Identification and comparison of two risk models based on characteristic gene sets of breast cancer for predictive prognosis

Chaoqun Ma (zhouhong_sznh@163.com)
Traditional Chinese Medicine Hospital of Kunshan
Ying Yu
Traditional Chinese Medicine Hospital of Kunshan
Jiaying Chen
Nanjing University of Chinese Medicine
Chang Yao
Traditional Chinese Medicine Hospital of Kunshan
Weihe Bian
Traditional Chinese Medicine Hospital of Kunshan
Cong Wang
Traditional Chinese Medicine Hospital of Kunshan
Bei Ye
Traditional Chinese Medicine Hospital of Kunshan
Tong Shen
Traditional Chinese Medicine Hospital of Kunshan
Mengmeng Guo
Nantong Hospital of Traditional Chinese Medicine
Xiping Zhang
Nanjing University of Chinese Medicine
Shihan Cao
Nanjing University of Chinese Medicine

Research Article

Keywords: breast cancer, risk models, characteristic gene set, prediction, prognosis

Posted Date: April 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1563580/v1

Abstract

Background: Breast cancer (BC) is one of the main causes of cancer related deaths in women, and also has a high mortality rate in men, resulting in great healthy problems worldwide. Because of its low cure rate, poor late prognosis, and high mortality, it is of great significance to find new biomarkers for diagnosis and prognosis.

Methods: In this study, 1030 cases of BCs from The Cancer Genome Atlas (TCGA) were obtained for differential expression analysis and short time-series expression miner (STEM) analysis to identify the BC development characteristic genes, divided into up-regulated genes and down-regulated genes with the development of BC. Two predictive prognosis models were both defined by Lasso. Survival analysis and ROC curve analysis were used respectively to determine the diagnostic and prognostic abilities of the two gene set models scores. Univariate and multivariate Cox regression analyses were used to determine whether the risk models are independent prognostic indicators and whether our models are better than the present clinicopathological features in prognosis.

Results: Our findings from this study suggest both the breast cancer unfavorable (BC1) and the breast cancer favorable (BC2) gene sets are reliable biomarkers for the diagnosis and independent biomarkers for predictive prognosis of BC, while BC1 model presents better diagnostic and prognostic value than BC2 model. Associations between the models and macrophages M2, the sensitivity to Bortezomib were also found, implying that the BC unfavorable genes are significantly involved in the tumor immune microenvironment of BC.

Conclusion: In conclusion, we successfully established two predictive prognosis models based on characteristic gene sets of BC, and one prognosis model (BC1) to diagnosis and predict the survival time of BC patients using a cluster of 12 DEGs. Additionally, the high infiltration of macrophages M2 and high sensitivity to Bortezomib in high risk group scored by our risk models provided theoretical support for further basic and clinical research on the tumor immune microenvironment of BC.

Introduction

Breast cancer (BC) is a major health problem worldwide and has been considered as the leading cause of cancer related deaths in females, accountable for approximately 14% of all cancer deaths (1). While in men, breast cancer is a rare disease, the incidence of breast cancer is much lower than in women, but the mortality rate of males with BC is the quadruple of that in females (2), worthy calling attention to the matter. According to statistics, total new BC cases in 2009 was 194280 in the US and the BC related deaths was 40610 (3), while a decade later, the number has increased to 268670 and 41400, respectively (2).

To date, there have been multiple advances in the diagnosis of BC, which might result in the rising incidence of BC. The self- examination of BC is considered to be useful in diagnosis in early stages and reduction of early deaths (4), but still numerous cases diagnosed in their advanced stages due to the

negligence of women regarding the self-inspection and clinical examination of the breast. Screening examinations are gradually becoming routines for BC diagnosis, including ultrasound breast imaging, single photon emission computerized tomography, and digital mammography. Mammography is considered as the gold standard test for early detection of breast cancer (5), while there are debates like frequent screening examinations might cause radical harm, the false-positive results and overdiagnosis, which lead to a negative impact on a woman's life(6). Biomarkers of BC are also used in diagnosis during different stages, including immunohistochemical markers (ER, PR, HER2 [ERBB2], and proliferation marker protein Ki-67 [MKI67]), genomic markers (BRCA1, BRCA2, and PIK3CA), and immunomarkers (tumor-infiltrating lymphocytes and PD-L1) (7). At the same time, progressions in the treatment have been made, whereas the mortality rate doesn't decrease, indicating that new biomarkers or models for BC survival prediction and prognosis should be established to help provide more efficient and personalized measures handle with BC.

In our study, we identified two characteristic gene sets of BC development, BC unfavorable gene set and BC favorable gene set. Survival analysis and ROC analysis were used respectively to determine the diagnostic and prognostic abilities of the two gene set models scores. Both the BC unfavorable and the BC favorable gene set models scores are reliable biomarkers for the BC diagnosis and independent biomarkers for prediction and prognosis.

Materials And Methods

Datasets

The gene expression profiles of BC and healthy breast samples were obtained from The Cancer Genome Atlas (TCGA), containing 173 stage□BCs, 597 stage□BCs, 240 stage □ BCs, 20 stage □BCs and 113 normal samples. TCGA is an international database publicly accessible and freely available for researches, containing a large collection of human cancer genome sequencing data. Previous studies using a group of genes with known alterations in BC to validate TCGA data suggested that the TCGA had accurately documented the genomic abnormalities of multiple malignancies (8).

