Exploration of Tumor Microenvironment-Related Biomarkers for the Prognosis of Triple Negative Breast Cancer

Xiaorui Han
South China University of Technology
Zaiyi Liu
Guangdong Provincial People's Hospital
Changhong Liang (liangchanghong@gdph.org.cn)
Guangzhou Province People's Hospital https://orcid.org/0000-0001-8267-150X

Primary research

Keywords: TNBC, TME, overall survival, risk score

Posted Date: May 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-510169/v1

Abstract

Background: Triple negative breast cancer (TNBC) is one of the most disastrous breast cancer subtypes world widely. The tumor microenvironment (TME), especially the infiltration of immune and stromal cells, are highly related to the occurrence, development and prognosis of breast cancer. Therefore, exploration of TME-related biomarkers is greatly important for improving overall survival rate of TNBC patients.

Methods: The open-assess Cancer Genome Atlas (TCGA) database provides gene expression profile for a variety of malignant tumors allowing researchers to explore genes demonstrating TME in the prognosis prediction of TNBC. In our present study, ESTIMATE algorithm was used to calculate the immune and stromal scores in accordance with the characteristics of specific genes in immune and stromal cells, and divide them into high and low-score groups. Limma R package was then utilized to screen differentially expressed genes (DEGs). After that, functional enrichment analysis and protein-protein interaction (PPI) network were performed to explore the bio-information of the DEGs and their encoded proteins. Subsequently, the identified these genes were further verified in the Gene Expression Omnibus (GEO) datasets.

Results: Eight genes, including *ACAP1, DUSP1, LYZGZMA, SASH3, CCL5, CD74,* and *DPT*, were explored to closely related to the TME of TNBC. A prognostic model containing these selected genes was established with a high efficiency in the prediction of the poor prognosis of TNBC patients.

Conclusion: An eight-gene prognostic model was a considerably reliable approach for predicting the overall survival of TNBC patients, and could help clinicians selecting personalized treatments for their TNBC patients.

Introduction

Accounting for about 15–20% of breast cancers, triple negative breast cancer (TNBC) is currently the most aggressive and lethal breast cancer subtype, posing a great threat to women's health[1]. Previous studies have showed that the TNBC had an increased likelihood of distant metastasis, the peak being within the first 3 years after onset. The mortality of TNBC within 5 years is significantly higher than other types of breast cancer[2, 3]. Multiple factors like age, menopause, tumor size, histological grade, and pathological stage play roles in the short and long term of clinical outcomes[4]. However, due to the extremely complex molecular mechanisms in tumor as well as the limited predictive strength of conventional clinical information[5], there is a greatly urgent need to explore more advanced techniques to accurately predict the prognosis of TNBC patients and develop personalized precise therapies.

Genomic sequencings and practicable datasets are emerging in this era of genomics [6]. Those tools have made great contributions to tumor diagnosis and prognosis prediction. The genes inherent in tumor cells, especially the main transcription factor, determine the initiation, progression and evolution of TNBC[7, 8]. Being similar to the solid tumor portions manifested the progressive level of tumor, tumor microenvironment (TME) contributing to tumor occurrence, development, and treatment have been widely

reported in recent year [9–11]. TME is a complex biosystem composed of extracellular matrix, stromal cells (such as fibroblasts, adipocytes and mesenchymal stromal cells) and immune cells (such as B and T lymphocytes, macrophages and natural killer cells)[12]. Immune cells and stromal cells are two main types of non-tumor components in TME, which are considered to be of great significance for the diagnosis and prognostic evaluation of tumors [13]. The range of stromal cells is a sensitive prognostic predictor for patients with solid cancer [16]. To be specific in breast cancer, higher immune infiltration indicates better clinical outcomes for the patients, especially in ER-negative patients. Higher CD8 + T cell infiltration indicates better overall survival (OS) of breast cancer patients [14]. Patients with hyperimmune infiltration showed satisfactory response to the neoadjuvant chemotherapy and adjuvant chemotherapy [15]. Therefore, in term of the special role of TME, it is crucial to accurately predict the purity and prognosis of TNBC and develop individualized treatment plans for different individuals with different TME.

