Identification of a 9-Gene Prognostic signature for breast cancer

Zelin Tian
Wuhan University Zhongnan Hospital https://orcid.org/0000-0002-3450-2469
Jianing Tang
Wuhan University Zhongnan Hospital
Xing Liao
Wuhan University Zhongnan Hospital
Qian Yang
Wuhan University Zhongnan Hospital
Yumin Wu
Wuhan University Zhongnan Hospital
Gaosong Wu (wugaosongtj@126.com)
https://orcid.org/0000-0002-7955-4537

Primary research

Keywords: Breast cancer, overall survival, risk score, GEO, prognostic signature

Posted Date: May 12th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-26949/v1

Abstract

Background Breast cancer (BRCA) is the most common cancer among women worldwide and results in the second leading cause of woman cancer death.

Methods This study sought to develop a prognostic gene signature to predict the prognosis of patients with BRCA. Studies were performed using the genome-wide data of BRCA patients from the Gene Expression Omnibus dataset (GSE20685, GSE42568, GSE20711, GSE88770). Univariate COX regression analysis was used to determine the association between gene expression levels and overall survival(OS) in each dataset. Taking P value < 0.05 as the inclusion criterion, the common genes in all datasets were selected as prognostic genes, and a 9-gene prognostic signature was developed.

Results The Kaplan-Meier survival curve was constructed using log-rank test to assess survival differences. The overall survival of patients in the low-risk group was significantly higher than that in the high-risk group. ROC analysis showed that this 9-gene signature showed good diagnostic efficiency both in overall survival(OS) and disease free survival(DFS). The 9-gene signature was further validated using GSE16446 dataset. In addition, multiple Cox regression analysis showed that this 9-gene signature was an independent risk factor. Finally, we established a nomogram that integrates conventional clinicopathological features and 9-gene signature. The analysis of the calibration plots showed that the nomogram has good performance.

Conclusions This study has developed a reliable 9-gene prognostic signature, which is of great value in predicting the prognosis of BRCA and will help to make personalized treatment decisions for patients at different risk score.

Background

Breast cancer is the most commonly diagnosed cancer in women and the 2018 global cancer statistics showed that there will be approximately 2.1 million newly diagnosed female breast cancer cases in 2018, accounting for one quarter of total cancer cases among women[1]. The molecular typing of breast cancer is closely related to the prognosis. In 2011, at the St. Gallen International Breast Cancer Conference, breast cancer was classified into Luminal A, Luminal B, HER2 positive and triple negative breast cancer (TNBC) four major parts based on the detection results of ER, PR, HER2 and Ki67 [2]. Among these four molecular types, Luminal A breast cancer is the most common molecular subtype. The results of research by Ihemelandu et al. showed that Luminal A type accounts for 50% of breast cancer patients, Luminal B, HER2 positive and TNBC types accounted for 14.1%, 12.7%, and 23.2%, respectively[3]. These subtypes show different histopathological characteristics and therapeutic sensitivities. Patients with Luminal A breast cancer usually have a better prognosis, while patients with TNBC have the worst prognosis[4]. For ER receptor-positive breast cancer, tamoxifen and aromatase inhibitors are effective endocrine therapies[5]. For HER2-positive breast cancer, the monoclonal antibody trastuzumab and lapatinib have been approved as the most specific molecular targeting drugs[6, 7]. Common therapies for BRCA include

surgical resection, chemotherapy, endocrine therapy, radiation therapy and molecular targeted therapy. With the development of medical technology, the prognosis of BRCA has been significantly improved. However, the prognosis of advanced breast cancer is still not optimistic[8].

In recent years, detailed information on prognostic assessment of cancer patients has been available through microarray analysis and whole-genome sequencing[9, 10]. Genomics and molecular characteristics research has significantly improved the biological understanding of breast cancer, elucidated the inherent molecular subtypes and genetic driving mechanisms of breast cancer[10, 11]. Many studies have further demonstrated the prognostic role of gene expression characteristics based on tumor arrays[12–14]. However, most are not clinically used. Therefore, it is urgent to identify novel and practical gene signatures to predict the prognosis of patients. In this study, we have mined from several published datasets of GEO to provide reliable prognostic signals for breast cancer. 9-gene prognostic signature were eventually identified, providing hope for more personalized treatment interventions for patients..

Materials And Methods

Data processing

The work flow of data acquisition, preprocessing, gene signature generation and verification was shown in Figure 1. The original gene expression data and clinical information were obtained from the GEO database. We used GSE20685, GSE42568, GSE20711, GSE88770 four independent datasets for analysis. GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array was used for data collection. After removing incomplete clinical information and cases of normal samples, a total of 622 cases were included in this analysis (327 cases in GSE20685 dataset, 101 cases in GSE42568 dataset, 85 cases in GSE20711 dataset, and 109 cases in GSE88770 dataset). In this study, the prognostic genes related to the overall survival were first selected, then the common prognostic genes (9 genes) in four datasets were selected as the prognostic gene signature, which were then developed using a risk scoring model and verified using four datasets and the entire cohort.

Prognostic signature

A univariate Cox proportional hazard regression model was used to screen genes associated with prognosis in each dataset. Hazard ratios | HR |> 1 and P-value <0.05 (P-value <0.05 in GSE20685、GSE20711 and GSE88770; P-value <0.01 in GSE42568) were used to screen candidate genes related to OS from each dataset. Genes with HR <1 are considered as protect genes, and genes with HR> 1 are considered as danger genes. To improve reliability, only the common genes in the four datasets were used as prognostic gene signature. Finally, we constructed a 9-gene prognostic signature

Validaton of signature

Each patient's risk score was constructed through the regression coefficient weighted prognostic gene expression values. Patients were divided into high-risk and low-risk groups based on the median risk score. This study using four GEO datasets(GSE20685, GSE20711, GSE88770 and GSE42568) for internal validation and GSE16446 datasets for external validation.

Construction and verification of the nomogram

In this study, a "rms" R package was used to generate a nomogram of each dataset containing clinical information and 9-gene signature. The C index and calibration plots were used to evaluate the accuracy of nomogram. The prediction efficiency of the nomogram is shown in the calibration plots, where the 45 ° dotted line indicates the best prediction.

Statistical analysis

The Kaplan-Meier method was used to evaluate the differences in OS and DFS in patients with low-risk and high-risk group, and log-rank tests were used to evaluate the statistical significance of the differences between groups. Multivariate Cox regression analysis and stratified analysis were used to assess whether the 9-gene signature was independent of other clinical characteristics. The "survivalROC" R package was used for time-dependent receiver operating characteristic (ROC) analysis, and the prognostic performance was verified by comparing the area under the ROC curve (AUC). $P < 0.05$ was considered statistically significant. All statistical tests were performed by R software (version 3.6.1).

Gene set enrichment analysis

According to the median risk value, 327 BRCA samples in the dataset GSE20685 were divided into high-risk and low-risk groups. We used the GSEA software (GSEA version 4.0.3) to perform a gene set enrichment analysis (GSEA) on high and low risk groups. The c2.cp.kegg.v6.2.symbols.gmt gene set was selected as the reference gene set. The most significant first 5 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were screened (FDR <0.05).

Result

Prognostic signature generation

In this study, in order to identify candidate prognostic genes that are significantly associated with OS, we performed univariate Cox proportional hazard regression analysis on each data. Using P <0.05 and HR <1 as the cutoff criteria, 1797 genes in GSE26085, 895 genes in GSE42568, 450 genes in GSE20711, and 666 genes in GSE88770 were identified as candidate protect genes. Using P <0.05 and HR> 1 as the cutoff criteria, there were 2528 genes in GSE26085, 1771 genes in GSE42568, 766 genes in GSE20711, and 1292 genes in GSE88770 were identified as candidate danger genes. Finally, the common genes in four datasets were retained as prognostic genes. Two protect genes (STXBP3, PKN2) and seven danger genes (TCAP, STARD3, CDR2L, PNMT, GPR4, ANGPT2, CAPN5) were finally obtained. Both them are mRNAs, and general information for these genes is shown in Table 1. The prognostic correlations of the 9-genes in each data set with respect to overall survival are shown in Table 2 (HR value, 95% confidence interval, P value).

9-gene prognostic signature validation

Regression coefficient weighted prognostic gene expression values were used to construct a risk score for each patient and a prediction model was constructed to predict overall survival of patients. Calculate the risk score for each patient, and then divide patients into high-risk and low-risk groups based on the median risk score. In Figure 2, the ranking is based on the risk score values of 9 gene signatures from low to high, the risk score distribution, risk gene expression and patient survival status of the four data sets GSE20685, GSE42568, GSE20711, and GSE88770 are shown respectively.