Identification of differentiated expressed genes (DEGs)

Preliminary identification of the characteristic genes promoting or inhibiting the development of BC was performed using Differential Expression Analysis and Short-Time Series Expression Miner (STEM) analysis. The RNA sequencing expression profile of BC was displayed as read counts, which was subsequently normalized by voom function in limma package. $P < 0.01$ adjusted by the false discovery rate (FDR) and $|\log \text{fold change (FC)}| > 1.5$ were set as a threshold. In the developing of BC, if a DEG was gradually upregulated (log FC stage I vs. control < log FC stage II vs. control < log FC stage III vs. control < log FC stage IV vs. control) or gradually downregulated (log FC stage I vs. control > log FC stage II vs. control > log FC stage III vs. control > log FC stage IV vs. control), and then it was considered to be BC-development characteristic gene.

STEM is a new software package for analyzing short time series expression data. The software can both find statistically significant patterns from short time series microarray experiments and compare data sets across experiments. STEM presents its analysis of the data in a highly visual and interactive manner (9).

Limma package

DEGs in the four stages of BC were identified separately using the limma package with RNA sequencing expression profiles of BC in TCGA. The package can now perform both differential expression and differential splicing analyses of RNA sequencing (RNA seq) data. This software provides an integrated data analysis solution, using advanced computational algorithms to deliver reliable performance on large data sets and represent expression data and simplify the user interface (10). Significant differential genes were selected by using the FDR and log (FC) values. In the development of BC, if the DEGs were gradually up-regulated (log (FC) value raised) or gradually down-regulated (log (FC) decreased), they were then considered to promote the BC development or suppress the BC development.

Cox regression analysis

Univariate Cox regression analysis was used to identify genes associated with BC prognosis. Multivariate Cox regression analysis was used to testify the correlation between our models, present clinicopathological features and BC prognosis.

Lasso regression analysis

The most refined prognostic prediction models were constructed by minimal compression using Lasso regression analysis based on two characteristic gene sets of BC development. With a constraint condition in the sum of the absolute terms, LASSO could reduce dimension of independent variable and produces a better model fit, it would be suitable for analyzing the genetic survival data, finding out more meaningful independent variables and model determination. The fixed coefficient R2 indicates the fraction of the variation that the covariate explains. The larger the percentage was, the better the model fit (11).

Calculation of BC-Development Characteristic GSVA Score

Both the two models based on up-regulated and down-regulated characteristic gene sets of BC development were scored. Gene set variation analysis is a popular method of scoring individual samples for molecular characteristics or gene sets. GSVA package in R was used to calculate BC-unfavorable GSVA score and BC-favorable GSVA score for individual samples.

We screened the manifestation levels for genes in the TCGA data group using univariate Cox regression analysis. Thus, we identified genes that were highly related to survival ($P < 0.01$). We chose these genes for more functional investigation and developed possible risk score using the LASSO Cox regression algorithm. Finally, minimum criteria defined genes and their constants, choosing the perfect penalty parameter λ related to the minimum 10-fold cross validation inside the training set. Using the following formula, the risk score was calculated:

Evaluation of the models

Receiver Operating Characteristic Curve (ROC) analysis and survival analysis were used to explore the diagnostic and prognostic capabilities of the two scoring systems. In ROC curve analysis, the area under the curve provides an objective parameter of the diagnostic or prognostic accuracy of a test, which is superior to comparing single combinations of sensitivity and specificity values, since the influence of the threshold value is eliminated. Because only part of the ROC curve represents clinically relevant combinations of sensitivity and specificity, comparing the ROC curves in the relevant sensitivity or specificity ranges is to be preferred over comparing the total area under the curve. In addition, ROC analysis can be used to determine the optimal threshold value for tests that generate continuous quantitative data (12). Multivariate Cox regression analysis was used to compare the relative prognostic value of the two score systems with that of routine clinicopathological features. Moreover, Correlation analysis was used to explore the correlation between these models and tumor immune microenvironment and drug responses.

Results

Identification of gradually up/down-regulated genes

Compared to normal breast tissue samples, a total of 5770 DEGs in stage I BCs, 6083 DEGs in stage II BCs, 5966 DEGs in stage III BCs, and 5855 DEGs stage IV BCs, a total of 4656 common DEGs were identified in stage I-IV BCs (Fig. 1A-B). We summarized 10 major variation tendencies of gene expression by STEM analysis in BC gene expression profiles in TCGA (Fig. 1C). Based on these tendencies, DEGs were classified into two groups: one group (BC1 cluster) containing gradually up-regulated genes with the development of BC, these genes showed an up-regulated tendency during the stage from □to □ (Fig. 1D-E), indicating the progression of BC, there were totally 476 up-regulated DEGs, named BC unfavorable gene set; while the second group (BC2 cluster) containing gradually down-regulated genes with the development of BC (Fig. 1F-G), which meant good prognosis, containing totally 262 down-regulated DEGs, named BC favorable gene set.

To identify BC DEGs associated with prognosis, we first used a univariate Cox regression analysis to assess the association between the expression levels of each DEG, and found that 26 genes were significantly associated with BC prognosis in BC unfavorable gene set (Supplemental Table 1; $P < 0.01$) and 14 genes in BC favorable gene set (Supplemental Table 2; $P < 0.01$).

