Aaltomaa S, Lipponen P, Eskelinen M, Ala-Opas M, Kosma VM: **Prognostic value and expression of p21(waf1/cip1) protein in prostate cancer**. *Prostate* 1999, **39**(1):8-15.
32. Korkolopoulou P, Kouzelis K, Christodoulou P, Papanikolaou A, Thomas-Tsagli E: **Expression of retinoblastoma gene product and p21 (WAF1/Cip 1) protein in gliomas: correlations with proliferation markers, p53 expression and survival**. *Acta Neuropathol* 1998, **95**(6):617-624.
33. Abbas T, Dutta A: **p21 in cancer: intricate networks and multiple activities**. *Nat Rev Cancer* 2009, **9**(6):400-414.
34. Lu X, Toki T, Konishi I, Nikaido T, Fujii S: **Expression of p21WAF1/CIP1 in adenocarcinoma of the uterine cervix: a possible immunohistochemical marker of a favorable prognosis**. *Cancer* 1998, **82**(12):2409-2417.
35. Yu Y, Zhang M, Wang N, Li Q, Yang J, Yan S, He X, Ji G, Miao L: **Epigenetic silencing of tumor suppressor gene CDKN1A by oncogenic long non-coding RNA SNHG1 in cholangiocarcinoma**. *Cell Death Dis* 2018, **9**(7):746.
36. Sun Y, Deng J, Xia P, Chen W, Wang L: **The Expression of TMEM74 in Liver Cancer and Lung Cancer Correlating With Survival Outcomes**. *Appl Immunohistochem Mol Morphol* 2019, **27**(8):618-625.