In order to assess the infiltration of immune cells and stromal cells, an online ESTIMATE (Estimation of Stromal and Immune cells in Malignant Tumor tissues using Expression data) algorithm was created to calculated the immune scores and stromal scores in tumor. By analyzing the expression characteristics of specific genes in immune and stromal cells, ESTIMATE algorithm is able to predict the infiltration of non-tumor cell [17]. Until now, it has been applied to the analysis of multiple solid cancers, including prostate cancer [18], breast cancer[11, 19], and colon cancer [20], demonstrating the potential efficiency of this data-based algorithm. In the most recent research, ESTIMATE was utilized in detection of TME of luminal breast cancer. A ceRNA network associated with the TME, consisting with miRNAs, lncRNAs and mRNAs was identified to have prognostic value for the luminal breast cancer.

In this study, we aimed to explore the expression profiles of high and low immune and stromal score groups in TNBC and identify potential genetic biomarkers using data in cBioPortal. By RNA-sequencing (RNA-seq) survival analysis, we explore eight gene eight-gene prognostic model including *ACAP1, DUSP1, LYZ, GZMA, SASH3, CCL5, CD74*, and *DPT*; the model was validated in the GEO data set. This prognostic model is promising to accurately predict the prognostic status of TNBC patients for the TME of different individuals.

Results

Demographic characteristics of patients

A total of 299 samples with pathological diagnosis of TNBC were downloaded and screened from the cBioPortal website. Table 1 describes the patient's detailed demographic characteristics and baseline characteristics.

Table 1
the baseline characteristics of the enrolled subjects

PARAMETER	Subtype	No.	percent(%)	Immune scores		Stromal scores	
				Median	p	Median	p
AGE AT DIAGNOSIS (YEARS)	60	180	60.2	1514.55	0.89	640.45	0.06
	≥ 60	119	39.8	1503.13		636.39	
LATERALITY	left	139	46.5	1555.89	0.61	647.88	0.73
	right	142	47.5	1512.93		670.49	
	NA	18	6	/	/	/	/
RISK	high	149	50	1464.5	0.27	722.99	0.01
	low	150	50	1555.21		555.23	
CELLULARITY	high	164	55	1463.64	0.59	489.39	0.00
	Moderate	90	30	1541.04		800.03	
	low	36	12	1601.81		887.49	
	NA	9	3	/	/	/	/
LYMPH NODES	N0	145	48.5	1469.26	0.51	619.96	0.65
	N1	103	34.5	1505.58		623.88	
	N2	30	10	1682.77		759.11	
	N3	21	7	1539.55		663.86	
TUMOR SIZE	T1	123	41	1661.03	0.006	703.62	0.18
	T2	154	52	1414.52		606.21	
	T3	18	6	1273.52		483.03	
	NA	4	1	/	/	/	/
OS STATUS	LIVING	138	46.2	1594.48	0.06	660.73	0.54
	DECEASED	161	53.8	1437.6		620.06	
CLAUDIN SUBTYPE	claudin-low	106	35.5	2108.5	0.00	1092.95	0.00
	Basal	151	50.5	1185.97		291.13	
	Her2	30	10	1286.39		725.68	
	LumA	2	0.7	558.17		944.06	

A $p < 0.05$ was considered statistically significant.

