Identification and Validation of a Novel 16-Gene Prognostic Signature for Patients with Breast Cancer

Zhenhua Zhong
Ningbo women and children's hospital
Wenqiang Jiang
Ningbo women and children's hospital
Jing Zhang
Ningbo women and children's hospital
Zhanwen Li
Ningbo women and children's hospital
Fengfeng Fan (Purplesu123@163.com)
Ningbo Women and Children's Hospital

Research

Keywords: BRCA, the 16-gene signature, nomogram, overall survival

Posted Date: September 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-837034/v1

Abstract

Background: Despite increased early diagnosis and improved treatment in breast cancer (BRCA) patients, prognosis prediction is still a challenging task due to the disease heterogeneity. This study was to identify a novel gene signature that can accurately evaluate BRCA patient survival.

Methods: The gene expression and clinical data of BRCA patients were collected from The Cancer Genome Atlas (TCGA) and the Molecular Taxonomy of BRCA International Consortium (METABRIC) databases. Genes associated with prognosis were determined by Kaplan–Meier survival analysis and multivariate Cox regression analysis. A prognostic 16-gene score was established with linear combination of 16 genes. The prognostic value of the signature was validated in the METABRIC dataset. Gene expression analysis was performed to investigate the diagnostic values of 16 genes.

Results: The 16-gene score was associated with shortened overall survival in BRCA patients independently of clinicopathological characteristics. The signalling pathways of cell cycle, oocyte meiosis, RNA degradation, progesterone mediated oocyte maturation and DNA replication were the top five most enriched pathways in the high 16-gene score group. The 16-gene nomogram incorporating the survival-related clinical factors showed improved prediction accuracies for 1-year, 3-year and 5-year survival (area under curve [AUC] = 0.91, 0.79 and 0.77 respectively). *MORN3, IGJ, DERL1* exhibited high accuracy in differentiating BRCA tissues from normal breast tissues (AUC > 0.80 for all cases).

Conclusions: The 16-gene profile provides novel insights into the identification of BRCA with a high risk of death, which eventually guides treatment decision making.

Background

BRCA (BRCA) is the most prevalent female malignancy in US and China. An estimated 284,200 cases will be diagnosed and 44,130 patients will die of the disease in 2021, accounting for more than 15% of newly diagnosed cancer cases and 7.3% of cancer-related mortalities (1, 2). According to the molecular classifications, BRCA can be mainly divided into five subtypes: luminal A, luminal B/ human epidermal growth factor receptor 2[Her2] negative, triple positive (ER+, Progesterone receptor [PR]+, Her2+), Her2-enriched, and triple negative (ER-, PR-, Her2-) (3). With the significant progresses of medical technology, the prognosis of BRCA has been remarkably ameliorated. However, the prognosis is still not optimistic for BRCA patients diagnosed at late stages.

The development of methods for risk stratification in BRCA has been a hotspot of research. Several studies demonstrate that multigene signatures might be more accurate for risk stratification than the traditional approaches in BRCA(4, 5). The MammaPrint, a 70-gene signature, is a prognostic model to stratify node-negative BRCA patients with different survival probabilities (4). Oncotype DX is a 21-gene signature that provides information of the likelihood of recurrence and weighs the potential benefits of chemotherapy in the node-negative, estrogen receptor positive BRCA (5). These multigene assays show potential clinical utility, but still need to be validated in large, randomized trials (6). Moreover, the

established methods are applicable to only limited disease subtypes, there is still lack of an effective prognostic model that could be used for almost all BRCA subtypes.

In the current study, we aimed to develop a novel gene profile to accurately estimate disease prognosis. We first examined all genes for their association with overall survival (OS) using the expression and clinical data of The Cancer Genome Atlas (TCGA) database (7)and validated the results in the Molecular Taxonomy of BRCA International Consortium (METABRIC) database(8). We next established a 16-gene score based on a linear combination of 16 gene expression levels and 16-gene nomogram to precisely predict the overall survival (OS) of BRCA patients. Lastly, we performed expression analysis of 16 genes and demonstrated their diagnostic values in BRCA.

Methods And Materials

Data acquisition and processing

We obtained RNA-seq expression data and clinical data of BRCA patients from the two different sources, the first of which was the TCGA database (n=1080 patients), the second source was the METABRIC study which was used to validate the associations between gene expression and OS (n=1,904 patients). Clinical features of BRCA patients are summarized and presented in Table1 and supplementary Table1 respectively. As the gene expression unit of the TCGA dataset differs from that of the METABRIC cohort, normalization of gene expression was performed using the formula $z = (x- \bar{x})/s$. where x, $\bar{x}$ and s are the gene expression value, mean and standard deviation of gene expression values.

Identification of survival-related clinical features and genes

We aimed to identify survival-related clinical features using different statistical methods. For quantitative variables, we utilized student t test to characterize their associations with OS. With respects to qualitative variables, we implemented fisher exact test to investigate their associations with OS. We followed Sha et al's methods (9) to identify and classify survival-related genes. In brief, we firstly split BRCA samples into two subgroups, the low-expression and high-expression groups, based on the median expression value. We performed Kaplan–Meier survival analysis to evaluate the statistical significance of the differences in OS with the survival package (10,11) and conducted multivariate Cox regression model to further validate the survival analysis. Survival-related genes with odd ratio [OR] > 1were considered as risk genes, while genes with 0<OR<=1 were defined as protective genes. To further evaluate the prognostic importance of the 16-gene score, we drew receiver operating characteristic (ROC) curves and computed the area under curve (AUC) values using the R package pROC (12). To investigate the potential biological function of prognosis-related genes, we analyzed the enrichment of Gene Ontology (GO) term and Kyoto Encyclopedia of Genes and Genomes (KEGG) signalling pathway using the online tool g:profiler(13).

Establishment and validation of the 16-gene score

We followed Lai et al's methods (14) to choose the set of genes which performed best in prognosis prediction and develop the 16-gene risk score. In brief, the least absolute shrinkage and selection operator (LASSO) models comprising different number of genes were evaluated for prediction accuracies of OS using glmnet in the TCGA dataset (15). The 16-gene score was created using the following formula: 16-gene score = -1.91 + expression of gene 1 × β1 + expression of gene 2 × β2 +⋯+ expression of gene n × βn. β values represented the coefficients generated from the optimal LASSO model. We then implemented Kaplan–Meier survival analysis, multivariate Cox regression analysis and stratification analysis to further investigate the association between the 16-gene score and OS in BRCA. We also analyzed the prediction capability of the 16-gene score for progression-free survival (PFS) and disease-free survival (DFS) in the TCGA cohort using Kaplan–Meier survival analysis. Lastly, we utilized linear regression model to investigate the correlations between clinical characteristics and the 16-gene score in the TCGA and METABRIC cohorts. $P < 0.05$ was considered statistically significant.