Figures

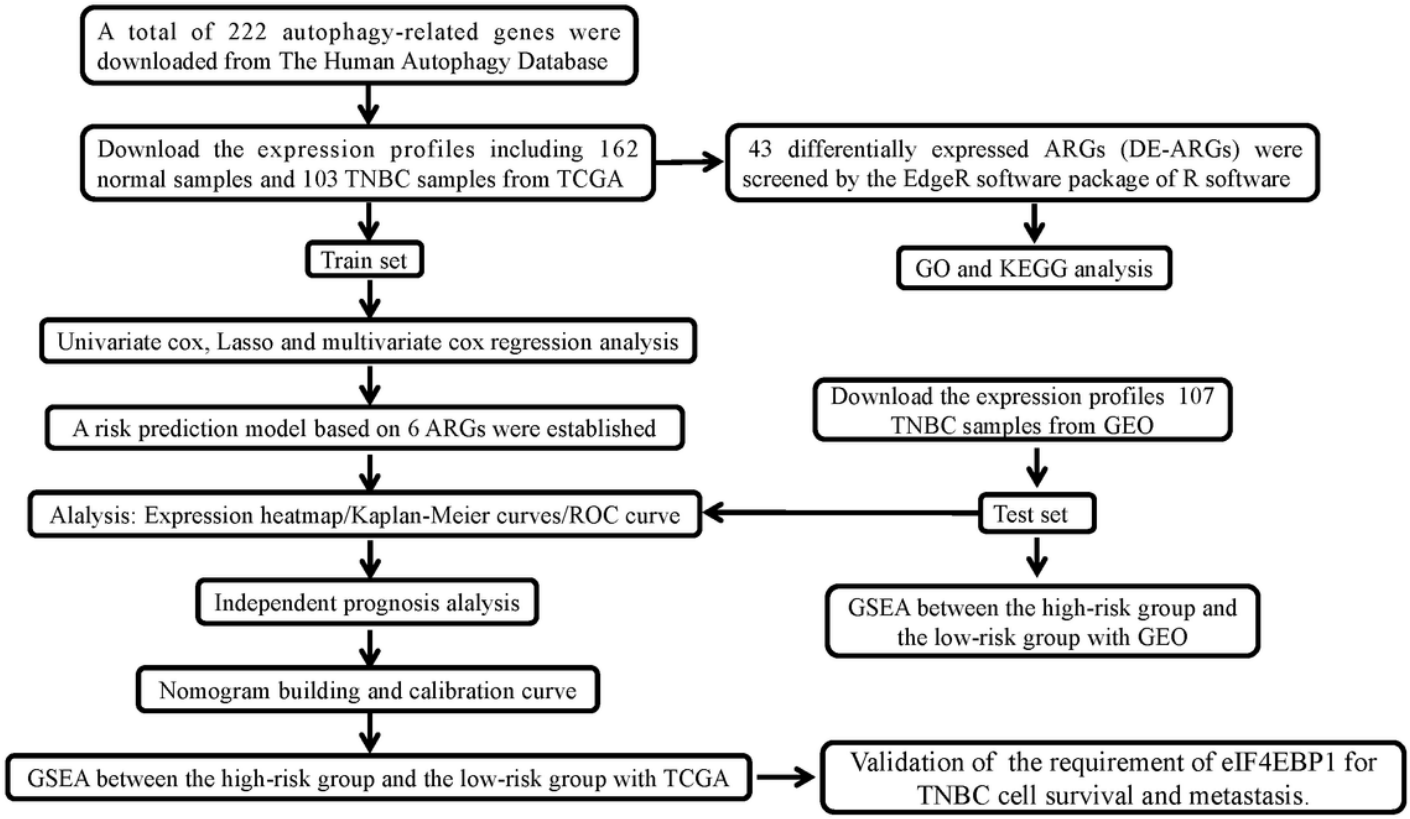

Figure 1

The flowchart describing the experimental design to establish and validate the prognostic signature in the study.