In Figure 3, Kaplan-Meier survival curves and ROC curves generated from 9-gene signature in four data sets are shown. Outcome data indicate that the overall survival of patients in the high-risk group is shorter than in the low-risk group(GSE20685: HR=2.68(1.67-4.29), P value=1.99e-05 ; GSE42568: HR=9.75(3.42-27.79), P value=1.82e-07; GSE20711: HR=4.38(1.74-11.02), P value=6.38e-04; GSE88770: HR=3.26(1.42-7.51), P value=3.35e-03) (Figure 3 Left panel). For further verification, we divided patients into three groups: high, medium, and low risk based on the value of the risk score. The results still show that the higher the risk score, the worse the patient's overall survival (Figure 3 Middle panel). Finally, we constructed a ROC curve based on 9-gene signature. The results show that, as time went on, the AUC values of the four datasets have remained at a relatively satisfactory value, can effectively predict overall survival (Figure 3 Right panel).

Of these 9 genes, two of them are protect genes (STXBP3, PKN2; HR <1) and seven of them are danger genes (TCAP, STARD3, CDR2L, PNMT, GPR4, ANGPT2 and CAPN5; HR> 1). Figure 4 showed the difference in expression of these 9 prognostic genes in the low-risk and high-risk groups. The results

showed that high-risk group patients had higher danger gene expression, while low-risk group patients had lower protect gene expression.

9-gene prognostic signature is independent of other clinicopathological factors

Multiple Cox regression analysis was used to assess whether 9 gene markers could be used as independent prognostic factors. The covariates in dataset GSE20685 include T_stage, M_stage, lymph node metastasis (nodes), in dataset GSE42568 include T_stage, nodes, grade, ER status, in dataset GSE20711 include T_stage, nodes, grade, ER status, HER2 status, and in dataset GSE88770 include nodes, grade, ER status, Ki67. Multivariate COX regression results showed that in four independent datasets, the 9-gene signature can be used as an independent prognostic factor, and its predictive ability has nothing to do with other clinicopathological factors(GSE20685: HR= 2.234(1.3668-3.651), P value= 0.001; GSE42568: HR= 8.388(2.908-24.196), P value= 8.33e-05; GSE20711: HR= 3.857(1.522-9.772), P value= 0.012; GSE88770: HR= 2.860(1.239-6.600), P value= 0.014)(Table 3). In addition, in four independent datasets, lymph node metastasis status can also be used as an independent prognostic factor.

Nomogram development and validation

In four independent datasets, we constructed a nomogram including statistically significant clinicopathological factors and 9-gene prognostic signature, to better quantitatively predict the three-year and five-year survival rate(Figure 5. Top panel And Supplementary figure 1. Left panel). The calibration curve (Figure 5. Lower panel And Supplementary figure 1. Right panel)and C index (GSE20685: Concordance= 0.735 (se=0.028); GSE42568: Concordance= 0.828(se=0.034); GSE20711: Concordance= 0.726(se=0.047); GSE88770: Concordance= 0.76(se =0.055)) indicate that the prediction results of the nomogram shows high accuracy.

Stratification analysis

In the above-mentioned multiple Cox regression analysis, some clinicopathological factors were identified as independent prognostic factors. In order to determine whether the 9-gene signature can be used to predict the overall survival of patients within the same clinical factor subgroup, we combined the four datasets for stratification analysis. We analyzed the patients in the entire cohort according to the lymph node metastasis status (nodes), ER status, T_stage, grade (due to the number of patients in grade I was too samll, only grade II and grade III were analyzed), and divided the patients into high-risk and low-risk groups. The results of the Kaplan-Meier survival curve show that in the same clinical subgroup, the overall survival of patients in the low-risk group is higher than that in the high-risk group(nodes(-): HR=

3.72(2.47-5.60), P value= 1.93e-11; nodes(+): HR= 2.09(1.05-4.13), P value= 3.10e-02; ER(+): HR= 4.80(2.50-9.21), P_value= 2.08e-07; ER(-): HR= 4.73(1.96-11.44), P value= 1.45e-03; T_stage I : HR= 2.32(1.21-4.44), P value= 9.22e-03; T_stage II: HR= 4.82(2.75-8.46), P value= 1.41e-09; Grade II: HR= 4.52(1.94-10.54), P value= 1.32e-04; Grade III: HR= 5.75(2.58-12.83), P value= 1.39e-06)(Figure 6).

Relationship between the 9-gene signature and Disease-Free Survival

We used disease-free survival (DFS) data in GSE42568 and GSE20711 datasets and distant recurrence-free survival (DRFS) data in GSE88770 to determine the role of the 9-gene prognostic signature in predicting disease relapse. The results showed that the disease-free survival (DFS) of BRCA patients in the high-risk group was significantly lower than in the low-risk group (GSE42568: HR = 3.71 (1.91-7.21), P = 3.28e-05; GSE20711: HR = 1.82 (0.95- 3.51), P = 6.88e-02; and GSE88770: HR = 2.59 (1.09-6.17), P = 2.59e-02) (Figure 7. Left panel). The time-varying ROC curve also further verified that the 9-gene prognostic signature has a substantially effective performance in predicting disease-free survival（Figure 7. Right panel）.

9-gene prognostic signature external validation

We used the GSE16446 dataset for external verification. The KM survival curves for OS and DFS showed that the 9-gene prognostic signature have good predictive ability, and the ROC working curve also showed that the gene signature has good working efficiency(Figure 8.).

Gene Set Enrichment Analysis

Finally, we used GSEA enrichment analysis to better determine the biological function of the 9-gene prognostic signature. The top 5 KEGG pathways enriched in high-risk and low-risk sample groups are shown according to the FDR <0.05 cut-off criteria: bladder cancer, glycosaminoglycan biosynthesis chondroitin sulfate, nicotinate and nicotinamide metabolism, steroid biosynthesis, and steroid hormone biosynthesis (Figure 9.).

Discussion

Breast cancer (BRCA) is the highest incidence of cancer in women worldwide [1]. The comprehensive treatment strategy for breast cancer mainly includes surgical resection, chemotherapy, radiation therapy and targeted therapy. The choice of treatment strategy mainly depends on the tumor stage and molecular subtype[15–18]. Screening methods for BRCA mainly include screening mammography, breast

ultrasound, magnetic resonance imaging and clinical breast examination[19, 20]. Effective screening methods can detect BRCA as early as possible and reduce BRCA mortality. However, in recent years, the benefits and harms of breast cancer screening have been hotly debated. According to Løberg M's research, the relative number of over-diagnosis (including ductal carcinoma in situ and invasiveness carcinoma) was 31%. From the age of fifty, fifteen out of one thousand women who undergo mammogram screening are over-diagnosed[21, 22]. At the same time, the prognosis of advanced breast cancer is not optimistic. Even after standardized treatment, many patients eventually develop distant metastases and die from this disease[16]. New breast cancer prognostic markers must be developed to provide guidance and direction for the risk stratification and individualized treatment of BRCA patients.

In this study, we constructed and validated a 9-gene prognostic signature (STXBP3, PKN2, TCAP, STARD3, CDR2L, PNMT, GPR4, ANGPT2, CAPN5) to predict the overall survival of BRCA patients. We used GSE20685, GSE42568, GSE20711, and GSE88770 dataset samples for analysis, and finally selected 9 genes common to these four datasets to build a prognostic gene signature model. The Kaplan-Meier survival curve showed that the overall survival and disease-free survival of patients in the low-risk group were significantly higher than those in the high-risk group. The time-based ROC curve showed that the 9-gene signature shows excellent diagnostic efficiency for OS and DFS events. The nomogram was developed, which combined the 9-gene prognostic signature and other clinicopathological risk factors, and accurately predicted the 3- and 5-year survival probability of BRCA patients. The calibration plots and C index verification showed that the nomogram has good prediction performance. In addition, the results of multivariate COX regression analysis and stratification analysis showed that the 9-gene prognostic signature can exist as an independent risk factor. All of the above results showed that this 9-gene prognostic signature can successfully divide patients into high-risk and low-risk groups, and is also an effective prognostic indicator for BRCA patients.