Gene set enrichment analysis

On the basis of the median 16-gene score, the BRCA patients were split into two subgroups: the high and low 16-gene score groups. Gene set enrichment analysis (GSEA) (16) was implemented to determine the dysregulated signalling pathways related to the 16-gene score using the default parameters. Q value < 0.25 was considered statistically significant.

Construction and validation of the 16-gene nomogram

Nomogram was constructed using the rms package in R, and included patient's age, tumor stage, menopause status, number of positive lymph nodes and 16-gene signature as they are significantly correlated with OS of BRCA. The performance of the nomogram developed was evaluated in the TCGA cohort and validated in the METABRIC cohort using the R package pROC. AUC values were computed accordingly for the nomogram in the prediction of one-year, three-year and five-year survival.

Expression analysis of prognosis-related genes

The online server cbioportal (17) was utilized to analyze the mutational profiles of the 16 genes in the TCGA cohort. Furthermore, the expression data of 779 BRCA tissues and 100 paired non-cancerous tissues were downloaded from the TCGA database. Differentially expressed genes were determined between BRCA tissues and paired normal tissues using student t test. To investigate the diagnostic values of the 16 genes, the pROC package was used to determine whether the gene expression could effectively distinguish cancer tissues from paired normal ones. P value was adjusted using false discovery rate. Adjusted $P < 0.05$ indicated statistical significance.

Results

Identification of survival-related clinical features in BRCA

We initially performed survival analysis between clinical features and OS and revealed higher patient's age, more positive lymph nodes, higher cancer stage, clinical T stage, clinical N stage, clinical M stage, post-menopause were high risk prognosticators for OS in the TCGA cohort ($P < 0.05$ for all cases, Table1). Furthermore, the inverse correlations between overall survival and patient's age, more positive lymph nodes, higher cancer stage, post-menopause, tumor size, radiotherapy were independently validated in the METABRIC cohort ($P < 0.05$ for all cases, supplementary Table1).

Identification and validation of survival-related genes in BRCA

We first examined the relation between gene expression and OS in the TCGA data set. The results showed that high expression levels of 1374 genes were related to significantly prolonged OS. While, high expression levels of 678 genes were related to significantly reduced survival in the TCGA cohort ($P <0.05$ for all cases, log rank test, Figure1). Multivariate Cox regression analysis confirmed 432 protective prognostic genes and 219 risk prognostic genes following the adjustment of clinical characteristics. Furthermore, the association between 651 gene expression and OS was analyzed in the METABRIC dataset (n=1904). The results validated 80 protective genes and 34 risk genes in the METABRIC cohort respectively ($P <0.05$ for all cases, log rank test, Figure1). Then, we analyzed the functional involvement of the protective and risk genes with g.profiler and uncovered the 80 protective genes were significantly enriched in the KEGG pathway of focal adhesion. While, the risk genes were significantly over-represented in GO terms, such as nuclear division, organelle fission, DNA metabolic process and nucleic acid metabolic process (adjusted P value < 0.05 for all cases, supplementary Figure1).

Construction of a 16-gene signature and its prognostic value in BRCA

The LASSO model comprising 16 genes showed the highest AUC value and was deemed the best model for survival prediction (Figure2A). Then we established the 16-gene score formula and computed the risk score for each BRCA patient. The Kaplan-Meier survival analysis and multivariate Cox regression analysis indicated that the high 16-gene score was indicative of worse OS in BRCA ($P<0.05$ for all cases, OR: 3.47, 95% confidence interval: 2.08-5.78, supplementary Figure2A). We also analyzed the association between the 16-gene score and DFS and PFS in the TCGA cohort. Similarly, we demonstrated that the high 16-gene score was significantly associated with shorter DFS and PFS (P value < 0.05 for all cases, supplementary Figure3). For further verification, the 16-gene score was calculated in the METABRIC dataset. The results also confirmed the negative correlation between the 16-gene score and patient's OS (Figure2D, supplementary Figure2B). Furthermore, the 16-gene score (AUC = 0.72, 0.71, 0.73, respectively) outperformed cancer stage (AUC = 0.71, 0.69, 0.66, respectively, supplementary Figure4) in predicting 1-year survival, 3-year survival and 5-year survival in the TCGA cohort. The results were also validated in the METABRIC cohort (supplementary Figure4) and suggested the 16-gene score is superior to cancer stage in the prediction of prognosis of BRCA patients.

Correlations between the 16-gene score and clinical factors in BRCA

The linear regression model analysis showed the 16-gene score was significantly positively associated with age, HER2 status, menopause status, clinical stage, clinical T stage, clinical M stage and negatively correlated with PR status, ER status, hormone therapy and radiotherapy in the TCGA cohort ($p<0.05$ for all cases, Figure3A). Moreover, the 16-gene score also exhibited positive correlation with age, HER2 status, menopause status, clinical stage and negative correlation with PR status, ER status, hormone therapy and radiotherapy in the METABRIC cohort ($p< 0.05$ for all cases, Figure3B). Next, we split BRCA patients into subgroups according to the clinical characteristics and conducted the Kaplan-Meier survival analysis to assess the prognostic value of the 16-gene score in clinical factor-specific subgroups. Overall, the results demonstrated that the high-risk was significantly correlated with worse OS in the same clinical subgroup of the TCGA cohort ($P<0.05$ for all cases, log rank test, supplementary Table2). Similar findings were also observed in the METABRIC cohort (supplementary Table3), suggesting that the implication of 16-gene score with OS is independent of clinicopathological characteristics.

Identification of signalling pathways associated with the 16-gene score

We performed the GSEA analysis to understand the biological functions related to the 16-gene score. The results exhibited thirteen signalling pathways were significantly over-represented in the high 16-gene score group of the TCGA cohort. Cell cycle, RNA degradation, oocyte meiosis, progesterone mediated oocyte maturation and DNA replication were the top five most enriched pathways (Figure4, q value <0.25 for all cases, supplementary Table4). While, up-regulation of arachidonic acid metabolism pathway genes were significantly associated with the low 16-gene score in the TCGA cohort (Figure4, q value <0.25, supplementary Table5). These results suggest that the aforementioned pathways probably are implicated in the association between 16-gene score and OS in BRCA.