Establishment and Evaluation of Prognostic Model Based on BC DEGs

Lasso regression analysis was performed to establish risk models concerning BC prediction and prognosis based on two characteristic gene sets of BC development. The results showed a total of 12 DEGs (Fig. 2A-D) of BC unfavorable gene set and 7 DEGs (Fig. 2E-H) of BC favorable gene set were

significantly associated with BC prognosis. On the basis, two risk models were obtained, consisting of 12 genes and 7 genes respectively, named BC unfavorable model (BC1 model) and BC favorable model (BC2 model).

The ROC analysis was performed to evaluate the prognostic efficacy of two risk models. In BC1 model, the AUC was 0.80 (Fig. 2I), presenting great capability of the model for BC prognosis; while in BC2 model, the AUC was 0.66 (Fig. 2L). Compared with the BC2 model, BC1 model was more efficient in prognostic capability. We also carried out ROC analysis on BC patients for 1, 2, 3 years after diagnosis, the AUC was 0.80, 0.77, and 0.75, respectively in BC1 model (Fig. 2J), and 0.66, 0.69, 0.74 in BC2 model (Fig. 2M). The results showed that both BC1 and BC2 models could accurately predict the survival rate of BC patients, especially for BC1 model.

Additionally, we compared the efficacy of risk score, age, stage, T, M and N for BC prognosis. In BC1 model, the AUC of risk score was 0.80, which remained the highest among these clinical characteristics, illustrating that BC1 model was more efficient than other clinical factors for prognosis (Fig. 2K). Whereas in BC2 model, the AUC of risk score was 0.64, which was lower than those clinical characteristics, including 0.71 in age, 0.64 in stage, 0.70 in T, 0.55 in M and 0.62 in N (Fig. 2N).

Survival analysis was performed to evaluate the survival possibility of BC patients divided into high risk group and low risk group based on these two models. In BC1 model, the results of survival analysis manifested that, the high risk group survival possibility was lower than low risk group, and the gap between two groups getting much wider with the survival years getting longer ($P < 0.001$; Fig. 3A). The final survival time of high risk group was less than 22 years, shorter than the low risk group, which was 23.5 years. In the scattergraphs, we ranked BC patients by the risk scores evaluated with our model from low scores to high scores in the horizontal axis. As a result, the vertical axis showed their survival time, green scatters represented alive cases and red represented dead, the graphs demonstrated that more dead cases appeared in patients with high risk scores and more alive cases appeared in patients with low risk scores, the survival time of patients with low risk scores was longer than that in patients with high risk scores. As shown in Fig. 3B, the survival time in high-risk group was significantly lower than that in low-risk group using BC1 model. In BC2 model, the results were similar (Fig. 3C-D), indicating that these two models can accurately predict the survival rate of BC patients.

Association of risk score from BC models with clinical characteristics

Multivariate Cox regression analysis and Univariate Cox regression analysis were used to identify factors independently associated with BC prognosis. For both BC1 and BC2 models, *P* values of age, N and risk score were all less than 0.001, suggesting the factors age, lymphatic metastasis and the risk score were all independently associated with BC prognosis. These findings demonstrated that the risk score derived from our model was an independent prognostic predictor of BC patients (Fig. 3E-H).

The performance of the risk scores was tested with each clinical characteristic. Significant differences were observed between risk score and age ($P < 0.001$; Fig. 3I) in BC1 model, while risk score and tumor topography was significantly correlated ($P < 0.001$; Fig. 3J) in BC2 model.

In the heatmap, there were obviously more aged (> = 60 years) in patients with high risk scores, yet there was no distinct regularity between risk scores and other features, which was consistent with the results of multivariate Cox regression analysis that age was independent factor correlated with BC1 model (Fig. 3I). In BC2 model, marked difference was observed between tumor topography and the risk score (Fig. 3J).

Immune infiltration analysis and drug susceptibility

To further explore the immune-related mechanism involved in BC patients, we performed the immune infiltration analysis using two BC prognosis models. It is strongly suggested that M2 macrophage was significantly correlated with BC1 and BC2 models (Fig. 4A-B). Moreover, more Macrophage M2 infiltration was remarkably higher in high risk group than in low risk group (Fig. 4C-D).

For BC chemotherapy, Bortezomib is a first-in-class selective and reversible proteasome inhibitor that targets the 26S proteasome (13). Previous studies have observed that Bortezomib combined with antiestrogen therapy might have therapeutic advantage in the management of early-stage breast cancer (14). Our study also found out that, patients in high risk group were more sensitive to Bortezomib than low risk group (Fig. 4E-F).

To further test these two models in cohorts from different populations, we assessed the testing power of two models of breast cancer development-related gene sets in GSE4922. The AUC of BC1 model was 0.79, higher than that of BC2 model, implying that the testing power of BC1 model was stronger than BC2 model (Fig. 4G-H).