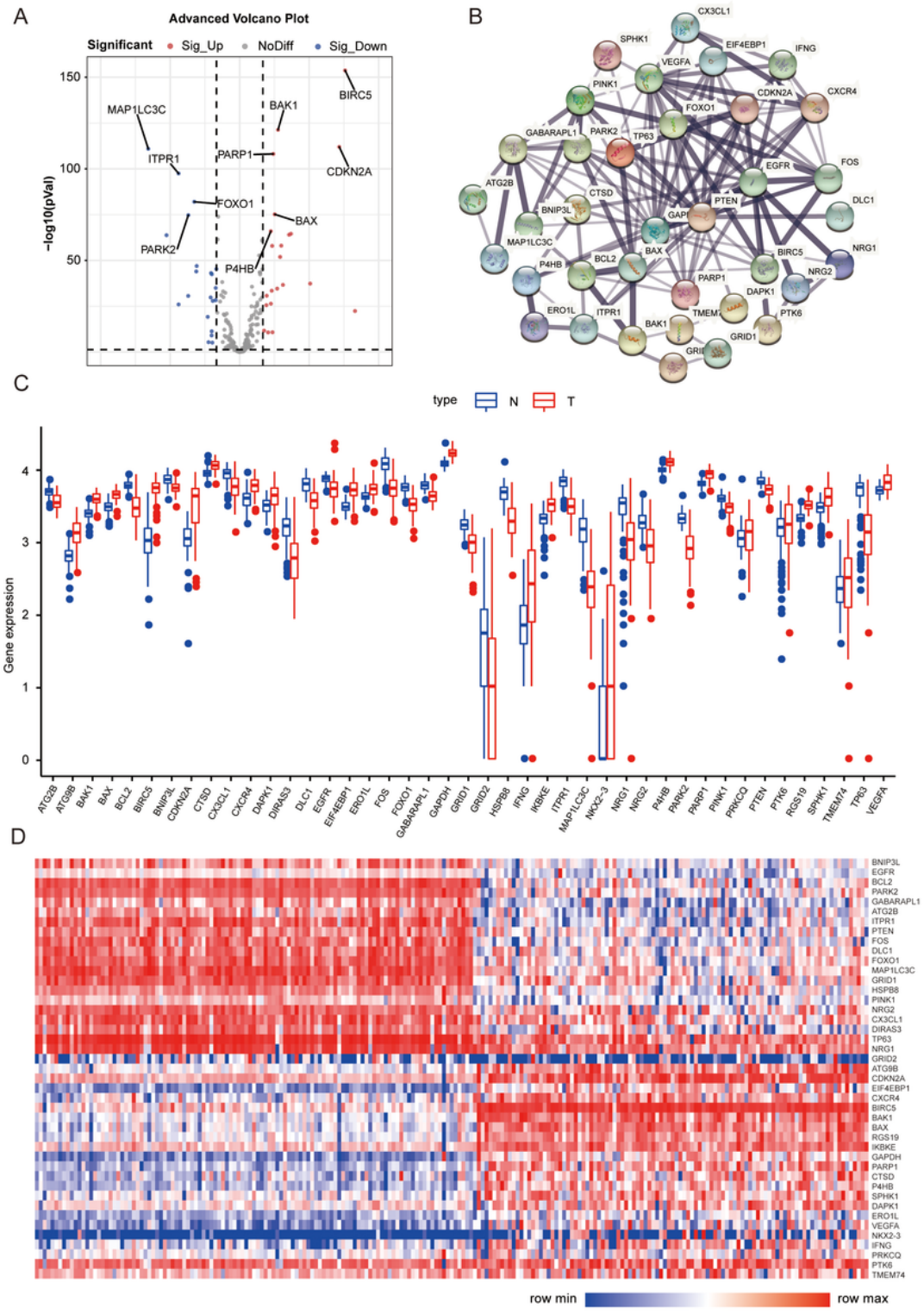

Figure 2

Differentially expressed autophagy-related genes. (A) Volcano map showed differentially expressed genes between normal samples and TNBC samples. Red dots represent significantly up-regulated genes, blue dots represent significantly down-regulated genes and gray dots represent ARGs with no difference. (B) Protein-protein interactions (PPI) network of ARGs using the STRING database. (C) Box plots showed gene expression values of ARGs. Blue represents normal samples, red represents tumor samples.(D)

Heatmap showed the expression levels of ARGs. The color scale represented the expression levels normalized by z-score.