PARAMETER	Subtype	No.	percent(%)	Immune scores		Stromal scores	
				Median	p	Median	p
	Normal	10	3.3	920.18		793.83	
HISTOLOGICAL SUBTYPE	Ductal/NST	254	85	1472.82	0.00	612.44	0.02
	Lobular	14	4.7	1362.14		630.82	
	Medullary	15	5	2138.19		752.74	
	Mucinous	1	0.3	1974.1		2141.37	
	Mixed	11	3.7	2008.9		1038.58	
	Other	4	1.3	545.5		440.44	
GRADE	1	3	1	1451.74	0.28	619.28	0.003
	2	36	12	1502.79		957.73	
	3	257	86	1521		591.58	
	NA	3	1	/	/	/	/
TUMOR_STAGE	1	62	20.7	1670.7	0.25	760.45	0.04
	2	130	43.5	1477.91		650.57	
	3	25	8.4	1491.62		726.04	
	NA	82	27.4	/	/	/	/
TOTAL	/	299	100	1510.01	/	638.83	/

A $p < 0.05$ was considered statistically significant.

Calculation of immune / stromal score and its different distribution in clinical characteristics

After ESTIMATE algorithm, the ranges of the immune score and stromal score were − 915.21 to 2141.37 and − 698.57 to 3079.68, respectively. Next, we compared the different distributions of these scores in terms of clinical characteristics, including age, left / right position, OS status, cellularity, claudin subtype, histological subtype, grade, tumor stage, tumor size, lymph node (Table 1, Fig. 1). In all the characteristics shown in Table 1, we found that the distribution of immune scores and stromal scores in histological subtypes and CLAUDIN subtypes are different ($p < 0.05$), and the comparison results of each subtype are shown in Table 2. In addition to this, the stromal infiltration score also varies in diagnostic cell count, grade and stage ($p < 0.05$), but the immune infiltration score is similar between them ($p > 0.05$); the immune infiltration score between different tumor sizes are also different ($p < 0.05$) (Table 1). Kaplan-Meier chart was used to analyze the relationship between the immune / stromal score and the

corresponding overall survival rate. The results showed that a lower immune score was significantly associated with a lower OS (Fig. 2, $p < 0.05$). However, no difference in survival data was observed in immune score and ESTIMATE score ($p = 0.08$ and 0.42, respectively) (Fig. 2).

Table 2
multiple comparisons

Immune score	Claudin subtype	Claudin subtype	Mean difference	P value	95% CI	
					Lower	Upper
	basal	Claudin-low	-922.52	0.00	-1122.56	-722.49
		Her-2	-100.42	1.00	-415.97	215.13
		LumA	627.81	1.00	-495.81	1751.42
		normal	265.79	1.00	-249.67	781.26
	Claudin-low	Her-2	822.10	0.00	495.64	1148.56
		LumA	1150.33	0.00	423.60	2677.06
		normal	1188.32	0.00	666.10	1710.53
	Her-2	LumA	728.22	0.75	-424.63	1881.08
		normal	366.21	0.73	-210.21	942.64
	LumA	normal	-362.01	1.00	-1584.80	860.77
CI: confidence interval						

Comparison of TNBC gene expression profile with immune score and stromal score

To reveal the correlation between the overall gene expression profile and the immune score and / or stromal score, we compared the Affymetrix microarray data of 299 TNBC patients obtained from the cBioPortal website. The heat map in Fig. 3 showed the different gene expression profiles of cases in the high and low immune / stromal score groups. 164 DEGs (| logFC |> 1, $p < 0.05$) were found in the comparison of immune scores. Similarly, 124 DEGs were found in the comparison of stromal scores (| logFC |> 1, $p < 0.05$). However, the DEGs extracted from the comparison of the high and low immune / stromal score groups were significantly different, the two groups of DEGs data were combined for the subsequent analysis.

Function enrichment analysis and construction of the PPI network

In order to explore the biological significance of DEGs, we performed GO function enrichment analysis on DEGs in the immune scores and stromal scores. The results showed that DEGs in the immune scores were significantly enriched in biological processes such as the T cell activation, the regulation of leukocyte activation, the side of membrane, the antigen binding ($p < 0.05$) (Fig. 4); DEGs in the stromal scores were significantly enriched in biological processes such as the extracellular matrix organization, the extracellular matrix, the extracellular matrix structural constituent ($p < 0.05$) (Fig. 4). In addition, in order to explore the relevant pathways of the TNBC microenvironment, we conducted a KEGG pathway analysis of all DEGs ($p < 0.05$). The Fig. 5 showed the main pathways involved such as the phagosome, the cytokine-cytokine receptor interaction, and the chemokine signaling pathway.