STXBP3 (syntaxin binding protein 3) has been shown to be involved in fatty acid-induced insulin resistance in skeletal muscle cells[23]. In chronic lymphocytic leukemia (CLL), after knockdown lipoprotein lipase (LPL), LPl can cooperate with STXBP3 to promote CLL cell apoptosis[24]. Protein kinase N2 (PKN2) is a serine / threonine protein kinase related to PKC. It plays an important role in transcription activation, cell cycle, cell adhesion and migration[25]. Koh H's research showed that the C-terminal region of PRK2 can interact with Akt, inhibit threonine phosphorylation at 308 and 473 site of Akt, and specifically down-regulate the activity of Akt protein kinase, blocking the activity of AKT signaling pathway and promoting tumor cell apoptosis[26]. At the same time, PKN2, as a potential tumor suppressor in colon cancer, can inhibit tumor growth by inhibiting the polarization of tumor-associated macrophages (TAMs) to M2-like phenotype[27]. The teneurin C-terminal associated peptides (TCAP) are encoded by four terminal exons of Tenurin, of which TCAP1 can be independently transcribed into a soluble peptide, and can be combined with Latrophilin to mediate cell adhesion[28, 29]. STARD3 can form a chimeric fusion transcript with PPP1R1B. By activating the PI3K / AKT signaling pathway, STARD3 can promote the proliferation and colony formation of gastric cancer cells, thereby promoting the occurrence and development of gastric cancer[30]. Yo antibody can bind endogenous CDR2L, and promote the occurrence of paraneoplastic cerebellar degeneration (PCD)[31]. Phenylethanolamine N-

methyltransferase (PNMT) is a rate-limiting enzyme in adrenaline synthesis and is specifically expressed in adrenergic neurons. In the brains of patients with Alzheimer's disease, PNMT protein at the axon end is reduced due to reduced transport of PNMT [32]. Zhong M's research found that the proton-sensing receptor GPR4 is highly expressed in colorectal cancer (CRC) and promotes the metastasis of CRC cells by inhibiting LATS activity and YAP1 nuclear translocation[33]. Chen Z's research showed that DARPP-32 can further regulate the angiogenic effect of ANGPT2 by inducing STAT3 phosphorylation in gastric tumors[34]. CAPN5 activation can promote the proteolysis and degradation of a variety of substrates, thereby inducing degeneration of the retina and nervous system[35]. CAPN5 can also regulate retinal pigment epithelial cell proliferation by regulating SLIT2 cleavage[36].

Conclusions

This study developed a novel 9-gene signature prediction model to predicting the prognosis of BRCA patients. The results show that the prediction model is a powerful predictor about OS and DFS. In addition, the prediction model does not depend on other clinical case factors such as lymph node metastasis, tumor stage(T,M), and grade. The establishment of this model may help BRCA patients to formulate more accurate treatment plans and improve the prognosis of BRCA.

Abbreviations

BRCA: breast cancer

OS: overall survival

DFS: disease free survival

GEO: Gene Expression Omnibus

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Written informed consent for publication was obtained from all participants.

Availability of data and materials

The datasets used and analyzed in the current study were downloaded from Gene Expression Omnibus databases (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi).

Competing interests

The authors claim no conflict of interest

Funding

Not applicable

Author's contributions

ZT write this article, GW revise this article. JT, XL, QY and YW conducted all statistical analyses.

Acknowledgements

Thanks to the research team of Wuhan University Zhongnan Hospital for their support. And also would like to acknowledge the GEO database.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes–dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol. 2011;22:1736–47.
3. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res. 2005;11:5678–85.
4. Prat A, Pineda E, Adamo B, Galvan P, Fernandez A, Gaba L, et al. Clinical implications of the intrinsic molecular subtypes of breast cancer. Breast. 2015;24(Suppl 2):26–35.

5. von Minckwitz G, Loibl S, Maisch A, Untch M. Lessons from the neoadjuvant setting on how best to choose adjuvant therapies. Breast. 2011;20(Suppl 3):142–5.
6. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med. 2001;344:783–92.
7. Dieras V, Miles D, Verma S, Pegram M, Welslau M, Baselga J, et al. Trastuzumab emtansine versus capecitabine plus lapatinib in patients with previously treated HER2-positive advanced breast cancer (EMILIA): a descriptive analysis of final overall survival results from a randomised, open-label, phase 3 trial. Lancet Oncol. 2017;18:732–42.
8. McDonald ES, Clark AS, Tchou J, Zhang P, Freedman GM. Clinical Diagnosis and Management of Breast Cancer. *J Nucl Med* 2016; 57Suppl 1: 9 s-16 s.
9. Kim I, Choi S, Kim S. BRCA-Pathway: a structural integration and visualization system of TCGA breast cancer data on KEGG pathways. BMC Bioinformatics. 2018;19:42.
10. Kan Z, Ding Y, Kim J, Jung HH, Chung W, Lal S, et al. Multi-omics profiling of younger Asian breast cancers reveals distinctive molecular signatures. Nat Commun. 2018;9:1725.
11. Gyorffy B, Pongor L, Bottai G, Li X, Budczies J, Szabo A, et al. An integrative bioinformatics approach reveals coding and non-coding gene variants associated with gene expression profiles and outcome in breast cancer molecular subtypes. Br J Cancer. 2018;118:1107–14.
12. Kim SY, Kawaguchi T, Yan L, Young J, Qi Q, Takabe K. Clinical Relevance of microRNA Expressions in Breast Cancer Validated Using the Cancer Genome Atlas (TCGA). Ann Surg Oncol. 2017;24:2943–9.
13. Tang XR, Li YQ, Liang SB, Jiang W, Liu F, Ge WX, et al. Development and validation of a gene expression-based signature to predict distant metastasis in locoregionally advanced nasopharyngeal carcinoma: a retrospective, multicentre, cohort study. Lancet Oncol. 2018;19:382–93.
14. Li J, Wang W, Xia P, Wan L, Zhang L, Yu L, et al. Identification of a five-lncRNA signature for predicting the risk of tumor recurrence in patients with breast cancer. Int J Cancer. 2018;143:2150–60.
15. Greenberg S, Stopeck A, Rugo HS. Systemic treatment of early breast cancer–a biological perspective. J Surg Oncol. 2011;103:619–26.
16. Giordano SH. Update on locally advanced breast cancer. Oncologist. 2003;8:521–30.
17. Smith I, Chua S. Medical treatment of early breast cancer. III: chemotherapy. Bmj. 2006;332:161–2.
18. Danoff BF, Haller DG, Glick JH, Goodman RL. Conservative surgery and irradiation in the treatment of early breast cancer. Ann Intern Med. 1985;102:634–42.
19. Nattinger AB, Mitchell JL. Breast Cancer Screening and Prevention. Ann Intern Med. 2016;164:ltc81–96.
20. Peairs KS, Choi Y, Stewart RW, Sateia HF. Screening for breast cancer. Semin Oncol. 2017;44:60–72.
21. Loberg M, Lousdal ML, Bretthauer M, Kalager M. Benefits and harms of mammography screening. Breast Cancer Res. 2015;17:63.

22. The benefits. and harms of breast cancer screening: an independent review. Lancet. 2012;380:1778–86.
23. Schlaepfer IR, Pulawa LK, Ferreira LD, James DE, Capell WH, Eckel RH. Increased expression of the SNARE accessory protein Munc18c in lipid-mediated insulin resistance. J Lipid Res. 2003;44:1174–81.
24. Porpaczy E, Tauber S, Bilban M, Kostner G, Gruber M, Eder S, et al. Lipoprotein lipase in chronic lymphocytic leukaemia - strong biomarker with lack of functional significance. Leuk Res. 2013;37:631–6.
25. Vincent S, Settleman J. The PRK2 kinase is a potential effector target of both Rho and Rac GTPases and regulates actin cytoskeletal organization. Mol Cell Biol. 1997;17:2247–56.
26. Koh H, Lee KH, Kim D, Kim S, Kim JW, Chung J. Inhibition of Akt and its anti-apoptotic activities by tumor necrosis factor-induced protein kinase C-related kinase 2 (PRK2) cleavage. J Biol Chem. 2000;275:34451–8.
27. Cheng Y, Zhu Y, Xu J, Yang M, Chen P, Xu W, et al. PKN2 in colon cancer cells inhibits M2 phenotype polarization of tumor-associated macrophages via regulating DUSP6-Erk1/2 pathway. Mol Cancer. 2018;17:13.
28. Woelfle R, D'Aquila AL, Pavlovic T, Husic M, Lovejoy DA. Ancient interaction between the teneurin C-terminal associated peptides (TCAP) and latrophilin ligand-receptor coupling: a role in behavior. Front Neurosci. 2015;9:146.
29. Husic M, Barsyte-Lovejoy D, Lovejoy DA. Teneurin C-Terminal Associated Peptide (TCAP)-1 and Latrophilin Interaction in HEK293 Cells: Evidence for Modulation of Intercellular Adhesion. Front Endocrinol (Lausanne). 2019;10:22.
30. Yun SM, Yoon K, Lee S, Kim E, Kong SH, Choe J, et al. PPP1R1B-STARD3 chimeric fusion transcript in human gastric cancer promotes tumorigenesis through activation of PI3K/AKT signaling. Oncogene. 2014;33:5341–7.
31. Krakenes T, Herdlevaer I, Raspotnig M, Haugen M, Schubert M, Vedeler CA. CDR2L Is the Major Yo Antibody Target in Paraneoplastic Cerebellar Degeneration. Ann Neurol. 2019;86:316–21.
32. Burke WJ, Chung HD, Marshall GL, Gillespie KN, Joh TH. Evidence for decreased transport of PNMT protein in advanced Alzheimer's disease. J Am Geriatr Soc. 1990;38:1275–82.
33. Yu M, Cui R, Huang Y, Luo Y, Qin S, Zhong M. Increased proton-sensing receptor GPR4 signalling promotes colorectal cancer progression by activating the hippo pathway. EBioMedicine. 2019;48:264–76.
34. Chen Z, Zhu S, Hong J, Soutto M, Peng D, Belkhiri A, et al. Gastric tumour-derived ANGPT2 regulation by DARPP-32 promotes angiogenesis. Gut. 2016;65:925–34.
35. Wang Y, Zhang X, Song Z, Gu F. An anti-CAPN5 intracellular antibody acts as an inhibitor of CAPN5-mediated neuronal degeneration. Oncotarget. 2017;8:100312–25.
36. Wang Y, Li H, Zang S, Li F, Chen Y, Zhang X, et al. Photoreceptor Cell-Derived CAPN5 Regulates Retinal Pigment Epithelium Cell Proliferation Through Direct Regulation of SLIT2 Cleavage. Invest

Tables

General information on constructing prognostic signature for 9 genes.