Discussion

In the world, especially for women, breast cancer is the main cause of cancer-related death(15). Even with gradually progressing diagnosis and surgical treatments, the recurrence rate and cancer-related death rate of breast cancer are still very high. Hence, it is of great urgency to explore new biomarkers which can precisely diagnose breast cancer and predict prognosis for the treatment and management of breast cancer. In this study, we have identified two prognostic models (BC1 and BC2) using bioinformatics analysis on the basis of BC-related DEGs, and further explored the prognostic value of BC survival time and mechanism involved. Moreover, one model (BC1) was observed to derive a better prognostic risk factor, which classifies the survival time in BC patients into low and high-risk categories.

An amount of previous studies have shown that abnormal expression of genes in breast cancer is closely-related to disease prognosis, and can be used as a potential biomarker of prognosis (16, 17). In the present study, we found a number of genes were differentially expressed in BC different stages. This indicated gene expression patterns varied along with the BC development. Compared to normal breast tissue, a gene may be differentially expressed in early BC but not in advanced stage. We identified 738

BC-development characteristic genes, including 476 genes gradually up-regulated and 262 genes gradually down-regulated with BC-development. The development of BC results from interactive effects of multiple genes. Prominently, not all BC development characteristic genes are associated with the prognosis of BC. BC-unfavorable gene set comprised 26 gradually up-regulated DEGs and BC-favorable gene set contained 14 gradually down-regulated DEGs. Unsurprisingly, previous studies have suggested that some of them are associated with BC development (18–20). These results confirmed the possibility that the BC unfavorable gene set and BC favorable gene set could be used as a prognostic model for BC.

In the previous studies, a gene often got a coefficient from a Cox regression analysis or other method in the training set (21). However, due to the limitations of the sample size and the heterogeneity of the tumor, we may never know the true coefficient of a gene. Therefore, GSVA was used to score individual samples against gene sets (BC-unfavorable gene set and BC-favorable gene set) in our study. ROC curve analysis suggested that both BC-unfavorable GSVA score and BC-favorable GSVA score exhibited strong prognostic capacity of BC and which was verified in other two independent data sets. Univariate and multivariate Cox regression analysis suggested that BC unfavorable, BC favorable gene set and the risk score systems were independent prognostic factors for BC. This result was also verified in an independent data set. However, further studies are needed to investigate and validate the functions of these genes, the synergy between the genes of these two gene sets concerning BC development still requires molecular experimental validation. Their application to the clinic still needs to wait for further decline in sequencing costs.

Besides, the interaction between immune microenvironment and cancer cells is important for tumor progression (22). Both BC development-promoting gene model and BC development inhibitory gene model were associated with high infiltration of M2 macrophages. Both the BC development-promoting gene model and the BC development-suppressing gene model are related to patient drug responsiveness to Bortezomib. Naturally, studies aimed at revealing the exact mechanisms and DEGs in BC are advocated. The potential of molecularly targeted immunotherapy to intervene breast cancer may be promising.

Conclusion

In conclusion, this study presents and validated one prognosis model (BC1) to diagnosis and predict the survival time of BC patients using a cluster of 12 DEGs, and extensive functional exploration revealed the model closely related to tumor immune microenvironment, especially M2 macrophage infiltration. Our study provided new insights into further studies in BC, which requires further research.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors have reviewed the manuscript and agreed for the publication. Written informed consent was obtained from each participant.

Availability of data and materials

The datasets analyzed were acquired from The Cancer Genome Atlas (TCGA) database (https://portal.gdc.cancer.gov/) and GEO database (http://www.ncbi.nlm.nih.gov/geo/).

Competing interests

There is no conflict of interest in this manuscript.

Funding

None.

Authors' contributions

Ying Yu, Chen Jiaying*, and Yao Chang prepared for the manuscript

Ying Yu and Bian Weihe performed the statistical analysis

Ying Yu, Cong Wang, Ye Bei performed the bioinformatics analysis

Shen Tong, Guo Mengmeng, and Zhang Xiping prepared for Figure 1-3

Cao Shihan prepared for Figure 4

Ma Chaoqun revised the manuscript

Acknowledgements

Not applicable.

References

1. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 2012;36(3):237-48.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30.
3. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin. 2009;59(4):225-49.