Nomogram combined 16-gene signature and clinical-related variables predicts patients' OS

In the TCGA and METABRIC cohorts, patient's age, tumor stage, menopause status, number of positive lymph nodes and 16-gene signature were significantly associated with OS. Then based on the above analysis results, we established a 16-gene nomogram that incorporated the survival-related clinical factors and 16-gene signature (Figure 5A). The nomogram predicted well the 1-year, 3-year and 5-year survival for BRCA patients in the TCGA cohort, ROC plot revealed the 16-gene nomogram showed improved prediction accuracies for 1-year, 3-year and 5-year survival as compared to the 16-gene score alone (AUC: 0.91, 0.79 and 0.77 respectively, Figure5B). The improved prognosis prediction was also validated in the METABRIC cohort (AUC: 0.83, 0.77 and 0.76 respectively, Figure5C), demonstrating the clinical value and validity of the 16-gene nomogram for OS evaluation of BRCA patients.

Assessment of diagnostic value

We utilized the online server cBioPortal to investigate the genomics variants of 16 genes from the TCGA datasets. The results showed that *DERL1, TNN, PXDNL, PCSK6* and *KLRB1*were the top five most frequently mutated genes, with mutation frequencies of 19%,10%, 9% 4%, 3% respectively in BRCA (supplementary Figure5). Similar mutation distribution was observed in the METABRIC cohort

(supplementary Figure6). By comparing expression levels of 16 genes between 779 BRCA samples and 100 paired normal breast tissues, 7 genes expression, such as *C7orf63, C9orf103, IGJ, ZNF385B* and *TNN,* was significantly lower in tumor tissues as compared with those in normal tissues. In contrast, 9 genes, such as *PXDNL, PCSK6, MORN3* and *DERL1*, were significantly higher expressed in BRCA tissues (adjusted P< 0.05 for all cases, student t test, Figure6A). ROC curves analysis further showed *MORN3, IGJ, DERL1* particularly were able to differentiate BRCA tissues from normal breast tissues with high accuracy (Figure6B, adjusted P values <0.05, AUC>0.80 for all cases).

Discussion

BRCA is a heterogeneous disease with several molecular subtypes, each of which has its distinct biological and clinical characteristics (18). The identification of reliable prognostic biomarkers would enable to prioritize patients at high risk for death and relapse and guide treatment. The traditional methods for the risk stratification include tumor size, tumor stage, lymph node metastasis and molecular subtype, which could be applicable to certain subgroup of BRCA, however, there is still lack of a prognostic model that could be applicable to almost all BRCA subtypes. Recent studies have shown gene expression profiles could serve as prognostic biomarkers in BRCA (19, 20). However, the accuracies of the previous gene profiles are still relatively low. In the current study, we have successfully established the 16-gene score which is correlated with poor OS, DFS and PRS in BRCA. We also demonstrated that the prognostic value of the 16-gene score was independent of clinical factors and applied to all subtypes of BRCA patients, which is advantageous to the MammaPrint model and Oncotype DX that show applicability to limited disease subtypes. Furthermore, we established a 16-gene nomogram that incorporated the survival-related clinical factors and 16-gene signature. As compared to established gene profiles, the 16-gene nomogram (AUC = 0.91, 0.79 and 0.77, respectively) performed better than the Teschendorff's (21)(AUC = 0.44, 0.47, 0.50, respectively) and Bianchini's (22) immune-related gene signatures (AUC = 0.53, 0.56, 0.51, respectively) (19) and cancer stage (AUC = 0.71, 0.69, 0.66, respectively) in predicting the 1-year, 3-year and 5-year survival of BRCA patients. Therefore, the 16-gene nomogram might be a reliable and useful prognostic tool for OS evaluation and will promote tailored therapy for all subtypes of BRCA patients.

The mechanisms by which the higher 16-gene score is associated with poor prognostic implication remain to be poorly understood. The GESA analysis uncovered cell cycle, RNA degradation, oocyte meiosis, progesterone mediated oocyte maturation and DNA replication were significantly over-represented in the high 16-gene score group. Cell cycle checkpoints are critical for ordered cell cycle progression, which ensures genomic stability and inhibits the process of carcinogenesis(23). The deregulation of the cyclin-dependent kinase inhibitors p21 and p27, cyclins D1 and E frequently exerts negative impacts on BRCA outcome and response to therapy (24). We believe the dysregulation of cell cycle signalling pathway largely contribute to the prognostic value of 16-gene score in BRCA.

Apart from prognostic value, the 16 genes might serve as diagnostic biomarkers for BRCA patients. Our study revealed that *MORN3, IGJ, DERL1* showed high accuracy in differentiating BRCA tissues from normal breast tissues. *DERL1* is involved in the elimination of misfolded proteins and has been implicated in the progression of human cancers. *DERL1* has been found to be overexpressed in several human cancers and exhibits oncogenic activities (25–28). Increased expression of *DERL1* is correlated with lymph node metastasis, advanced clinical stage, and unfavorable overall survival (25, 27, 28). Enhanced expression of *DERL1* promotes the progression of BRCA (25, 29) and colon cancer (28) and hepatocellular carcinoma (30). These results demonstrate *DERL1* functions as an oncogene in cancers, therefore, the gene may become a druggable target for BRCA patients.

Conclusion

Taken together, this study identified a novel 16 gene signature that could serve as an independent factor for predicting BRCA prognosis independently of clinical characteristics. The gene set related to the high-risk group participated in the cell cycle signal pathway.

Abbreviations

Breast cancer: BRCA

The Cancer Genome Altas: TCGA

The Molecular Taxonomy of BRCA International Consortium: METABRIC

Overall survival: OS

The least absolute shrinkage and selection operator: LASSO

Gene Ontology: GO

Kyoto Encyclopedia of Genes and Genomes: KEGG

Receiver operating curves: ROC

Area under curve: AUC

Odd ratio: OR

Confidence interval: CI

Gene set enrichment analysis: GSEA

Declarations

Acknowledgement

None

Funding

None

Ethics approval and consent to participate

None

Consent for publication

None

Competing interests

The authors declare no competing interests.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the figshare repository (figshare ID: 15048003, https://figshare.com/s/df0ee21997f1aa0da4bd).