In order to explore the protein interactions related to the TNBC microenvironment, we conducted a PPI network analysis (interaction score > 0.9) of DEGs on the STRING online website. We selected the hub genes by the number of node connections and displayed them with barplot charts (Fig. 6). We found that most of these hub genes were important immune-related factors, indicating that TNBC was very immunogenic and had strong immune characteristics.

Establishment of gene-related prognostic models

By using single factor Cox survival analysis of DEGs related to TNBC microenvironment, 39 genes significantly related to survival were obtained ($p < 0.05$). Then, the top 20 were taken to enter into the multifactor regression analysis, and the stepwise regression method was used to establish the best model. The model finally included eight genes (Fig. 7). The predictive model was characterized by linear combination based on the expression levels of eight genes weighted by their relative coefficients, resulting in the following model:

*risk score (RS) = -0.5076 * ACAP1 + 0.2347 * DUSP1 + 0.2336 * LYZ − 0.2973 * GZMA + 0.5710 * SASH3 -0.2512 * CCL5 + 0.5367 * CD74 -0.4422 * DPT.*

The risk value of each patient was calculated according to the risk scoring formula, and it was divided into a high-risk group (n = 149) and a low-risk group (n = 150) according to the median value of the patients. There were significant differences in the KM plot survival curves of the two groups (the median OS was 3.01 years and 4.54 years, $p < 0.05$; Fig. 8A). The prognostic survival of the high-risk group was significantly lower than that of the low-risk group. The ROC curve was used to evaluate the efficacy of RS in predicting 5-year survival of patients. The AUC value reached 0.737, indicating that the model was accurate and reliable (Fig. 8B).

Verification of the prognostic model

The breast cancer dataset (No. GSE103091) was downloaded from the GEO, and 107 TNBC samples was selected. After that, the established model was validated depending on the newly enrolled cases. In brief, the risk value of each patient was calculated using the above risk score formula, and the accuracy was evaluated to predict the 5-year survival of the subjects. As shown in the results, the AUC of the

established model reach as high as 0.636, indicating its broadly applied efficiency in TME estimation for TNBC patients (Fig. 9).

Independence analysis of prognostic model and other clinical characteristics

The clinical information of 277 patients in the cBioPortal TNBC cohort was downloaded, and the univariate and multivariate Cox regression analysis was used to evaluate the independent predictive value of the prognostic model of TNBC. The findings of the univariate Cox regression analysis indicated that the prognostic model and histological subtype showed considerable prognostic value ($p < 0.05$). In contrast, the diagnosis age, left/right position, grade, stage, tumor size, and number of lymph nodes were not related to OS. Therefore, the prognostic model and the histological subtype were taken into a multifactor Cox regression analysis to clarify their specific effectiveness with OS. In our further analysis, we concluded that prognostic model and histological subtype were the independent prognostic factors related to OS (Fig. 10).

Discussion

TNBC is still the serious subtype of breast cancer due to its complex molecular and cellular heterogeneity with a increasing incidence worldwider [24]. Previous studies have shown that TME is highly correlated with the occurrence, progression and prognosis of breast cancer [25]. TME is where the immune system interacts with the tumor, indicating that the plastic cells of the tumor and immune system are an important part of TME. Undoubtably, the TME play a vital role in the development of various kinds of cancer, expecially in breast cancer [13, 26]. Therefore, in our present study, we used the TNBC data obtained from the open-access cBioPortal website to identify specific TME-related genes. Those genes represent the bioactivities of immune and stromal components in TME and might pose a great impact on the prognosis of the TNBC.