4. Kosters JP, Gotzsche PC. Regular self-examination or clinical examination for early detection of breast cancer. Cochrane Database Syst Rev. 2003(2):CD003373.
5. Sardanelli F, Giuseppetti GM, Panizza P, Bazzocchi M, Fausto A, Simonetti G, et al. Sensitivity of MRI versus mammography for detecting foci of multifocal, multicentric breast cancer in fatty and dense breasts using the whole-breast pathologic examination as a gold standard. American Roentgen Ray Society. 2004;183:1149–57.
6. Broeders M, Paci E. The balance sheet of benefits and harms of breast cancer population-based screening in Europe: outcome research, practice and future challenges. Womens Health. 2015;11(6): 883–90.
7. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. The Lancet. 2021;397(10286):1750-69.
8. Ping Z, Siegal GP, Almeida JS, Schnitt SJ, Shen D. Mining genome sequencing data to identify the genomic features linked to breast cancer histopathology. J Pathol Inform. 2014;5(1):3.
9. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics. 2006;7:191.
10. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43(7):e47-e.
11. Lina Y, Ting Q, Tong W. The Application of LASSO in the Cox Model. Chinese Journal of Health Statistics. 2012;29(1).
12. Pattynama ARvEPMT. Receiver operating characteristic (ROC) analysis: Basic principles and applications in radiology. European Journal of Radiology. 1998;27:88-94.
13. Tan CRC, Abdul-Majeed S, Cael B, Barta SK. Clinical Pharmacokinetics and Pharmacodynamics of Bortezomib. Clin Pharmacokinet. 2019;58(2):157-68.
14. Periyasamy-Thandavan S, Jackson WH, Samaddar JS, Erickson B, Barrett JR, Raney L, et al. Bortezomib blocks the catabolic process of autophagy via a cathepsin-dependent mechanism, affects endoplasmic reticulum stress and induces caspase-dependent cell death in antiestrogen-sensitive and resistant ER+ breast cancer cells. Autophagy. 2010;6(1):19-35.
15. Ahmad A. Breast Cancer Statistics: Recent Trends. Adv Exp Med Biol. 2019;1152:1-7.
16. Joe S, Nam H. Prognostic factor analysis for breast cancer using gene expression profiles. BMC Med Inform Decis Mak. 2016;16 Suppl 1:56.
17. Mranda GM, Xue Y, Zhou XG, Yu W, Wei T, Xiang ZP, et al. Revisiting the 8th AJCC system for gastric cancer: A review on validations, nomograms, lymph nodes impact, and proposed modifications. Ann Med Surg (Lond). 2022;75:103411.
18. Singh Y, Subbarao N, Jaimini A, Hathaway QA, Kunovac A, Erickson B, et al. Genome-wide expression reveals potential biomarkers in breast cancer bone metastasis. J Integr Bioinform. 2022.
19. Du Y, Han Y, Wang X, Wang H, Qu Y, Guo K, et al. Identification of Immune-Related Breast Cancer Chemotherapy Resistance Genes via Bioinformatics Approaches. Front Oncol. 2022;12:772723.

20. Hou L, Hou S, Yin L, Zhao S, Li X. Epithelial-Mesenchymal Transition-Based Gene Signature and Distinct Molecular Subtypes for Predicting Clinical Outcomes in Breast Cancer. Int J Gen Med. 2022;15:3497-515.
21. Hao S, Huang M, Xu X, Wang X, Huo L, Wang L, et al. MDN1 Mutation Is Associated With High Tumor Mutation Burden and Unfavorable Prognosis in Breast Cancer. Front Genet. 2022;13:857836.
22. Ni Y, Tsang JY, Shao Y, Poon IK, Tam F, Shea KH, et al. Combining Analysis of Tumor-infiltrating Lymphocytes (TIL) and PD-L1 Refined the Prognostication of Breast Cancer Subtypes. Oncologist. 2022;27(4):e313-e27.