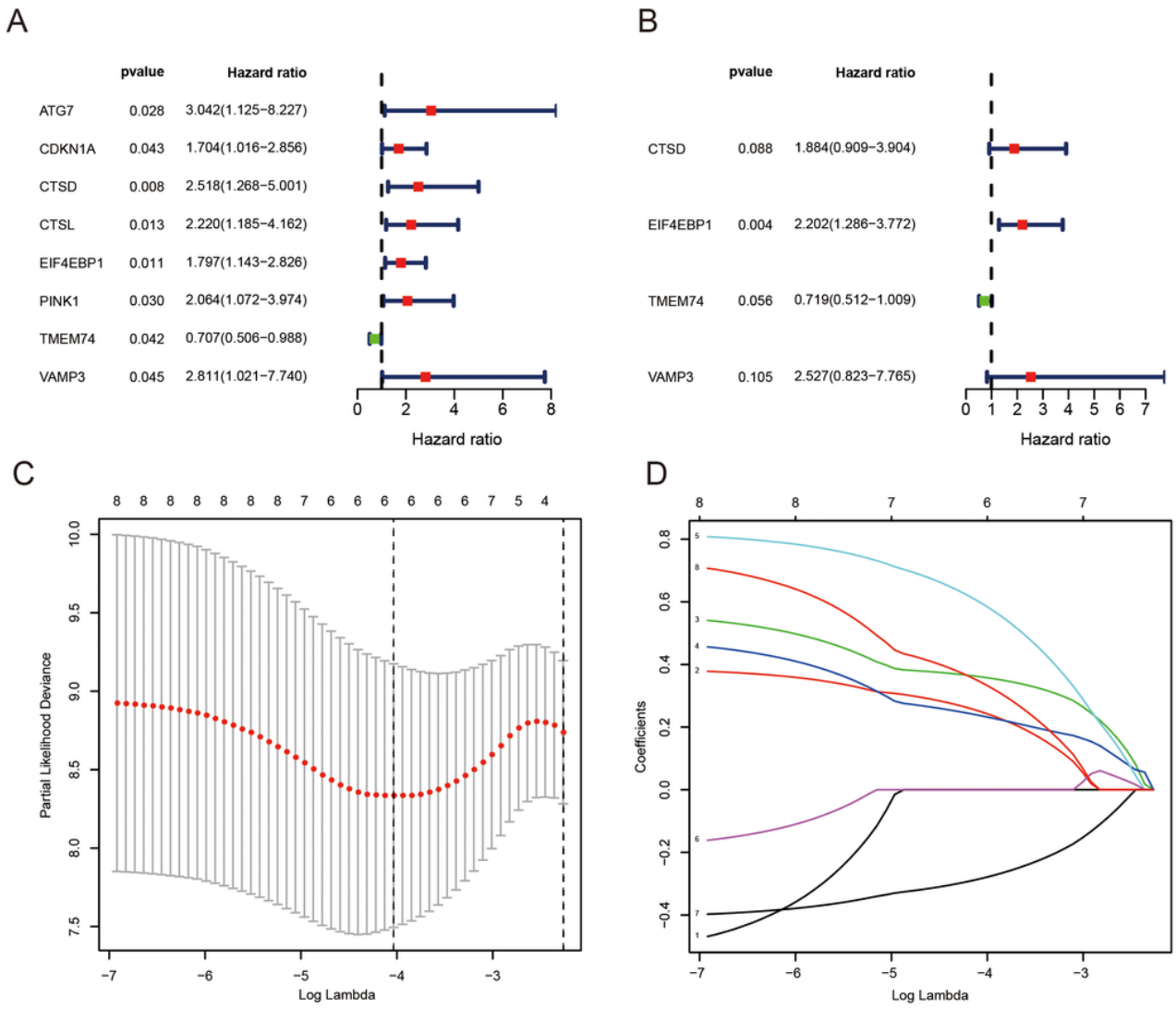

Figure 3

Identification of ARGs with prognostic value in breast cancer. (A) Univariate Cox regression hazard model for the overall survival in TNBC. (B) Multivariate Cox regression hazard model for the overall survival in TNBC. (C) Lasso regression analysis of ARGs based on univariate Cox regression analysis. The horizontal axis represents the log value of the independent variable λ, whilst the vertical axis represents the partial likelihood deviance of the log value of each independent variable λ. (D) Coefficients calculated for each lambda. Each line represents a gene confidence value.