In our present study, we calculated the immune and stromal scores in TME by the ESTIMATE algorithm to investigate the infiltrated level of immune and stromal cells in TME of TNBC. Our results illustrated that in the claudin subtype, the immune score of Claudin-low subtype was significantly higher than that of other subtypes ($p < 0.05$). Kaplan-Meier analysis also showed a higher immune score predicted OS in TNBC patients Good ($p < 0.05$). This outcomes was consistent with the results published by Kay Dias that Claudin-low cancer has different clinical pathology and prognostic characteristics from other types of breast cancer. In terms of disease free survival (DFS), Claudin-low cancer has the best 10-year prognosis (72.5%, $p = 0.002$) [27].

In order to explore the potential mechanism of TME changes, we performed GO function enrichment analysis on the screened 289 differentially expressed genes, and found that majorities were related to the TME. Then, a PPI network was constructed assess the interactions between the corresponding TME-related proteins. Multiple critical genes was determined as they had higher number of node connections

in PPI. Interesting, most of the genes are found to be immune-related, indicating that TNBC is highly immunogenic with strong immune characteristics. In the identified genes, *COL1A1* promote the metastasis of breast cancer especially in TNBC cell lines as previous reported. In metastasis, extracellular matrix (ECM) secreted more *COL1A1* than usual to regulate the bioprocess of cells and ECM, subsequently resulting in an invasive and metastatic phenotype [28]. Pre-immune markers *CXCL9*, *CXCL13* were positively correlated with enhanced number of tumor infiltrating lymphocyte (TIL), and were significantly associated with the longer DFS and (pathologic complete response, pCR) [29, 30].

In the study, a survival analysis was performed to explore the potential prognostic value of 289 DEGs and establish a risk model for predicting the prognosis of TNBC. We identified eight TME-related genes (*ACAP1, DUSP1, LYZ, GZMA, SASH3, CCL5, CD74, DPT*). The expression levels of these genes in TNBC patients had significant correlations with poor prognosis, indicating that our huge data analysis via ESTIMATE algorithm on the cBioPortal were greatly successful [31–34]. Among these eight genes, we are particularly interested in *ACAP1, DUSP1*, and *GZMA*. *ACAP1* (ArfGAP With Coiled-Coil, Ankyrin Repeat And PH Domains 1) is a gene encoding protein. Hoffman et al. reported that the expression of the six genes identified by the genes was related to the risk of breast cancer, including the expression of three genes in breast tissue (*RCCD1, DHODH* and *ANKLE1*), and three in whole blood genes (*RCCD1, ACAP1* and *LRRC25*). *ACAP1* played a role in cell proliferation and activation of Arf6 protein [20]. In addition, *ACAP1* might be correlated with *EEC / PI*, and regulated the recycling of integrin β1 during cell migration [7]. Dual-specificity phosphatase-1 (*DUSP1 / MKP1*) was a member of the triple-tyrosine dual-specificity phosphatase family. As a protein phosphatase, *DUSP1* downregulates p38 *MAPK* and *JNKs* signals by directly dephosphorylating threonine and tyrosine. *DUSP1* participated in multiple bioprocess such as cell proliferation, differentiation and apoptosis. It was showed that[33] TNBC had the highest frequency of *DUSP1* methylation if comparing with other breast cancer subtypes. Therefore, *DUSP1* methylation was considered as a unique subtype-specific marker in TNBC patients. Granzyme A, one of the five granzymes encoded in the human genome, was a human protein encoded by *GZMA*, which was closely related to immunity. In a meta-analysis, it was found that at least half of malignant tumors had low or missing *GZMA* protein expression. High levels of *GZMA* and *PRF1* synergistically affected the survival rate of tumors [16].