Figures

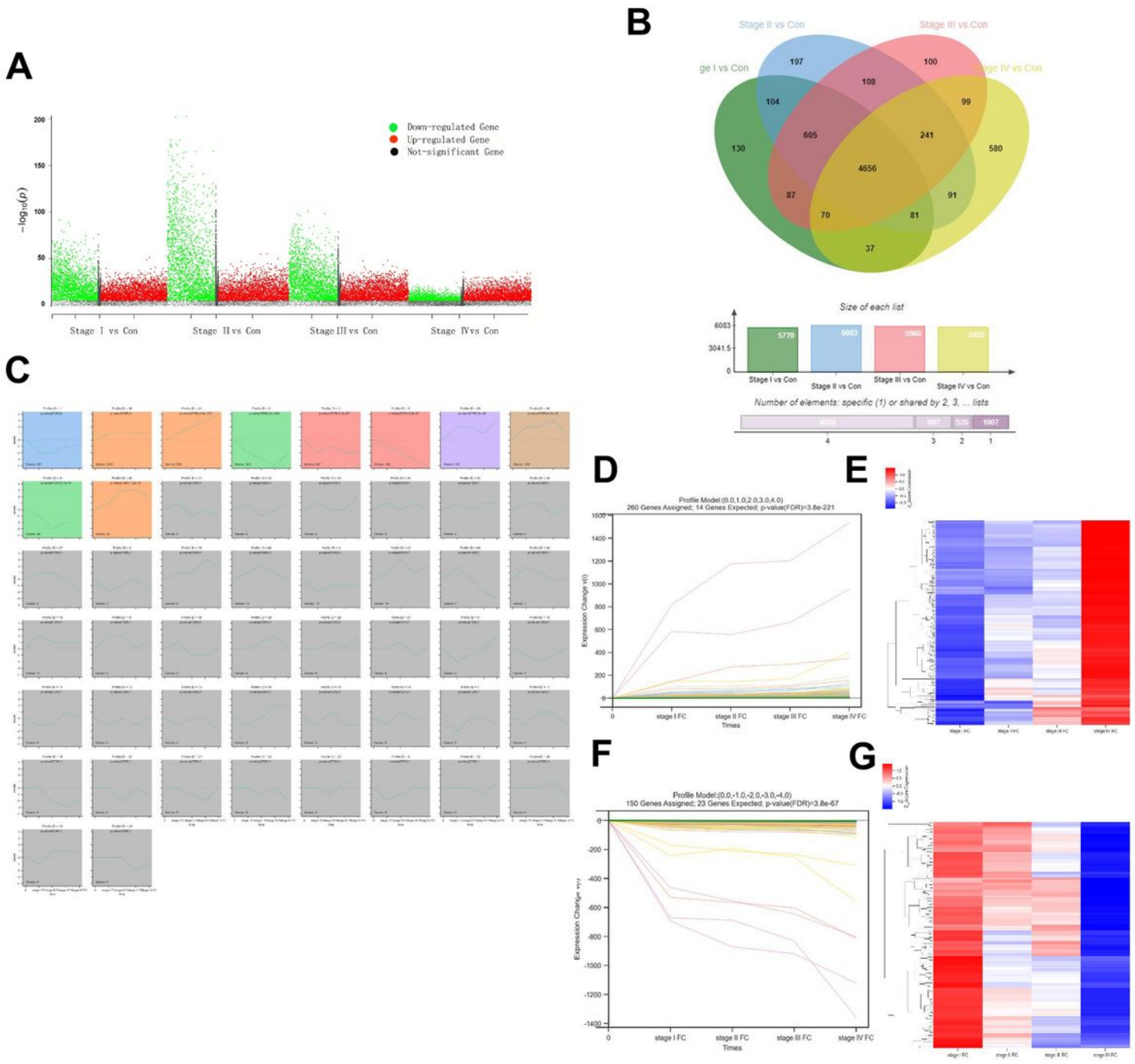

Figure 1

Identification of differentially expressed genes (DEGs) in different stage of breast cancer (BC). (A) Manhattan plot showed differentially expressed genes (DEGs) in different stage of breast cancer (BC); (B) Venn diagram showed common DEGs in BC stage I–IV; (C) 10 major variation tendencies of gene expression by STEM analysis; (D-E) Up-regulated genes with the development of BC; (F-G) Up-regulated genes with the development of BC.