Authors' contributions

Fengfeng Fan conceived the study. Zhenhua Zhong and Wenqiang Jiang performed the survival analyses. Zhenhua Zhong, Wenqiang Jiang, Jing Zhang developed the 16-gene score and nomogram and implemented the validation analysis. Zhanwen Li[1], Fengfeng Fan performed the GSEA analysis. Zhenhua Zhong wrote the manuscript. All authors read and approved the final manuscript.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin [Internet]. 2019,69(1):7–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30620402
2. Zheng RS, Sun KX, Zhang SW, Zeng HM, Zou XN, Chen R, et al. [Report of cancer epidemiology in China, 2015]. Zhonghua Zhong Liu Za Zhi. 2019 Jan,41(1):19–28.
3. Kast K, Link T, Friedrich K, Petzold A, Niedostatek A, Schoffer O, et al. Impact of breast cancer subtypes and patterns of metastasis on outcome. Breast Cancer Res Treat [Internet].

2015,150(3):621–9. Available from: https://doi.org/10.1007/s10549-015-3341-3

4. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer. J Natl Cancer Inst. 2006,98(17):1183–92.
5. Toole MJ, Kidwell KM, Van Poznak C. Oncotype Dx results in multiple primary breast cancers. Breast Cancer Basic Clin Res. 2014,8(1):1–6.
6. Kwa M, Makris A, Esteva FJ. Clinical utility of gene-expression signatures in early stage breast cancer. Nat Rev Clin Oncol [Internet]. 2017,14(10):595–610. Available from: http://dx.doi.org/10.1038/nrclinonc.2017.74
7. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018,173(2):291-304.e6.
8. Pereira B, Chin SF, Rueda OM, Vollan HKM, Provenzano E, Bardwell HA, et al. The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. Nat Commun. 2016,7(May).
9. Lai Y, Sheng L, Wang J, Zhou M, OuYang G. A Novel 85-Gene Expression Signature Predicts Unfavorable Prognosis in Acute Myeloid Leukemia. Technol Cancer Res Treat. 2021,20:15330338211004932.
10. Therneau T. Survival Analysis. Cran [Internet]. 2016, Available from: https://cran.r-project.org/web/packages/survival/survival.pdf
11. Fox J. Cox Proportional-Hazards Regression for Survival Data The Cox Proportional-Hazards Model. Most [Internet]. 2002,2008(June):1–18. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.2264&amp,rep=rep1&amp,type=pdf
12. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics [Internet]. 2011,12(1):77. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-12-77
13. Reimand J, Kull M, Peterson H, Hansen J, Vilo J. G:Profiler-a web-based toolset for functional profiling of gene lists from large-scale experiments. Nucleic Acids Res. 2007,35(SUPPL.2):193–200.
14. Sha K, Lu Y, Zhang P, Pei R, Shi X, Fan Z, et al. Identifying a novel 5-gene signature predicting clinical outcomes in acute myeloid leukemia. Clin Transl Oncol. 2020 Aug,
15. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. J Stat Softw. 2011 Mar,39(5):1–13.
16. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci [Internet]. 2005 Oct 25,102(43):15545–50. Available from: http://www.pnas.org/content/102/43/15545.abstract
17. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal [Internet]. 2013 Apr 2,6(269):pl1–pl1. Available from: https://pubmed.ncbi.nlm.nih.gov/23550210

18. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: A collaborative analysis of data for 10,159 cases from 12 studies. PLoS Med. 2010,7(5).
19. Xie P, Ma Y, Yu S, An R, He J, Zhang H. Development of an Immune-Related Prognostic Signature in Breast Cancer. Front Genet. 2020,10(January):1–14.
20. Shimizu H, Nakayama KI. A 23 gene–based molecular prognostic score precisely predicts overall survival of breast cancer patients. EBioMedicine [Internet]. 2019,46:150–9. Available from: https://doi.org/10.1016/j.ebiom.2019.07.046
21. Teschendorff AE, Gomez S, Arenas A, El-Ashry D, Schmidt M, Gehrmann M, et al. Improved prognostic classification of breast cancer defined by antagonistic activation patterns of immune response pathway modules. BMC Cancer. 2010 Nov,10:604.
22. Bianchini G, Iwamoto T, Qi Y, Coutant C, Shiang CY, Wang B, et al. Prognostic and therapeutic implications of distinct kinase expression patterns in different subtypes of breast cancer. Cancer Res. 2010 Nov,70(21):8852–62.
23. Malumbres M, Carnero A. Cell cycle deregulation: a common motif in cancer. Prog Cell Cycle Res. 2003,5(April 2021):5–18.
24. Caldon CE, Daly RJ, Sutherland RL, Musgrove EA. Cell cycle control in breast cancer cells. J Cell Biochem. 2006,97(2):261–74.
25. Zeng J, Tian Q, Zeng Z, Cai J, Ye M, Liu Y, et al. Derlin-1 exhibits oncogenic activities and indicates an unfavorable prognosis in breast cancer. Cell Biol Int [Internet]. 2020,44(2):593–602. Available from: http://dx.doi.org/10.1002/cbin.11259
26. Li L, Liu M, Zhang Z, Zhang W, Liu N, Sheng X, et al. Derlin1 functions as an oncogene in cervical cancer via AKT/mTOR signaling pathway. Vol. 52, Biological research. 2019. p. 8.
27. Mao M, Zhang J, Jiang J. Overexpression of Derlin-1 is Associated with Poor Prognosis in Patients with Non-small Cell Lung Cancer. Ann Clin Lab Sci. 2018 Jan,48(1):29–34.
28. Tan X, He X, Jiang Z, Wang X, Ma L, Liu L, et al. Derlin-1 is overexpressed in human colon cancer and promotes cancer cell proliferation. Mol Cell Biochem. 2015 Oct,408(1–2):205–13.
29. Liu Y, Wang Z, Liu H, Wang X, Zhang Z, Xiao B, et al. Derlin-1 functions as a growth promoter in breast cancer. Biol Chem. 2020 Feb,401(3):377–87.
30. Fan J, Tian L, Huang S, Zhang J, Zhao B. Derlin-1 Promotes the Progression of Human Hepatocellular Carcinoma via the Activation of AKT Pathway. Onco Targets Ther. 2020,13:5407–17.