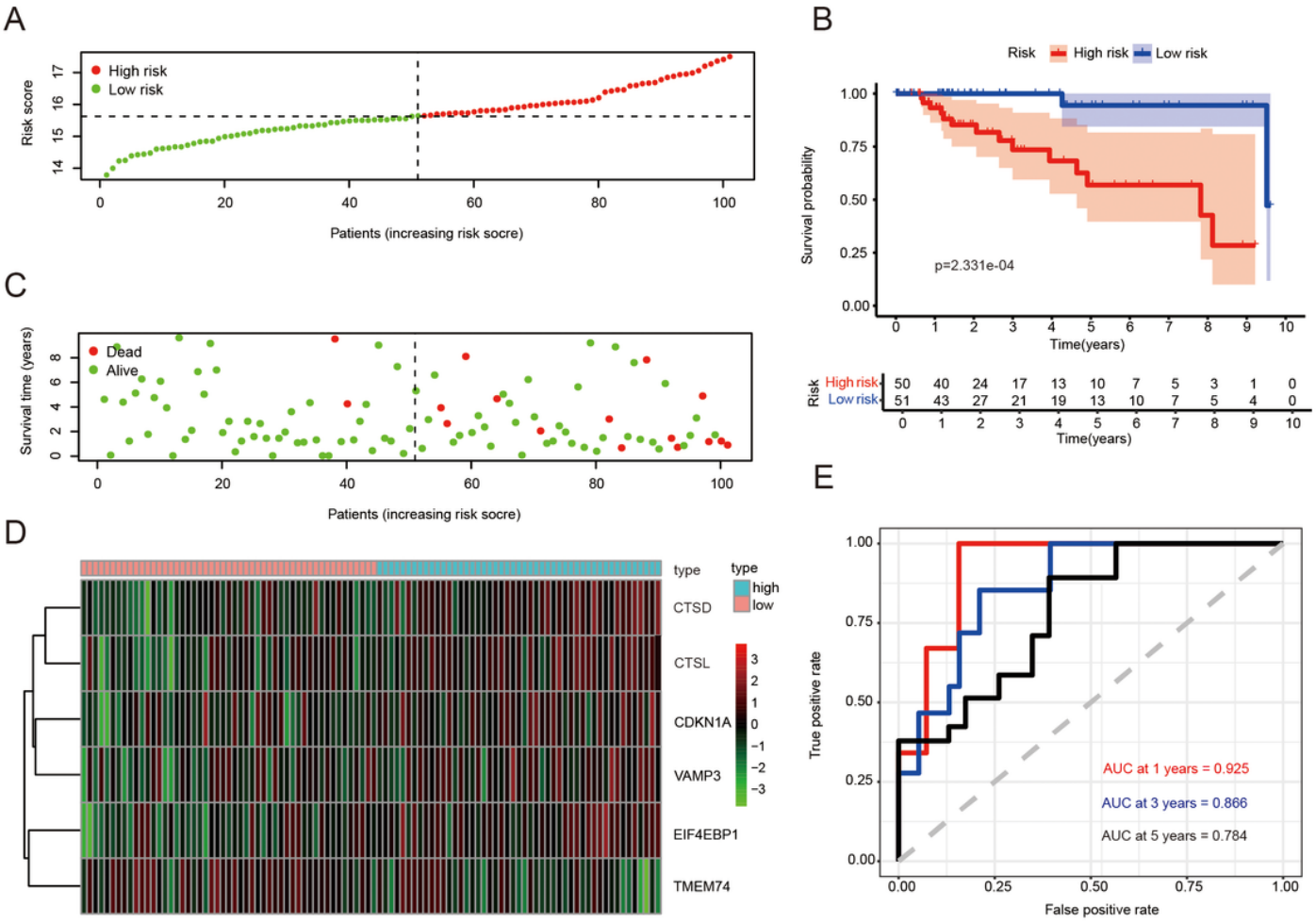

Figure 4

ARGs signature based risk analysis and overall survival in TNBC patients in the train set. (A) The risk score stratified the TNBC patients into high- risk groups (“High” red line) and low-risk groups (“Low” green line). (B) Kaplan-Meier survival curves to show survival probability comparing the high-risk groups with low-risk group. (C) Comparison of Survival time and survival status of patients in TNBC between high-risk groups and low-risk groups. Green plots for alive, red plots for dead. (D) Heatmap showing expression of the six genes screened from ARGs in TNBC. Blue color represents high group, while pink color represents low group. The color scale represented the expression levels normalized by z-score. (E) Time-dependent ROC curves for survival prediction by the risk score.