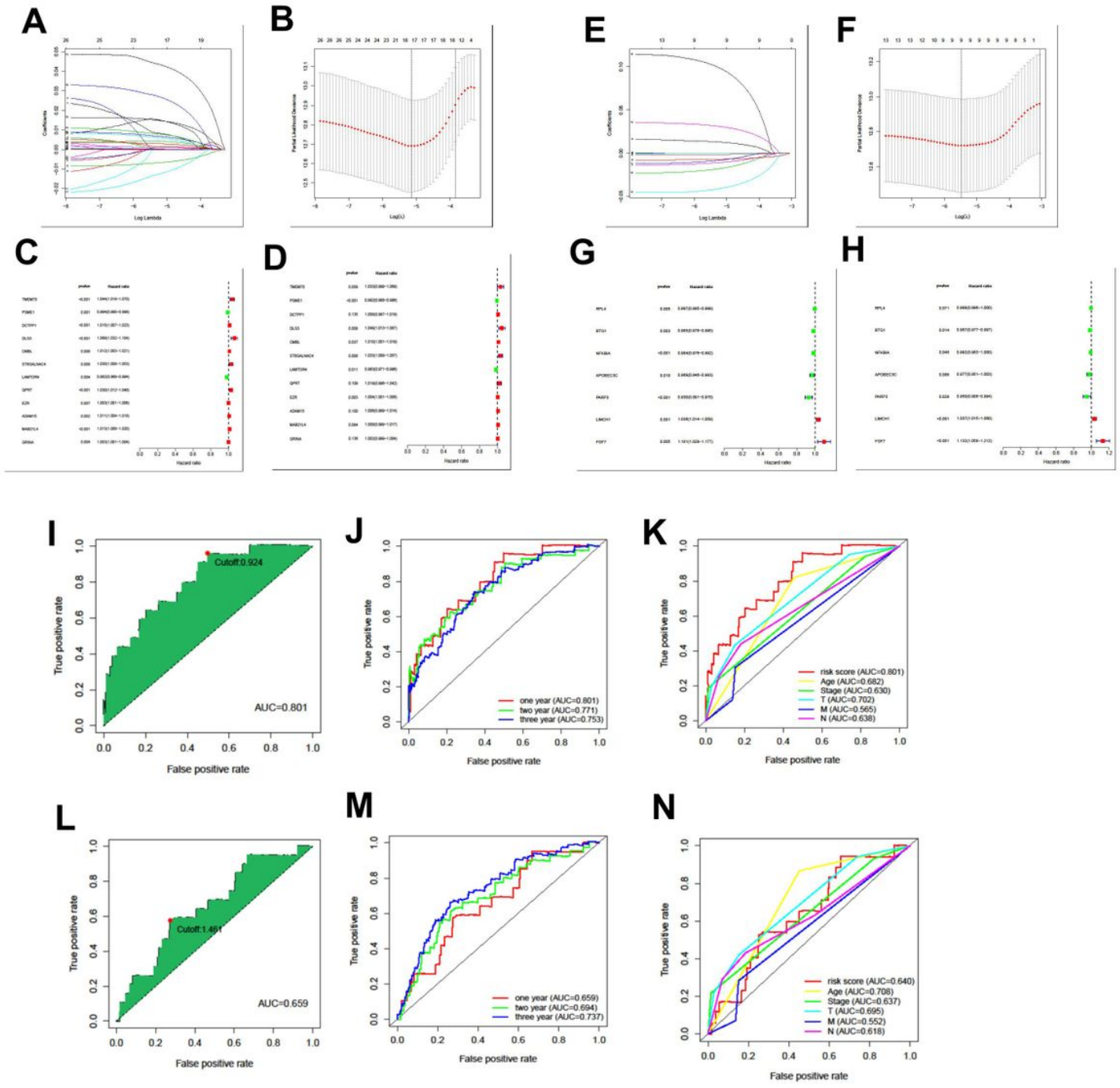

Figure 2

Establishment and Evaluation of Prognostic Model Based on BC DEGs. (A-B) Essential parameter for the establishment of BC1 model. (C-D) Forest plots of significant DEGs in BC1 model using uultivariate Cox regression analysis and multivariate Cox regression analysis. (E-F) Essential parameter for the establishment of BC2 model. (G-H) Forest plots of significant DEGs in BC2 model using uultivariate Cox regression analysis and multivariate Cox regression analysis. (I) Evaluation of prognostic efficacy in BC1 model by ROC analysis. (J) Evaluation of prognostic efficacy in BC1 model on the survival rate of 1, 2, and 3 years by ROC analysis. (K) Comparison of prognostic efficacy of BC1 model and clinical variables

by ROC analysis. (L) Evaluation of prognostic efficacy in BC2 model by ROC analysis. (M) Evaluation of prognostic efficacy in BC2 model on the survival rate of 1, 2, and 3 years by ROC analysis. (N) Comparison of prognostic efficacy of BC2 model and clinical variables by ROC analysis.