Gene name	Gene type	Gene type	Location
TXBP3	Protect gene	Protein Coding	Chromosome 3, NC_000069.6
PKN2	Protect gene	Protein Coding	Chromosome 1, NC_000001.11
TCAP	Danger gene	Protein Coding	Chromosome 17, NC_000017.11
STARD3	Danger gene	Protein Coding	Chromosome 17, NC_000017.11
CDR2L	Danger gene	Protein Coding	Chromosome 17, NC_000017.11
PNMT	Danger gene	Protein Coding	Chromosome 10, NC_005109.4
GPR4	Danger gene	Protein Coding	Chromosome 19, NC_000019.10
ANGPT2	Danger gene	Protein Coding	Chromosome 8, NC_000008.11
CAPN5	Danger gene	Protein Coding	Chromosome11, NC_000011.10

Univariate regression analysis was performed on the overall survival of 9 genes and BRCA patients in four datasets.

Genes	GSE26085		GSE42568		GSE20711		GSE88770	
	HR(95%CI)	P-value	HR(95%CI)	P-value	HR(95%CI)	P-value	HR(95%CI)	P-value
STXBP3	0.27 (0.12-0.60)	1.46E-03	0.17 (0.08-0.37)	4.46E-06	0.12(0.03-0.45	1.91E-03	0.12 (0.03-0.51)	4.01E-03
PKN2	0.58 (0.34-0.99)	4.86E-02	0.31 (0.14-0.69)	3.83E-03	0.24 (0.07-0.88)	3.08E-02	0.16 (0.03-0.95)	4.38E-02
TCAP	1.69 (1.25-2.28)	5.91E-04	1.86 (1.16-2.97)	9.81E-03	2.38 (1.35-4.19)	2.74E-03	6.87 (1.61-29.21)	9.08E-03
STARD3	1.33 (1.12-1.58)	1.36E-03	1.64 (1.20-2.25)	1.96E-03	1.35 (1.04-1.74)	2.31E-02	1.62 (1.01-2.60)	4.69E-02
CDR2L	1.66 (1.21-2.29)	1.91E-03	2.09 (1.32-3.32)	1.74E-03	2.02 (1.13-3.63)	1.79E-02	5.89 (2.21-15.70)	3.85E-04
PNMT	1.20 (1.04-1.38)	1.26E-02	1.35 (1.09-1.67)	5.41E-03	1.44 (1.21-1.72)	4.71E-05	2.01 (1.23-3.29)	5.26E-03
GPR4	2.46 (1.15-5.26)	2.02E-02	7.63 (2.85-20.39)	5.14E-05	13.26 (2.21-79.45)	4.65E-03	6.03 (1.80-20.26)	3.65E-03
ANGPT2	1.56 (1.07-2.26)	2.08E-02	4.05 (1.60-0.25)	3.09E-03	2.41 (1.21-4.81)	1.28E-02	2.70 (1.13-6.46)	2.54E-02
CAPN5	1.69 (1.08-2.63)	2.11E-02	6.00 (2.20-16.41)	4.78E-04	4.50 (2.12-9.55)	9.16E-05	2.49 (1.09-5.69)	3.13E-02

Univariate and multivariate Cox regression analyses were performed on the gene signatures and overall survival of BRCA patients in four datasets.

Variables		Patients(N)	Univariate analysis		Multivariate analysis	
			HR(95% CI)	*P*	HR(95% CI)	*P*
GSE20685						
T_stage	I/II	101/188	1.136(0.664-1.944)	0.642	0.732(0.4150-1.292)	0.281
T_stage	I/III	101/38	4.663(2.550-8.526)	**5.7e-07**	1.824(0.889-3.744)	0.101
M_stage	M0/M1	319/8	5.204(2.391-11.33)	**3.22e-05**	1.475(0.6029-3.609)	0.394
Nodes	-/+	137/190	3.785(2.163-6.623)	**3.12E-06**	3.391(1.869-6.155)	**5.91e-05**
Risk score	Low/High	163/164	2.678(1.672-4.287)	**4.11e-05**	2.234(1.3668-3.651)	**0.001**
GSE42568						
T_stage	I/II	34/67	2.311(1.002-5.332)	**0.049**	1.438(0.613-3.374)	0.404
Nodes	-/+	44/57	4.556(1.877-11.050)	**0.001**	3.601(1.460-8.882)	**0.005**
Grade	I/II	10/40	1.857(0.232-14.86)	0.559		
Grade	I/III	10/51	6.208(0.839-45.96)	0.073		
ER	-/+	34/67	0.532(0.2679-0.086)	**0.041**	0.461(0.230-0.922)	**0.028**
Risk score	Low/High	50/51	9.746(3.418-27.790)	**2.06e-05**	8.388(2.908-24.196)	**8.33e-05**
GSE20711						
T_stage	I/II	46/39	1.833(0.838-4.044)	0.133		
Nodes	-/+	29/56	3.115(1.067-9.095)	**0.037**	2.483(1.082-7.322)	**0.046**
Grade	I/II	13/4	1.277 (0.414-14.300)	0.843		
Grade	I/III	13/68	2.433(0.571-10.360)	0.229		
ER	-/+	43/42	0.557(0.245-1.265)	0.162		
HER2	-/+	61/24	2.533(1.146-5.597)	**0.021**	1.412(0.606-3.290)	0.424
Risk score	Low/High	42/43	4.378(1.739-11.020)	**0.001**	3.857(1.522-9.772)	**0.012**
GSE88770						
Nodes	-/+	62/47	2.455(1.100-5.476)	**0.028**	2.338(1.035-5.280)	**0.041**
Grade	I/II	13/90	2.190(0.513-9.338)	0.290		
ER	-/+	11/98	0.997(0.338-2.937)	0.996		
Ki67(%)	~15/15~30	71/23	3.565(1.525-8.333)	**0.003**	3.160(1.344-7.429)	**0.008**
Ki67(%)	~15/30~	71/15	1.827(0.642-5.200)	0.259	1.515(0.527-4.358)	0.441
Risk score	Low/High	54/55	3.263(1.418-7.509)	**0.005**	2.860(1.239-6.600)	**0.014**