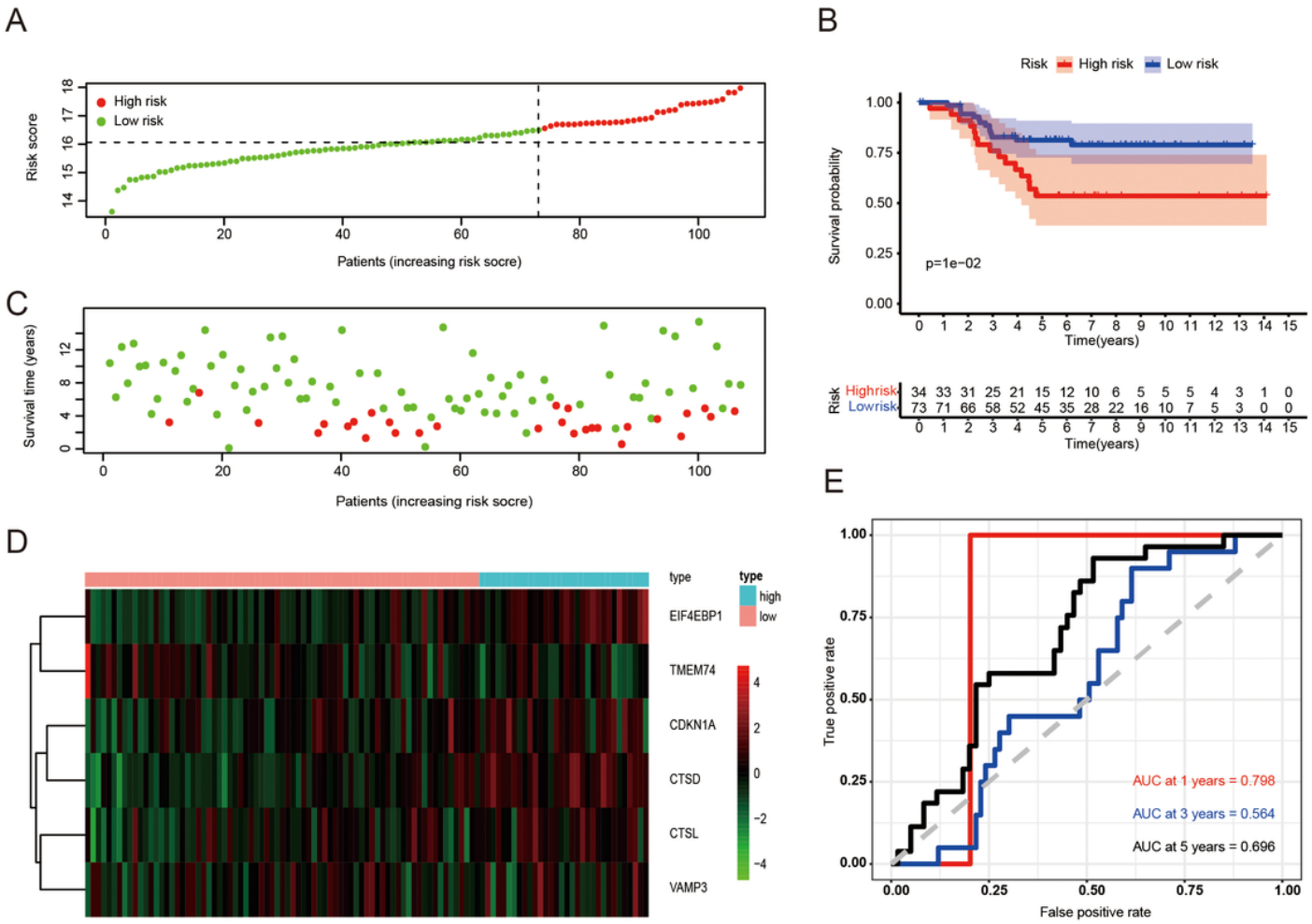

Figure 5

ARGs signature based risk analysis and overall survival in TNBC patients in the test set. (A) The risk score stratified the TNBC patients into high-risk groups (“High” red line) and low-risk groups (“Low” green line). (B) Kaplan-Meier survival curves to show survival probability comparing the high- risk groups with low-risk group. (C) Survival time and survival status of patients in TNBC comparing high-risk group with low-risk group. Green plots for alive, red plots for dead. (D) Heatmap showing expression of the six genes screened from ARGs in TNBC. Blue color represents high group, while pink color represents low group. The color scale represented the expression levels normalized by z-score. (E) Time-dependent ROC curves for survival prediction by the risk score.

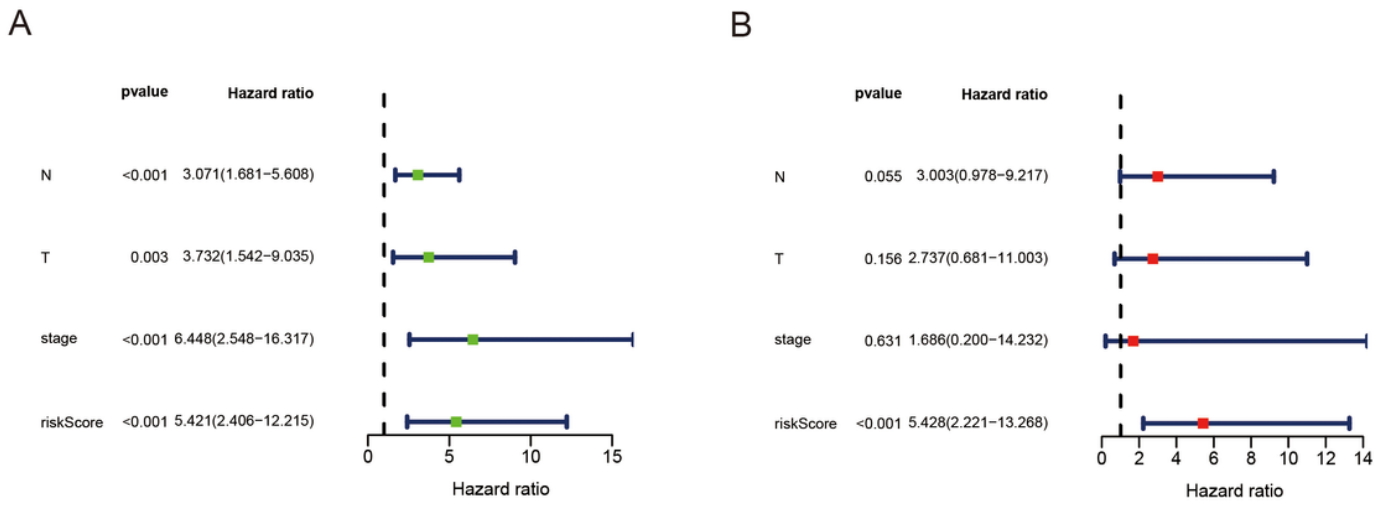

Figure 6

Analysis of the risk scores as an independent prognostic indicator. (A) Univariate Cox regression analysis identified that N, T, stage and risk score were significantly associated with OS prediction. (B) Multivariate Cox regression analysis identified that risk score was independent prognostic factors for TNBC.