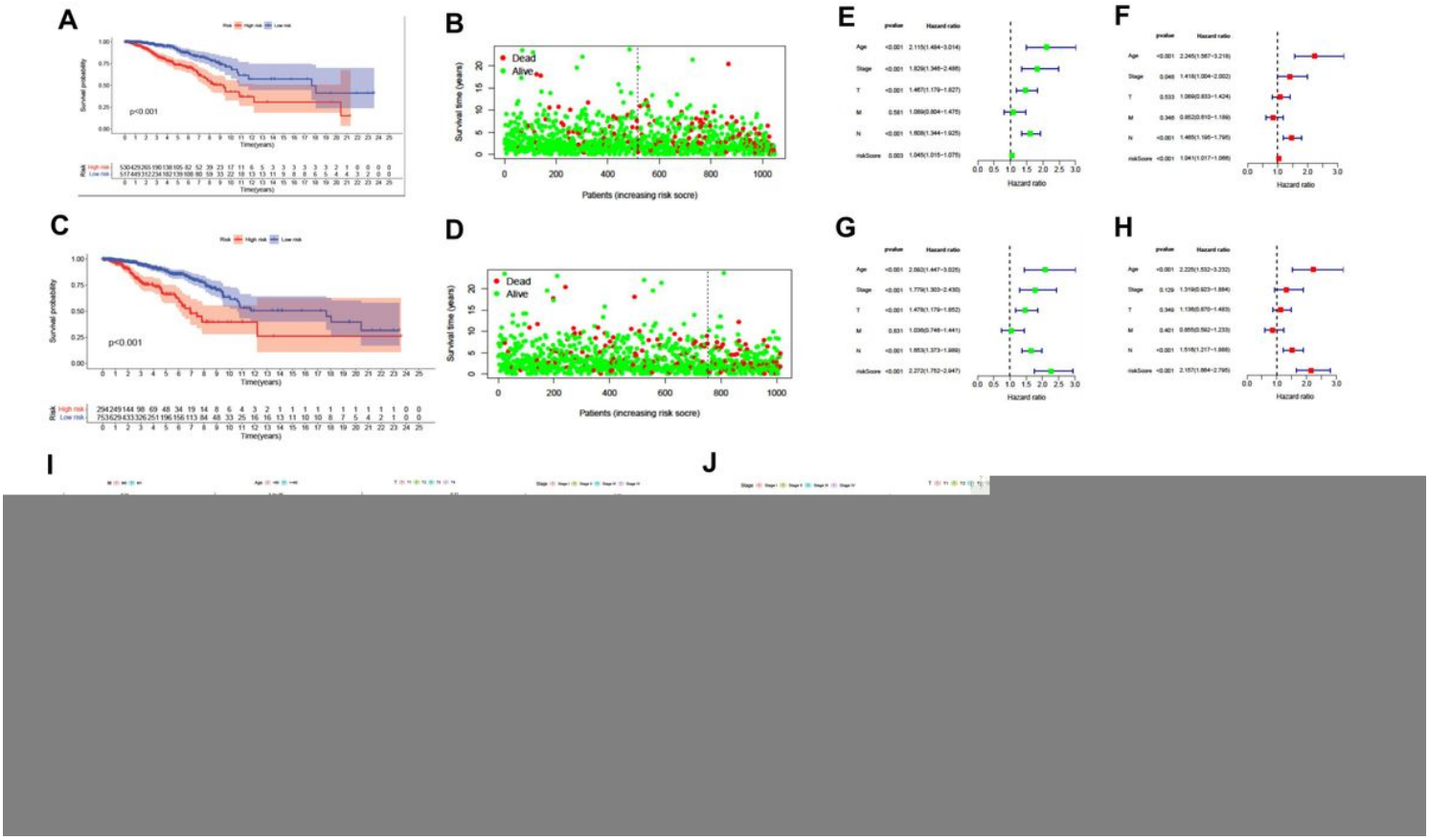

Figure 3

Association between BC models and clinical characteristics. (A) ROC analysis for the survival rate of BC patients using BC1 model. (B) scattergraph of risk scores derived from BC1 model to evaluate the survival rate. (C) ROC analysis for the survival rate of BC patients using BC2 model. (D) scattergraph of risk scores derived from BC1 model to evaluate the survival rate. (E-F) Univariate Cox regression analysis (E) and Multivariate Cox regression analysis (F) for BC1 model to explore the association between clinical characteristics and BC prognosis. (G-H) Univariate Cox regression analysis (G) and Multivariate Cox regression analysis (H) for BC1 model. (I-J) The comparison and heatmap of the distribution of clinicopathological features compared between the low-risk and high-risk groups with BC1 model (I) and BC2 model (J).

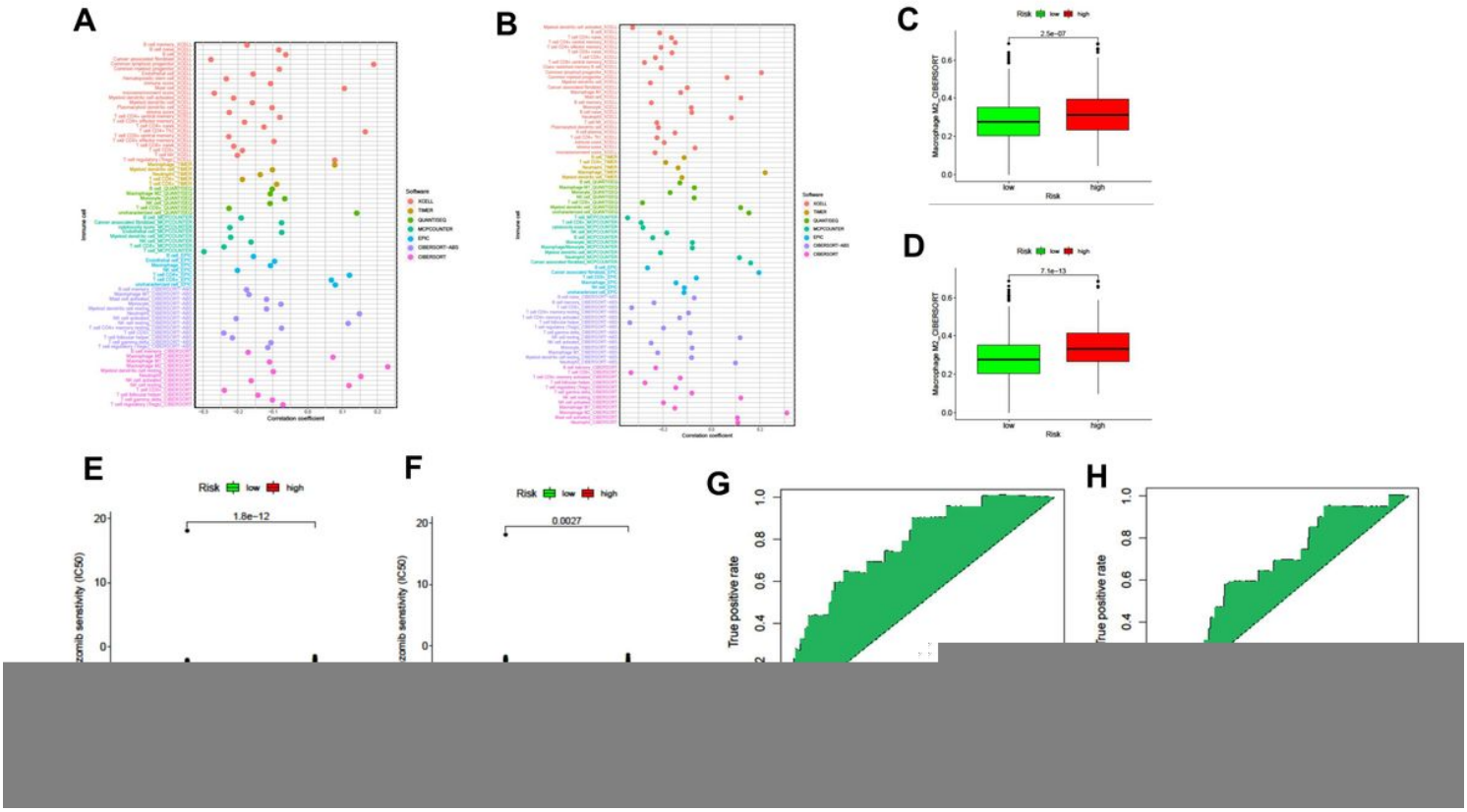

Figure 4

Immune infiltration analysis and drug susceptibility to Bortezomib. (A-B) Immune infiltration analysis using BC1 (A) and BC2 (B) models. (C-D) Quantitative analysis of M2 infiltration between low-risk and high-risk groups using BC1 (C) and BC2 (D) models. (E-F) Bortezomib susceptibility test of low-risk and high-risk groups using BC1 (E) and BC2 (F) models. (G-H) Validation of the prognostic efficacy of BC1 (G) and BC2 (H) models in GSE4922.

Supplementary Files

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

- SupplementalTable1.docx
- SupplementalTable2.docx