Figures

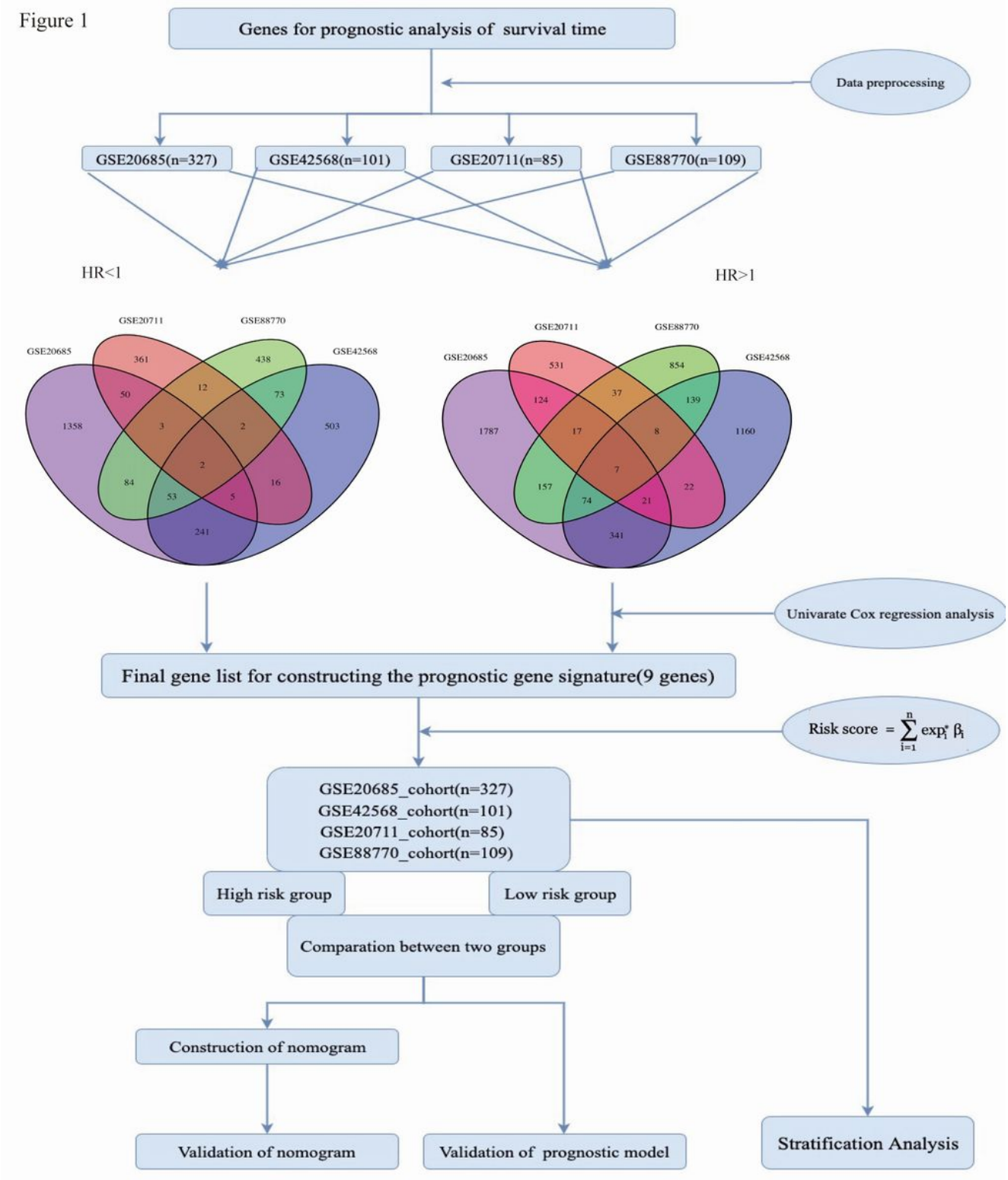

Figure 1

Flow diagram of date preparation, processing, analysis, and validation.

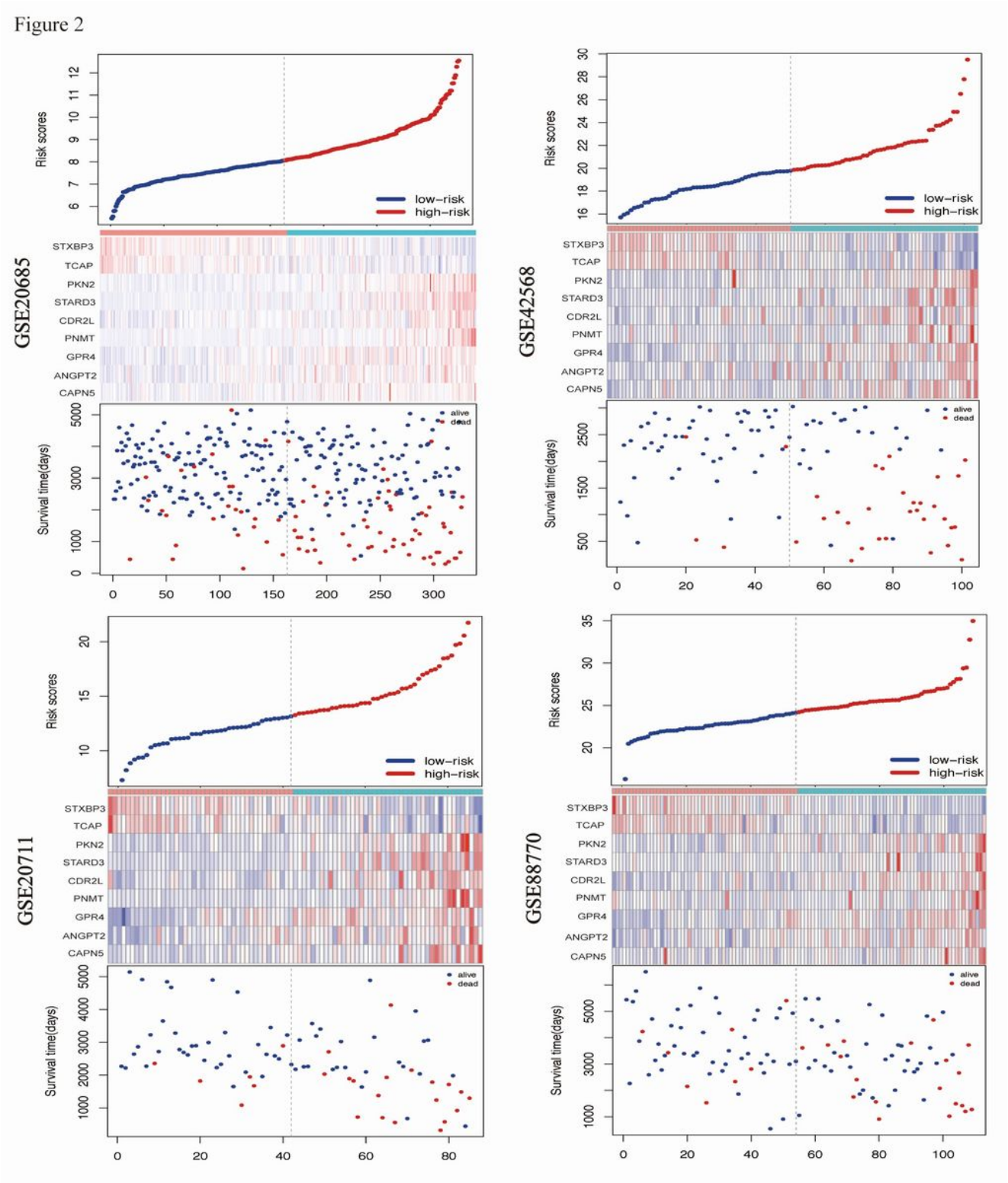

Figure 2

Analysis of risk score for BRCA patients in four datasets. From top to bottom are risk score distribution, gene expression profile and patient survival status. The black dashed line represents the median value of the risk score, which is used as a boundary to divide patients into high-risk and low-risk groups.