Tables

Table1. Association between the clinical features and patients' mortality in 1080 BRCA patients of the TCGA dataset

Variables	Alive	Dead	P value
Age	57.95	61.22	0.01[a]
Tumor weight	367.54	395.57	0.5[a]
Number of positive lymph nodes	2.07	4.3	<0.001[a]
Menopausal stage			<0.001[b]
Indeterminate	19	15	
Peri-menopause	38	1	
Post-menopause	602	91	
Pre-menopause	209	18	
ER status			
Positive	694	100	0.07 [b]
Negative	196	41	
HER2 status			
Positive	137	23	0.15 [b]
Negative	499	57	
PR status			
Positive	600	86	0.1 [b]
Negative	286	56	
Cancer stage			<0.001 [b]
I	163	16	
II	549	66	
III	202	46	
IV	4	15	
T stage			<0.001 [b]
T1	241	33	
T2	550	77	
T3	112	25	
T4	24	15	

N stage			
N0	467	44	<0.001
N1	296	59	
N2	96	22	
N3	61	15	
M stage			
M0	772	120	<0.001
M1	10	17	
Chemotherapy			
Yes	450	36	0.89 [b]
No	257	22	
Hormone therapy			
Yes	247	17	0.47 [b]
No	460	41	
Radiotherapy			
Yes	374	58	0.05 [b]
No	497	51	

a and b indicate student t test and fisher exact test respectively.

Figures

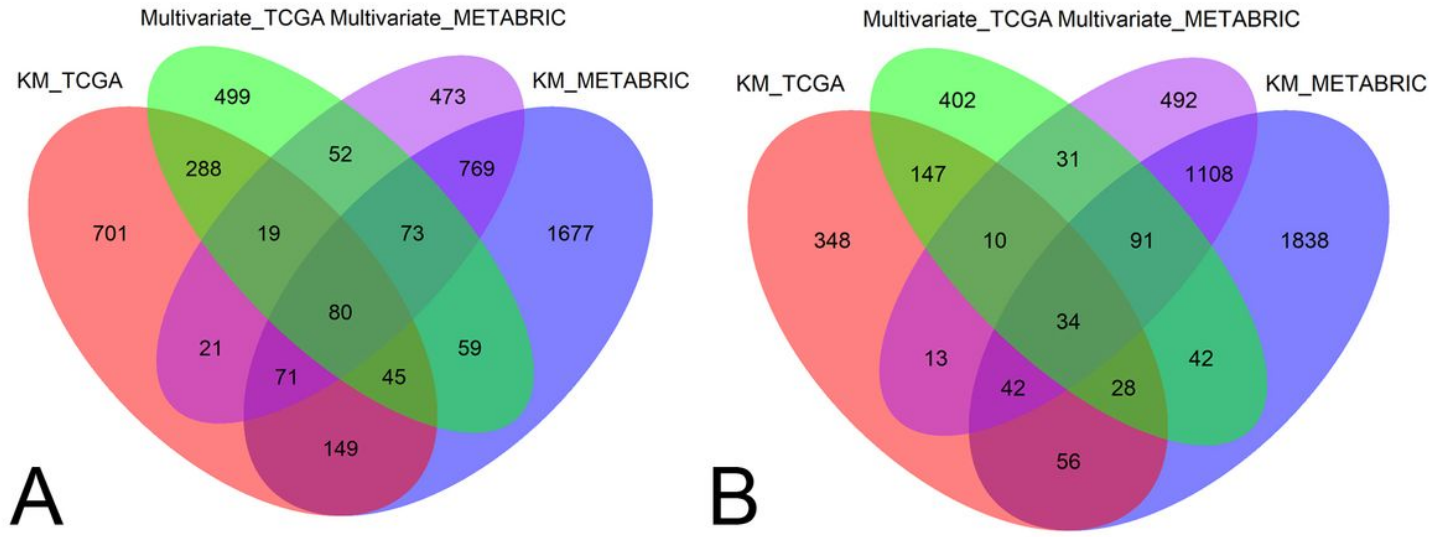

Figure 1

The prognosis-related genes common to the TCGA and METABRIC datasets. A. The protective prognostic genes common to the TCGA and METABRIC datasets, B. The risk prognostic genes common to the TCGA and METABRIC datasets. KM_TCGA and multivariate_TCGA represent prognosis-related genes determined by the Kaplan-Meier survival analysis and multivariate Cox regression analysis respectively in the TCGA cohort. Similarly, KM_METABRIC and multivariate_METABRIC denote survival-related genes in the METABRIC cohort.