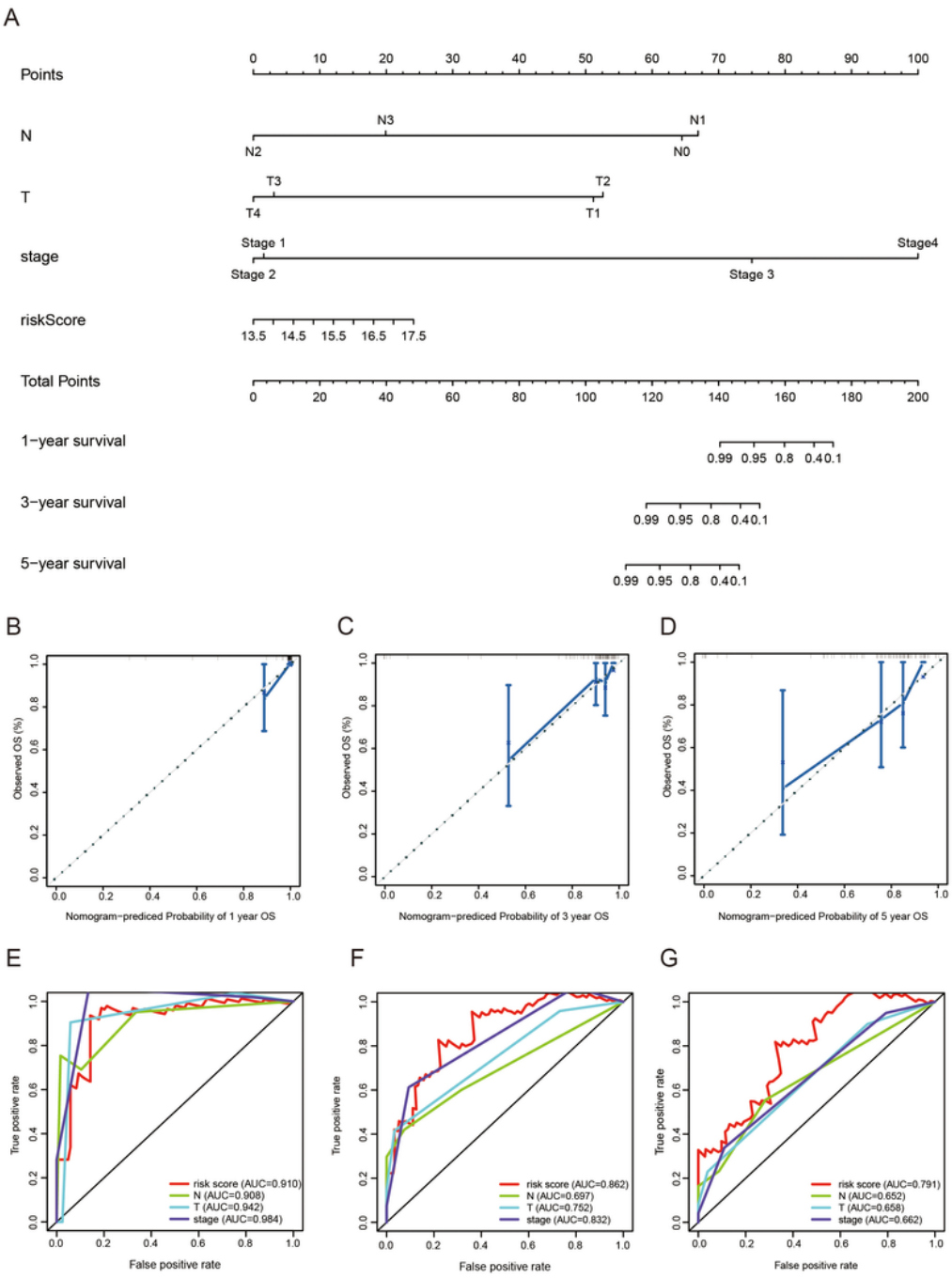

Figure 7

The nomogram to predict overall survival of TNBC patients of TAGC cohort. (A) The nomogram for predicting survival proportion of patients in 1-, 3-, and 5-year. (B-D) The calibration plots for predicting patient survival at 1-, 3- and 5-year. (E-G) The ROC curve of OS for risk score, N, T and stage at 1-, 3- and 5-year.