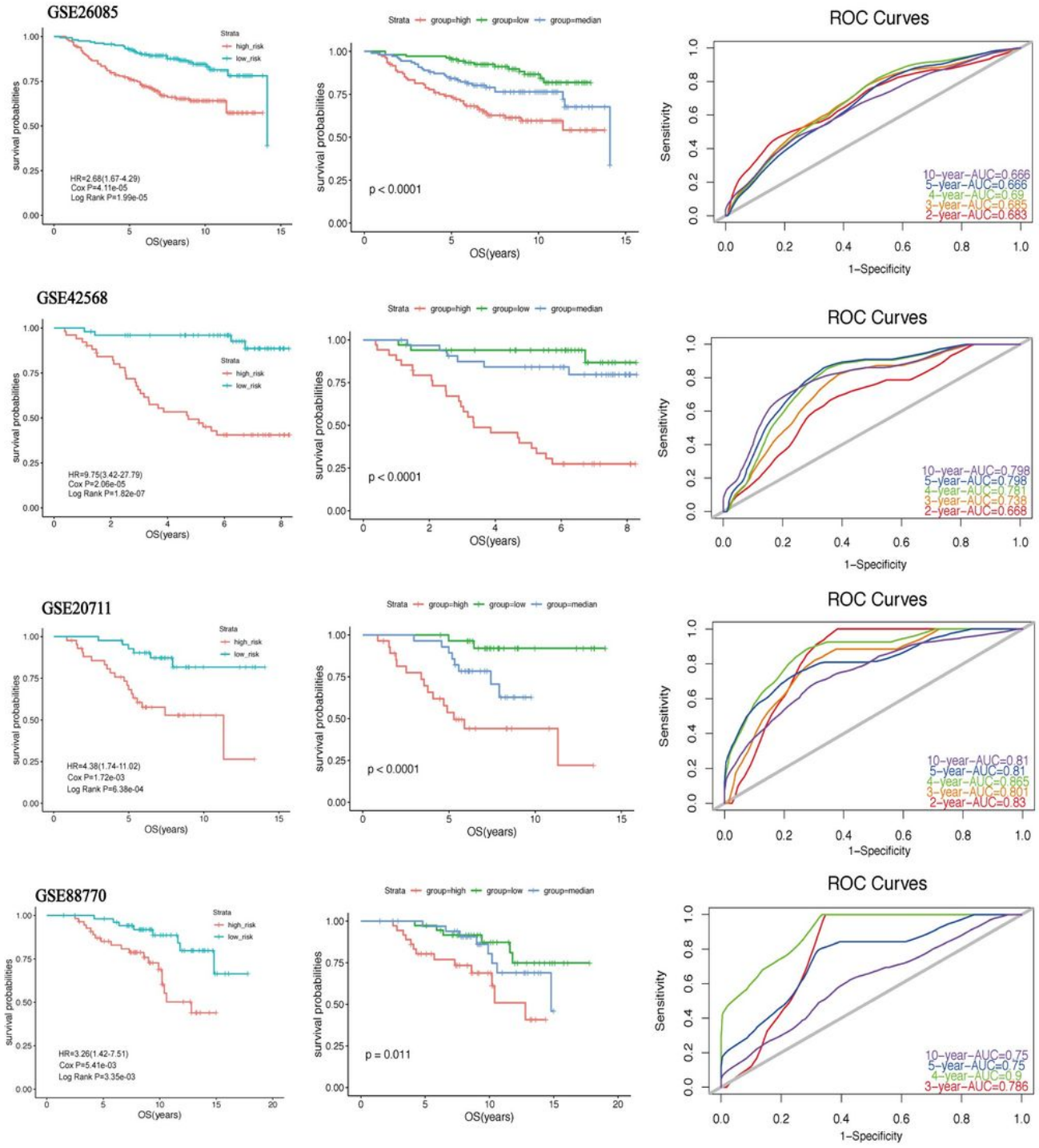

Figure 3

Kaplan-Meier survival curve and ROC curve for 9-gene signature in four datasets. The overall survival of patients in the high-risk group is lower than that in the low-risk group. P value <0.05 was considered statistically significant.

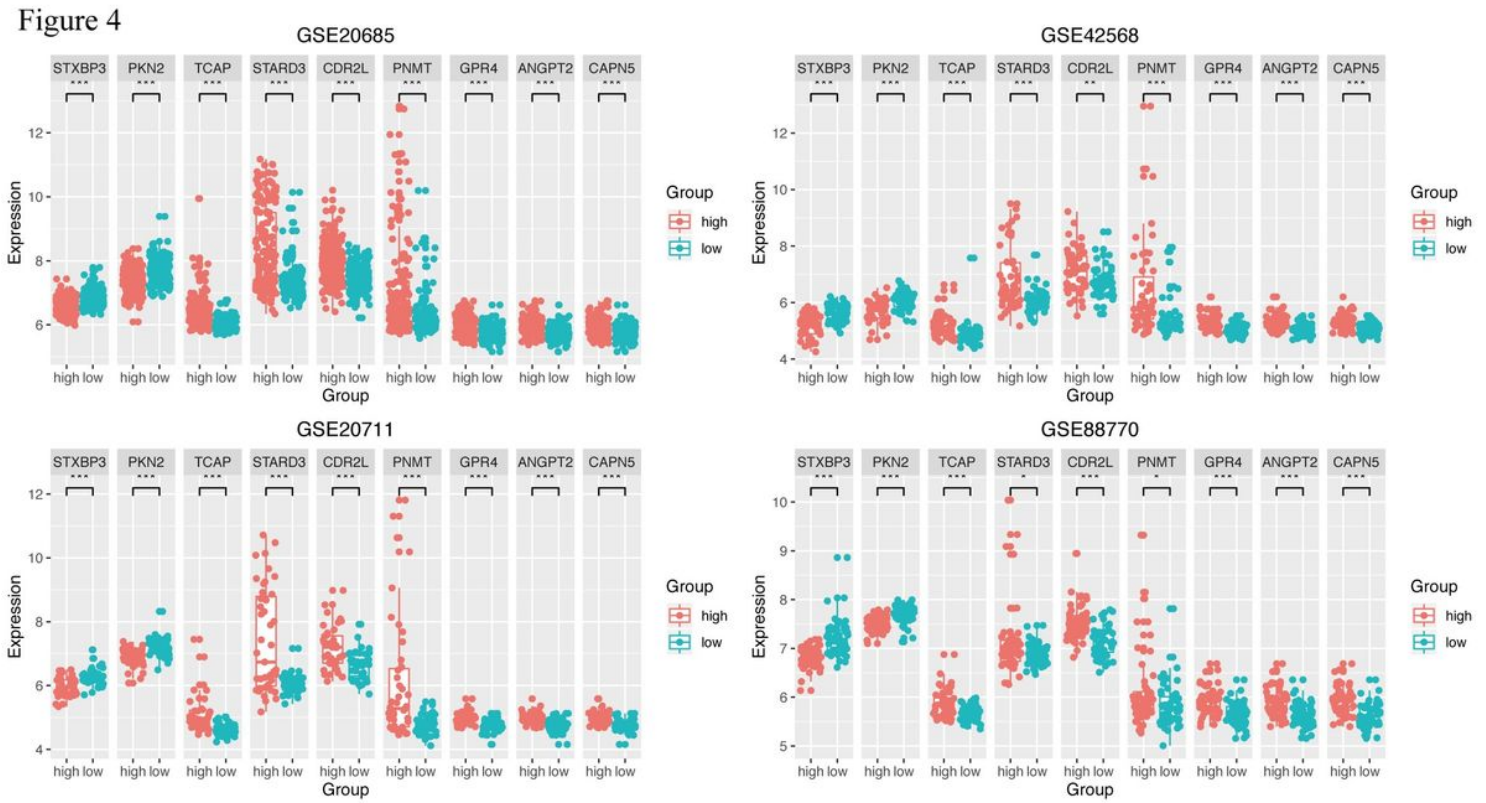

Figure 4

Box plot visualization of gene expression levels of 9-gene signature in four datasets. Patients in the high-risk group had higher expression of danger genes, while those in the low-risk group had lower expression of protect genes. P value <0.05 was considered statistically significant. * P <0.05; ** P <0.01; *** P <0.001.