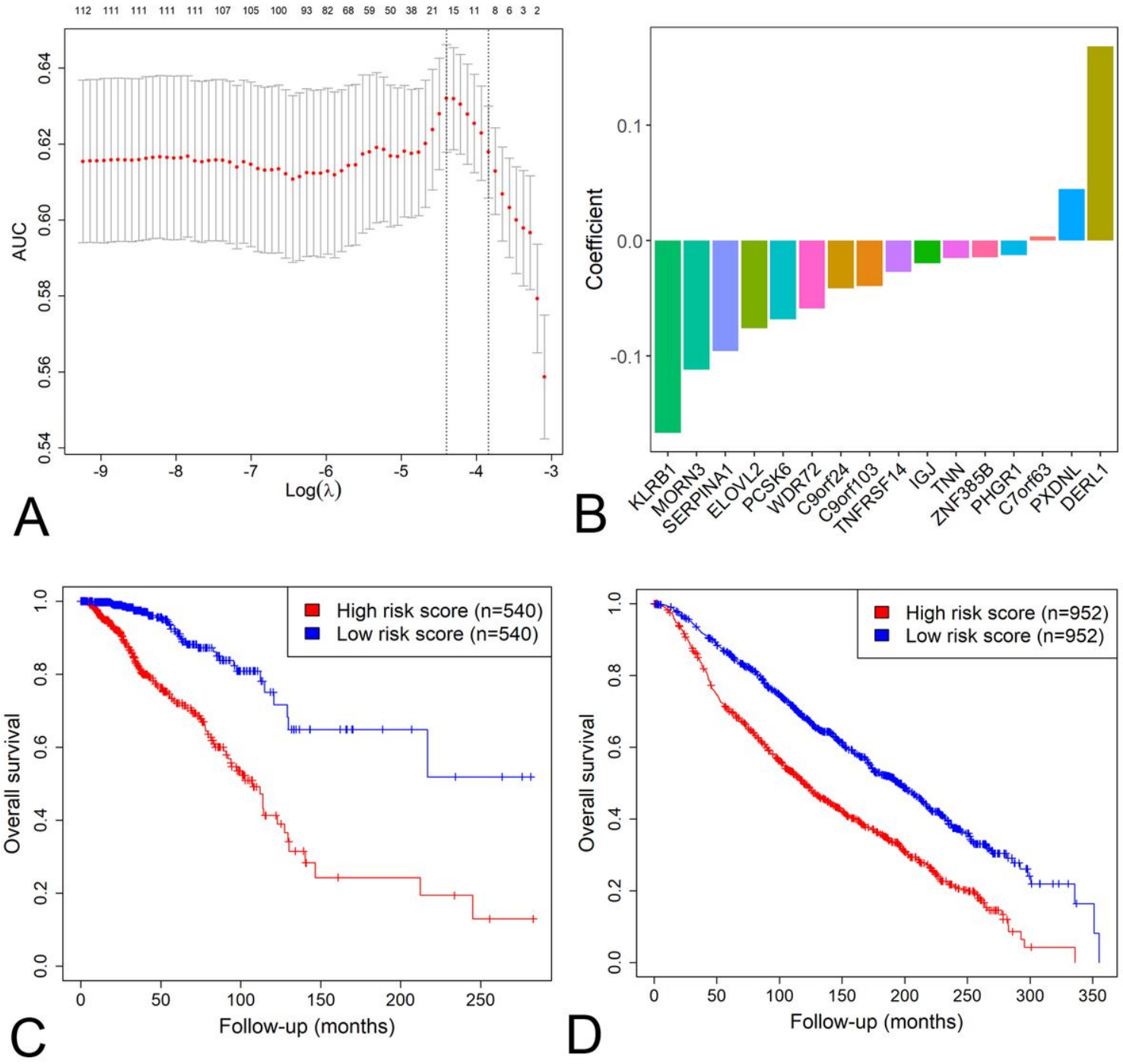

Figure 2

The 16-gene score is an indicator of infavorable survival in BRCA. A. The relationship among AUC, log scaled lambda values and number of genes with non-zero coefficients in the LASSO model. The x axis represents the log value of the independent variable λ, whilst the y axis represents the partial likelihood deviance of the log value of each independent variable λ. B. The coefficients of 16 genes in the LASSO model. C. Kaplan-Meier survival curve of patients' OS for BRCA patients with different 16-gene scores in

the TCGA cohort, D. Kaplan-Meier survival curve of patients' OS for BRCA patients with different 16-gene scores in the METABRIC dataset.

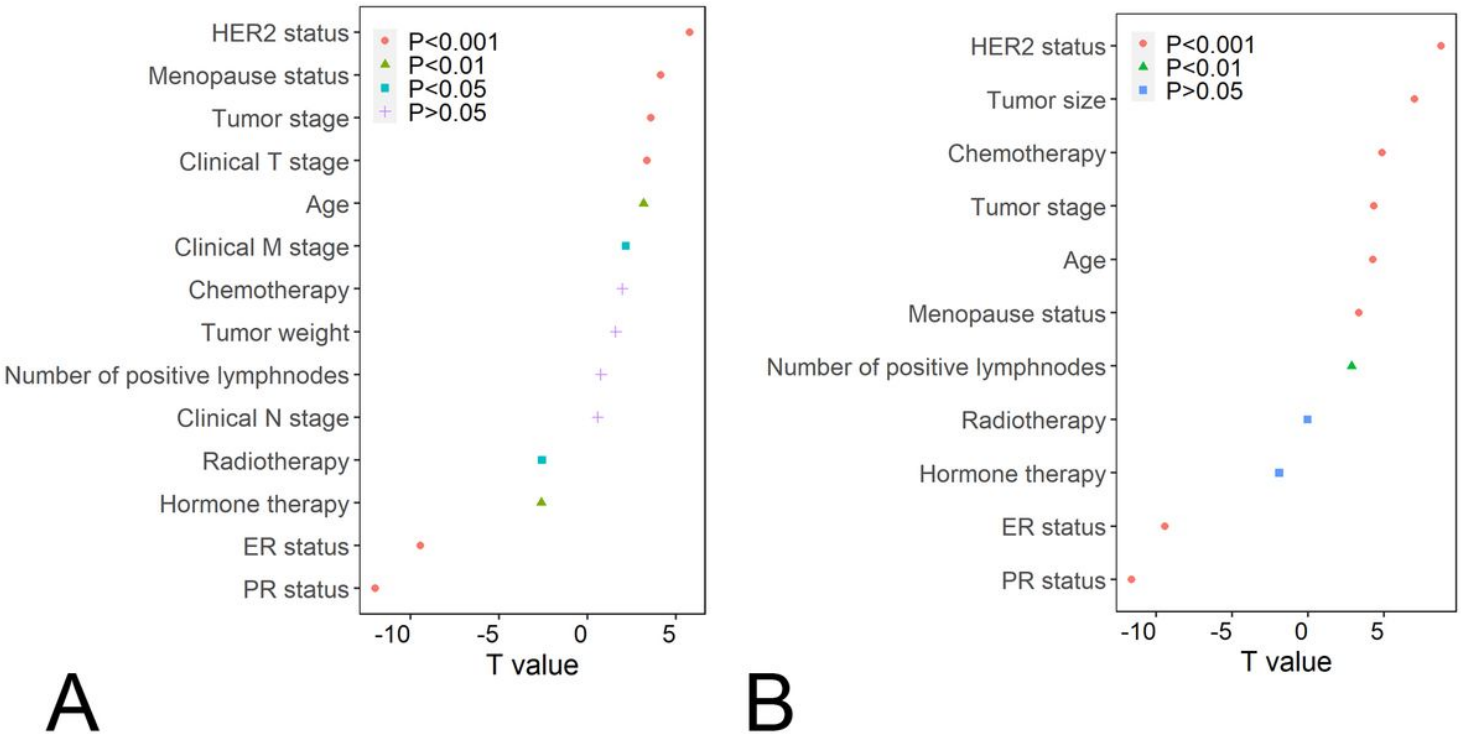

Figure 3

The correlations between the clinical characteristics and the 16-gene score. A. The associations between clinical characteristics and the 16-gene score in the TCGA cohort. B. The associations between clinical characteristics and the 16-gene score in the METABRIC cohort.