By using the established prognostic model to predict the 5-year survival rate of TNBC patients, we concluded that the AUC of the ROC curve was 0.737. It meant that the prognostic model had a good survival prediction performance. In the coming future practices, TNBC patients will be divided into high-risk groups and low-risk groups as the mRNA-based risk score prognostic models suggested. Clinicians can determine the therapies based on the predicted outcomes of the model, so as to achieve personalized treatments of TNBC patients. Particularly in high-risk populations, positive strategies should be adopted to prevent TNBC recurrence. Meanwhile, high-risk populations should also be followed up more frequently, and breast MRI scans should be performed routinely to detect the TNBC recurrence earlier. We also demonstrated that the prognostic model was independent of other clinical factors in TNBC. In the GEO dataset (No. GSE103091) to predict the patients' 5-year survival rate, in order to verify its predictive

ability, the AUC value of the ROC curve reaches 0.636. This shows that our model has significance in wide application.

As far as we concerned, these eight biomarkers have not yet been studied in TNBC before. Hence, our findings can provide a solid basic foundation for the development of these new prognostic factors for TNBC in clinical practices, particularly in diagnostic kits exploration. The advantages of the predictive genes we identified is that no further requirement of needed somatic mutation assessment were needed in patients. In addition, our method greatly reduces the cost of sequencing, which makes targeted sequencing applications more cost-effective and routine. Accurate prognosis was essential for appropriate treatment selection. In routine clinical practice, pathological stage classification was common evaluations of prognosis in TNBC patients [35]. However, the clinical outcomes of patients at same stages were usually various with each other due to the known tumor heterogeneity, which indicated that the current staging classification was far from sufficient to a comprehensive prognosis of TNBC [36]. Obviously, the current-used pathological stages in TNBC was entirely based on the anatomical scope of the diseases. This limited property indicated that they were unable to fully represent the biological heterogeneity of TNBC [36]. The tumor heterogeneity, as demonstrated in the previous report, is not only represent the numerous genetic mutations happening in tumor cells, but also dynamic changes of the TME. Those changes of TME mediated by large number of the recruited immune cells somehow determined the occurrence, development and prognosis of the TNBC [37]. As the conventional classifications failed to estimate the tumor heterogeneity, the prognostic model we proposed was expected to improve the prognosis accuracy in TNBC patients.

Honestly, our research also suffered from some limitations. First, the population of the TNBC samples obtained cBioPortal website was mainly limited to white and black people, so it is necessary to expend our study to other nationalities. Secondly, the AUC of the prognostic model we evaluated by GEO dataset (No. GSE103091) was not high enough. Therefore, more verifications in multicenter clinical trials and prospective studies were needed. In the future, we will also explore the possibility of more predictors to improve the predictive performance of our model. Other regression modeling methods will be used to determine whether the prediction accuracy can be improved or not.

In summary, it was clearly demonstrated the eight-gene prognostic model was a considerably reliable tool for predicting the OS of TNBC patients, and can help clinicians selecting personalized treatments for their TNBC patients.

Materials And Methods

Data resources

The metabric breast cancer datasets were download from the open-access cBioPortal data portal (http://www.cbioportal.org/), and TNBC samples was selected and categorized by ER / PR / HER2. The data extracted from the cBioPortal included mRNA expression profile of the TNBC patients and the

corresponding clinical information (age, left/right position, overall survival rate (OS) status, cellularity, claudin subtype, histological subtype, grade, tumor stage, tumor size, lymph node). Our research follows the cBioPortal data access policy and publishing guidelines, so no approval from the local ethics committee is required.

Different distributions of immune scores and stromal scores in clinical characteristics

We extracted the matrix data of the gene expression corresponding to the sample, standardized it with the Limma R package (Version 3.5.2) [21], and then calculated the immune scores and the stromal scores of each sample. The Wilcox rank sum test and the Kruskal–Wallis rank sum test were used to analyze the distribution of immune scores and stromal scores in their clinical characteristics. Then, we analyzed the immune scores, stromal scores, and ESTIMATE scores of those TNBC patients through the survival analysis software. $p < 0.05$ was considered statistically significant.

Screening of differentially expressed genes (DEGs)

The TNBC cases were divided into high/low groups according to the median values of immune and stromal scores, and the difference analysis of the data was performed using the software package limma[21] to obtain the corresponding DEGs (| logFC |> 1, $p < 0.05$).