Figure 5

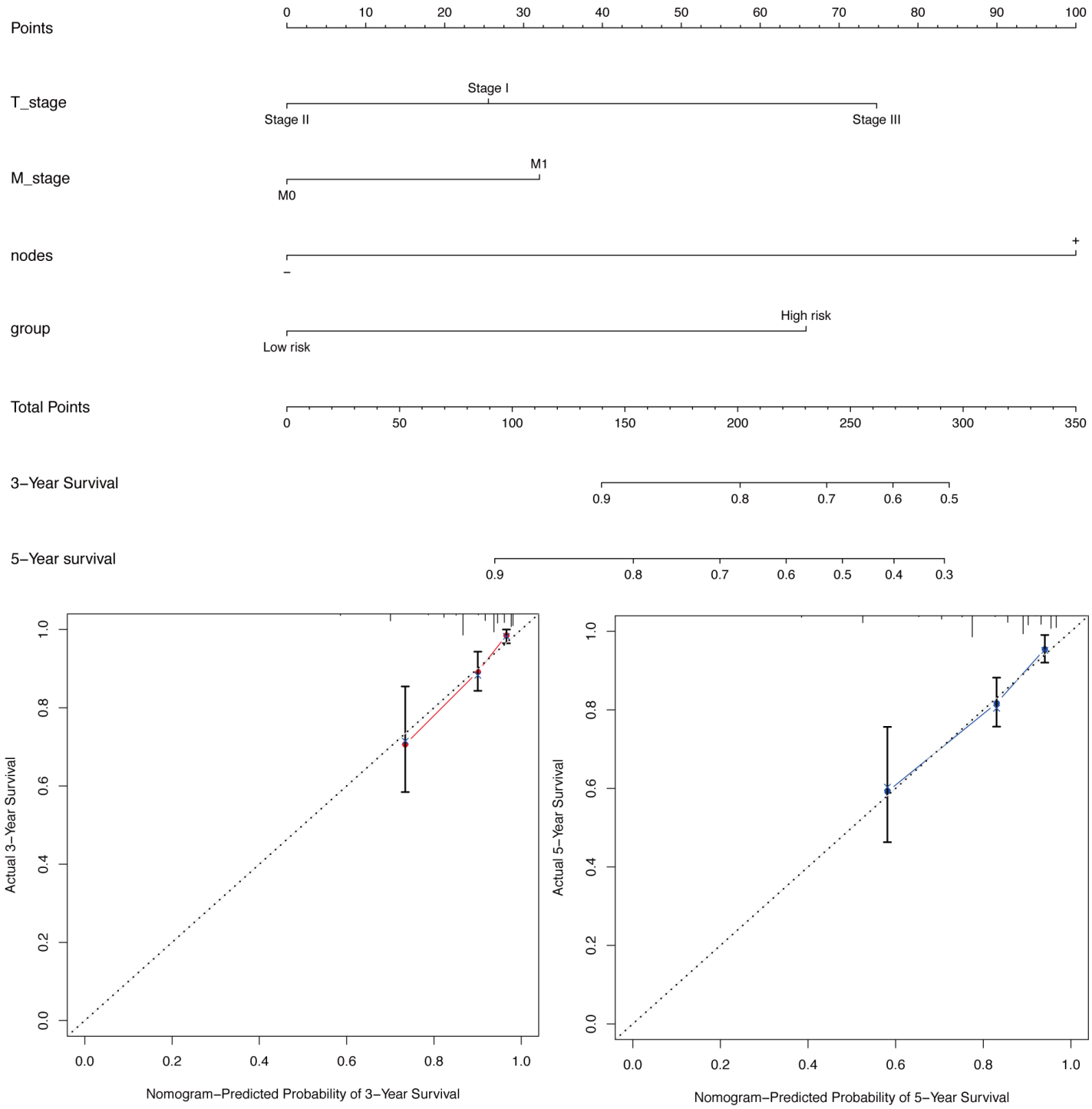

Figure 5

Nomogram of the GSE20685 datasets was used to predict the overall survival of BRCA patients. The 3-year and 5-year overall survival calibration plots show that the nomogram has good prediction accuracy.

Figure 6

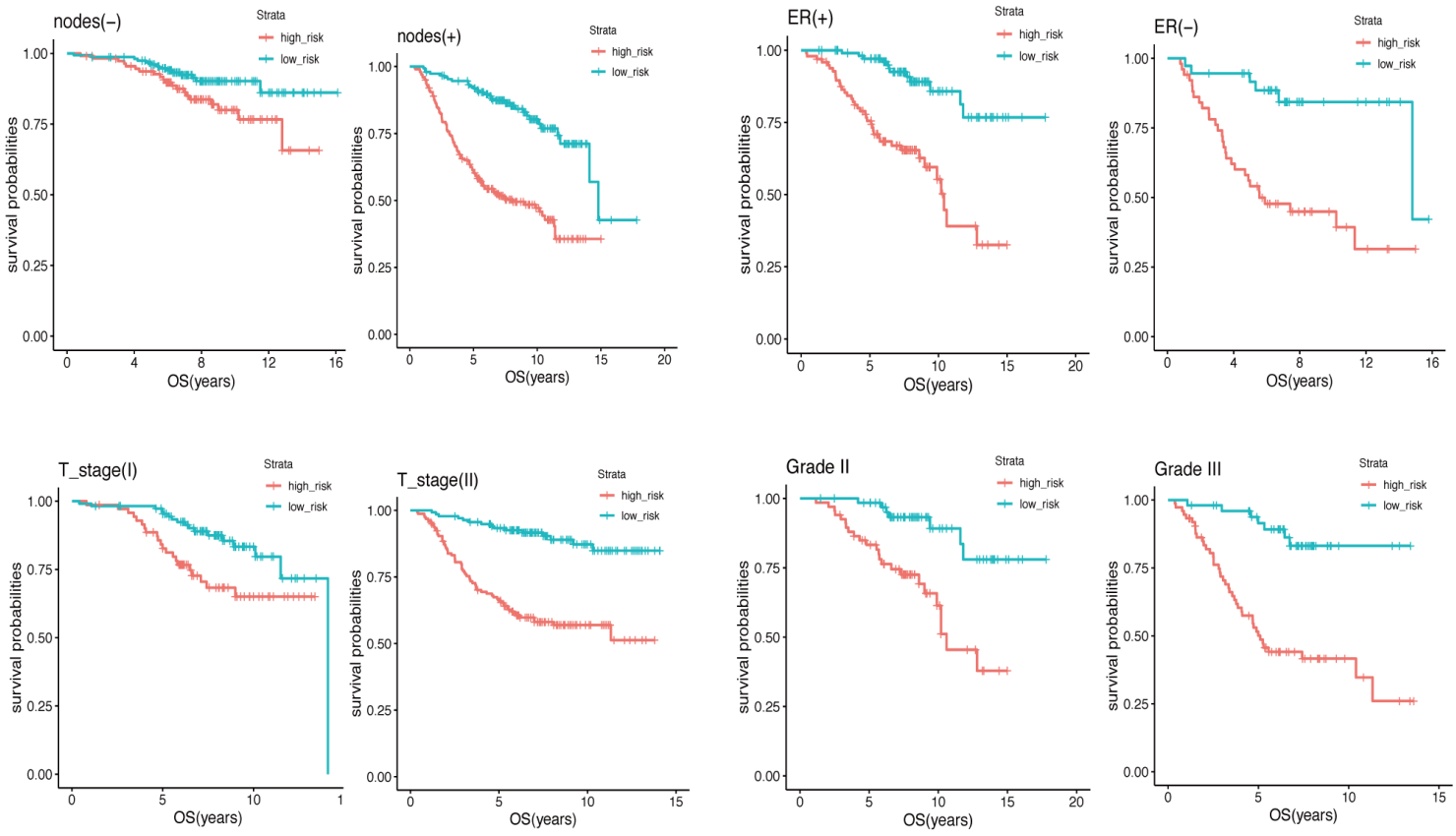

Figure 6

The Kaplan-Meier survival curve was drawn to predict the overall survival of patients by stratification analysis about nodes, ER stauts, T_stage and grade. The overall survival of patients in the high-risk group was lower than that in the low-risk group. P value <0.05 was considered statistically significant.

Figure 7

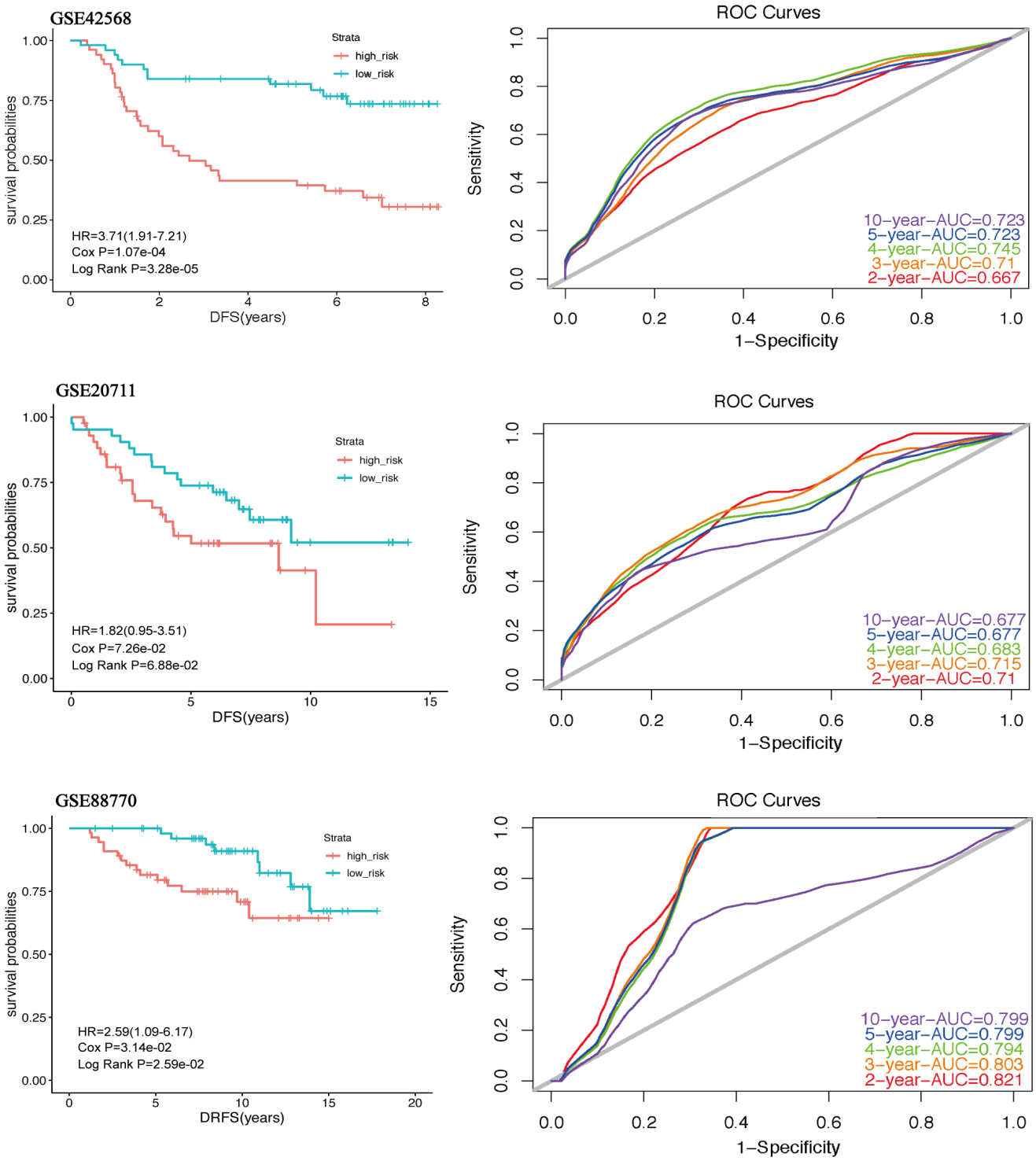

Figure 7

Kaplan-Meier survival curve and ROC curve of 9-gene signature in the three datasets of GSE42568, GSE20711 and GSE88770. Patients in the high-risk group had shorter disease-free survival. P value <0.05 was considered statistically significant.