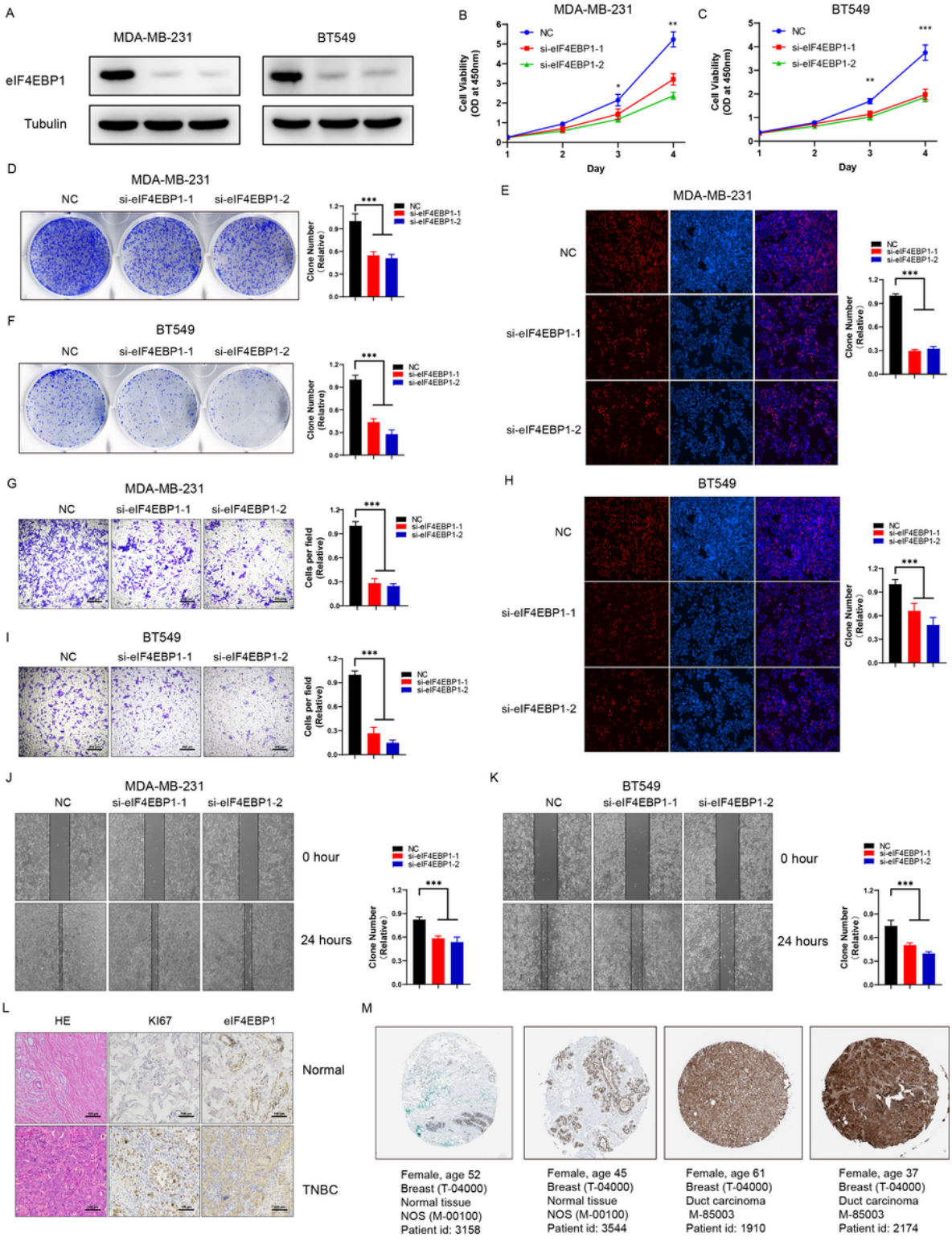

Figure 8

EIF4EBP1 is required for TNBC cell survival and migration. (A) Western blot to show knockdown efficiency of eIF4EBP1 in MDA-MB-231 and BT549 cells by two independent siRNAs. (B, C) Cell proliferation of MDA-MB-231 cells(B) or BT549 cells(C) transfected with control or eIF4EBP1 siRNAs was measured by CCK8. (D, F) Colony formation of MDA-MB-231 cells or BT549 cells transfected with control or eIF4EBP1 siRNAs was measured by ImageJ. (G, I) Trans-well assay to show the cell metastasis of control cells

comparing to eIF4EBP1 knockdown cells. (E, H) Edu assay to show the cell proliferation of control cells comparing to eIF4EBP1 knockdown cells. (J, K) Wound healing assay to show the cell migration of control cells comparing to eIF4EBP1 depleted cells. (L) Representative images of HE staining, immunostaining of KI67 and eIF4EBP1 in primary TNBC samples versus normal samples. (M) Representative images of immunostaining of eIF4EBP1 in primary TNBC samples compared to normal tissues from HPA database. All data are shown as mean ± s.e.m. *P < 0.05, **P < 0.01, ***P < 0.001 by one-way ANOVA.

Supplementary Files

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

- supplement1.jpg
- supplement2.jpg
- supplement3.jpg
- supplement4.jpg
- supplement5.jpg