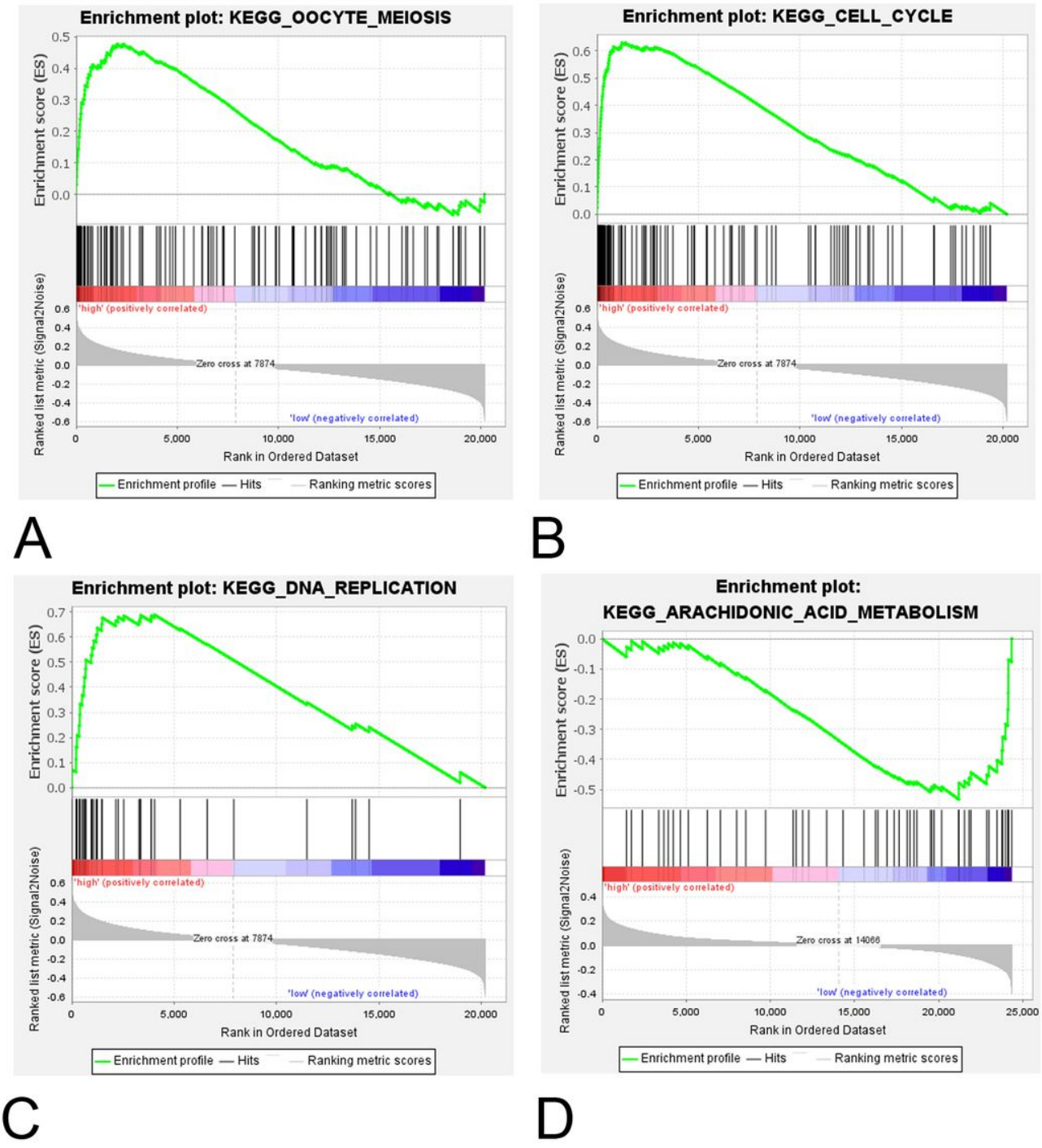

Figure 4

GSEA analysis revealed three significantly enriched pathways related to the high 16-gene score, including oocyte meiosis (A), cell cycle (B), DNA replication (C), and the significantly up-regulated arachidonic acid metabolism (D) associated with the low 16-gene score. For each gene set, vertical bars along the x-axis represent where the genes locate within the ranked list. Negative enrichment score indicates down-regulation, while, positive value denotes up-regulation of the gene set.