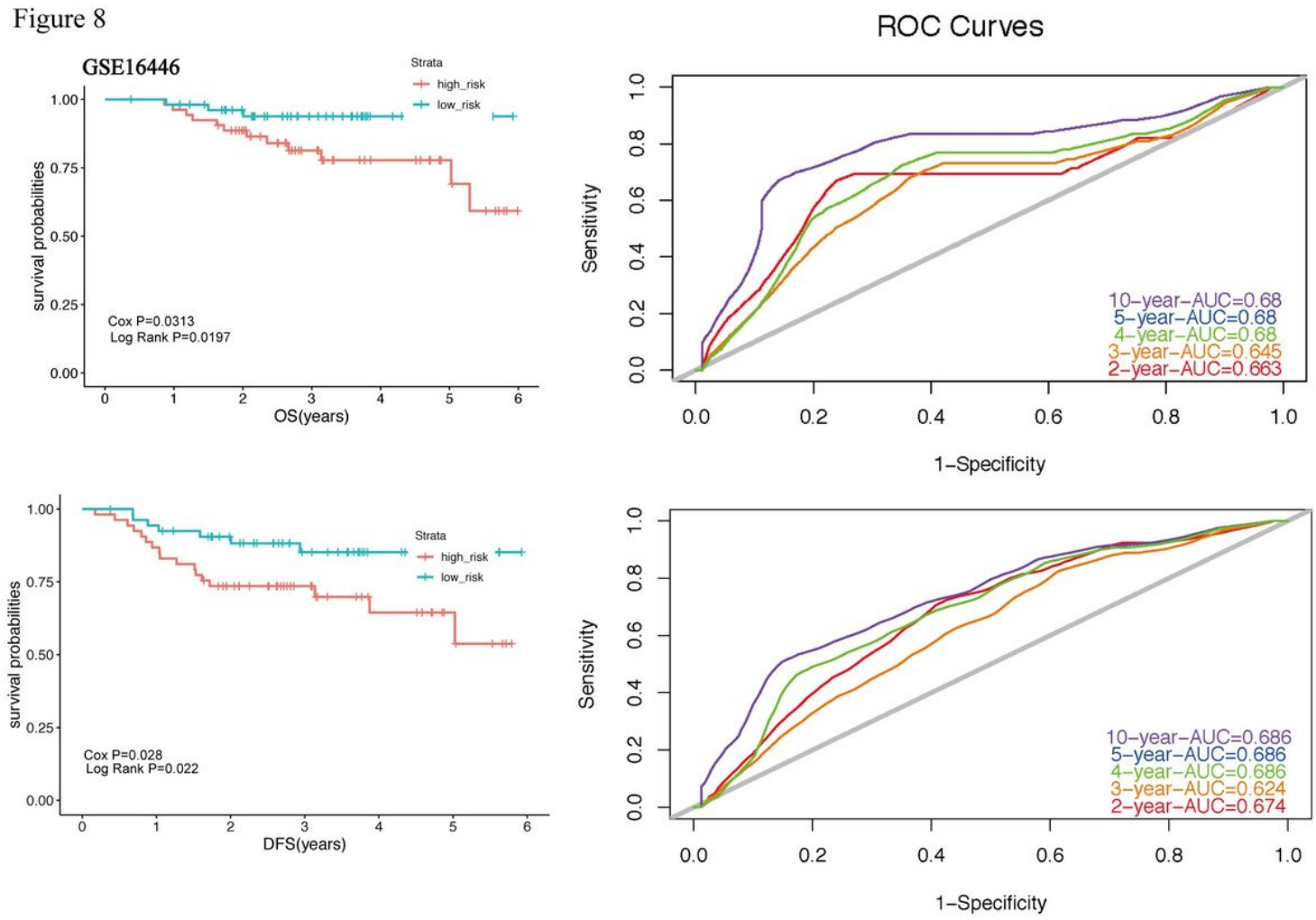

Figure 8

Exteral validation of 9-gene signature. Patients in the high-risk group had shorter OS and DFS. P value <0.05 was considered statistically significant.

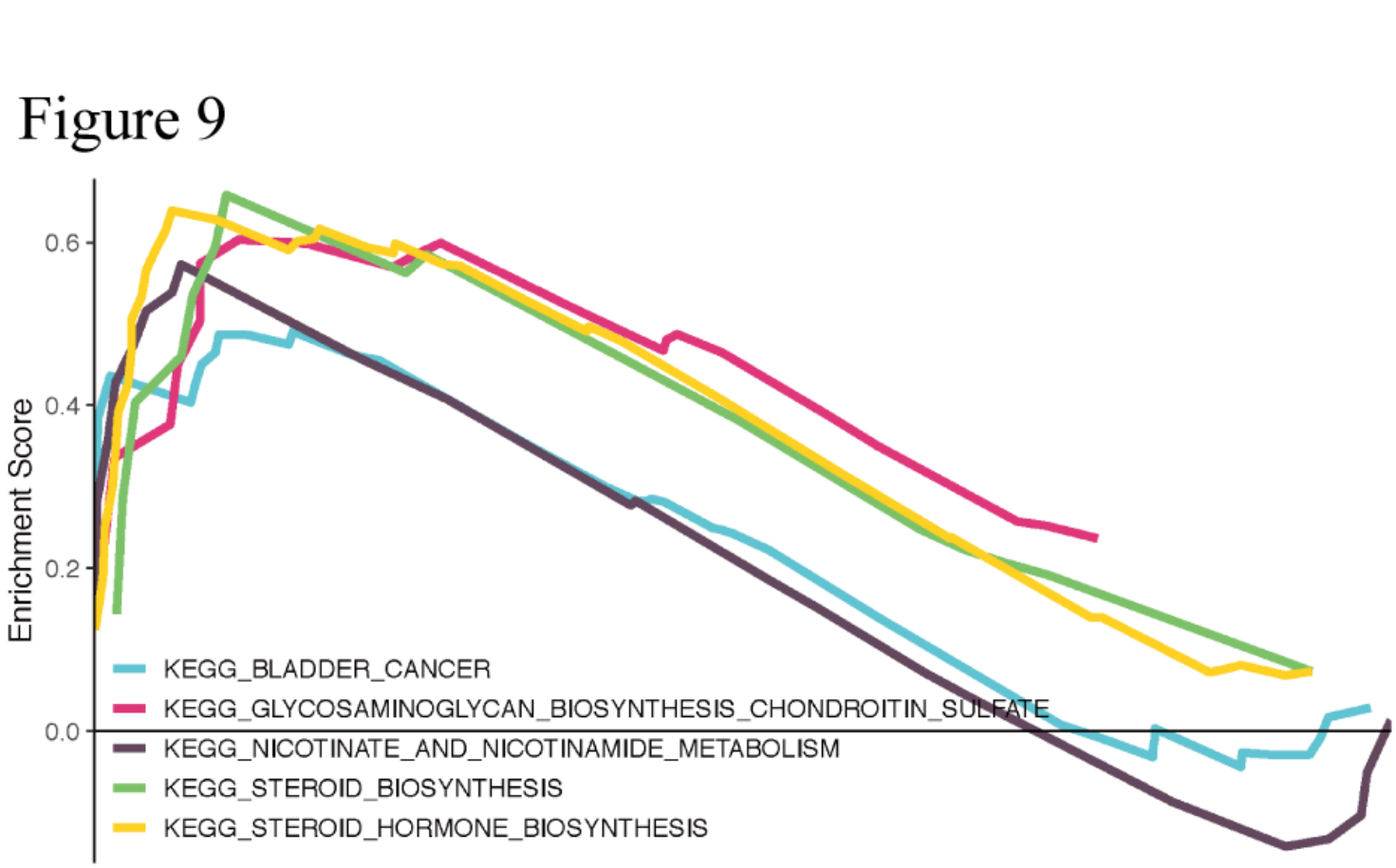

Figure 9

Gene set enrichment analysis.

Supplementary Files

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

- supplementaryFigure1.tif