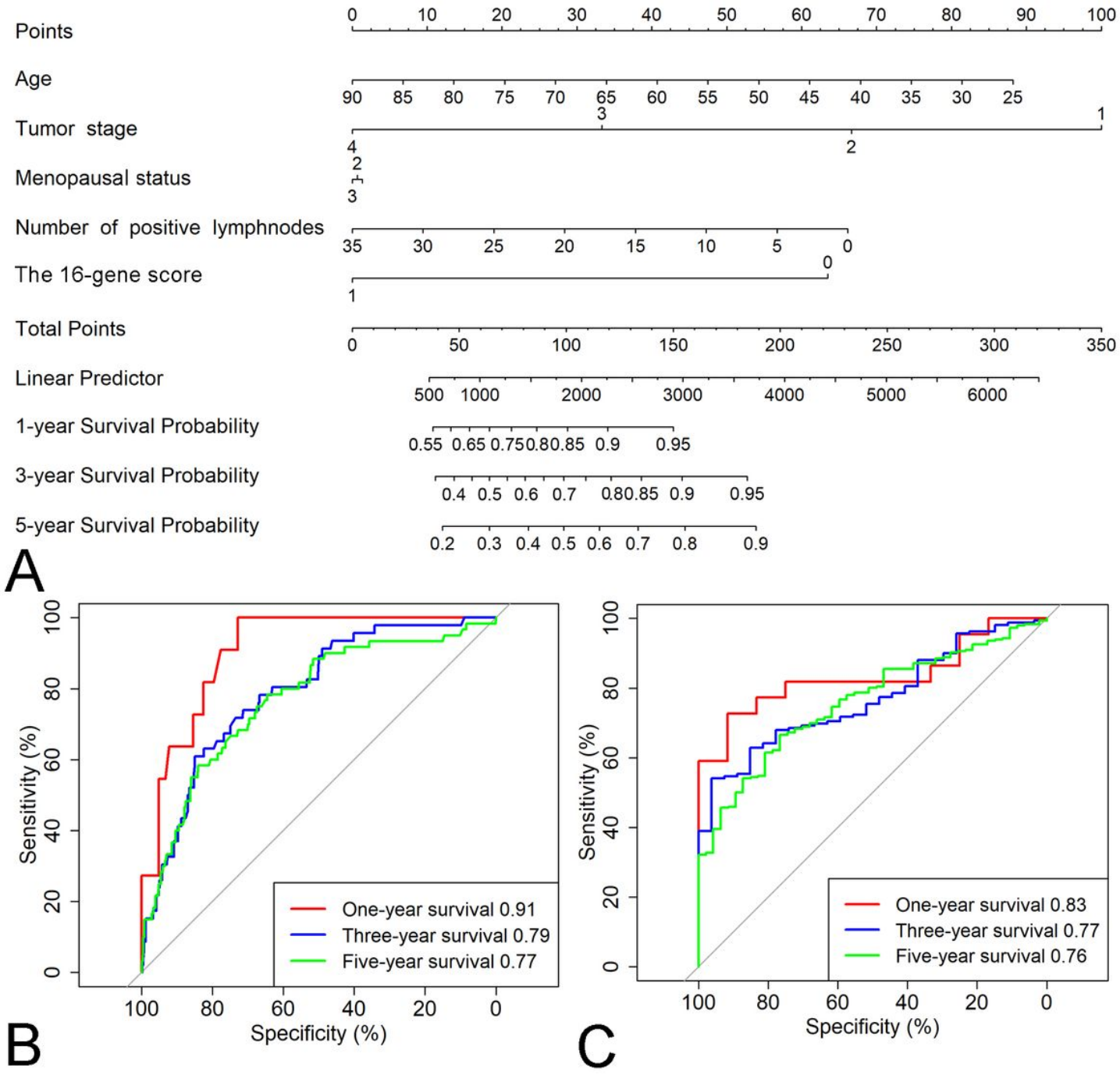

Figure 5

The 16-gene nomogram to predict the risk of disease in patients with BRCA. (A) mRNA nomogram to predict disease-free survival. 1,2,3 for the menopausal status denote pre-menopause, peri-menopause and post-menopause respectively. 0 and 1 for the 16-gene score represent high and low 16-gene scores respectively which were divided by the median 16-gene score. B. The ROC plot for the nomogram in predicting of 1-year, 3-year and 5-year survival in the TCGA cohort. C. The ROC plot for the nomogram in predicting of 1-year, 3-year and 5-year survival in the METABRIC cohort.

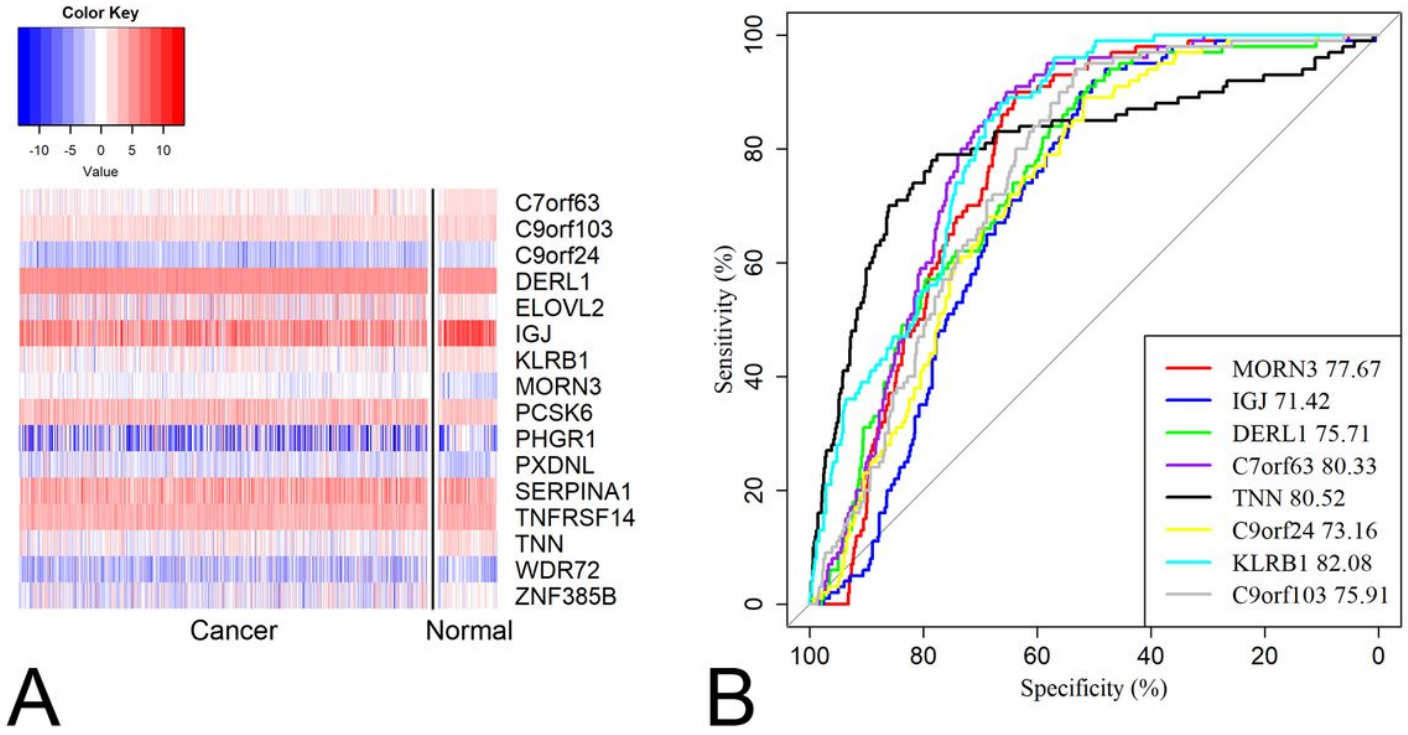

Figure 6

Gene expression analysis of prognosis-related genes. A. The heatmap shows the expression differences of 16 prognosis-related genes between BRCA tissues and paired normal breast tissues. B. The ROC curves for the top eight prognosis-related genes showing the highest diagnostic capability.

Supplementary Files

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

- Supplementaryfigures.docx
- Supplementarytables.docx