Enrichment path analysis and construction of PPI network

In our study, we used cluster profiler package to perform Gene Ontology (GO) function enrichment analysis on immune and stromal cell DEGs ($p < 0.05$). Then the two groups of DEGs were combined, and the cluster profiler package was used to perform Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses, which showed the pathways involved. Then, a protein-protein interaction (PPI) network was retrieved from the STRING database [22], reconstructed by Cytoscape software, and a single network with 10 or more nodes was selected for further analysis [23]. We calculated the number of connections of each node in the network and filtered out hub genes.

Establishment of gene-related prognostic models

A single factor Cox regression analysis of the above DEGs was performed using the survival software package to analyze the correlation between TNBC patients and the expression level of each gene. Those genes significantly related to survival were obtained ($p < 0.05$). Next, we used multifactor Cox regression analysis and the stepwise regression to create a most match model to evaluate the contribution of genes as independent prognostic factors for patient survival, calculated the risk value of each patient according to the risk formula, and divided the patients into high-risk and low-risk groups. The Kaplan-Meier (KM) survival curves of high-risk and low-risk cases were drawn, and the properties of the prediction model were evaluated by ROC curve.

Verification of the prognostic model

The breast cancer datasets were downloaded from the GEO database and TNBC samples were selected from it. The above risk scoring formula was used to calculate the risk value of each patient, and the ROC curve was used to evaluate its accuracy of predicting the 5-year survival of the patients, so as to validate the prognostic model. The p value is bilateral, and $p < 0.05$ indicates statistical significance.

Independence analysis of prognostic model and other clinical characteristics

In order to determine whether the predictive performance of the prognostic model of TNBC patients can be independent of other clinical variables (age, left / right position, OS status, cellularity, claudin subtype, histological subtype, grade, tumor stage, tumor size, lymph node), we conducted an univariate and multivariate Cox regression analysis. using other traditional clinical characteristics as independent variables and OS as the dependent variable, hazard ratio (HR) and 95% confidence interval.

Abbreviations

TNBC, triple negative breast cancer; TME, tumor microenvironment; TCGA, the open-assess Cancer Genome Atlas; DEGs: differentially expressed genes; GEO: Gene Expression Omnibus;

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

The National Key R&D Program of China (No. 2017YFC1309100); the National Science Fund for Distinguished Young Scholars (No. 81925023); National Natural Science Foundation of China (No. 81771912); National Natural Science Foundation of China (No. 82071892).

Authors' contributions

Xiaorui Han organizes the research, performs the data analysis and writes the manuscript. Zaiyi Liu revises the manuscript. Changhong Liang supervises the research.

Acknowledgements

Dr. Guoju Hong from Division of Orthopedic Surgery, University of Alberta provided valuable technical advices for the editing of the manuscript.

References

1. Boyle P. Triple-negative breast cancer: epidemiological considerations and recommendations, Ann Oncol 23 Suppl 6 (2012) vi7-12.
2. Lin NU, Vanderplas A, Hughes ME, Theriault RL, Edge SB, Wong YN, Blayney DW, Niland JC, Winer EP, Weeks JC. Clinicopathologic features, patterns of recurrence, and survival among women with triple-negative breast cancer in the National Comprehensive Cancer Network. Cancer. 2012;118(22):5463–72.
3. Trivers KF, Lund MJ, Porter PL, Liff JM, Flagg EW, Coates RJ, Eley JW. The epidemiology of triple-negative breast cancer, including race. Cancer Causes Control. 2009;20(7):1071–82.
4. Ovcaricek T, Frkovic SG, Matos E, Mozina B, Borstnar S. Triple negative breast cancer - prognostic factors and survival. Radiol Oncol. 2011;45(1):46–52.
5. Zhang F, Ren C, Zhao H, Yang L, Su F, Zhou MM, Han J, Sobie EA, Walsh MJ. Identification of novel prognostic indicators for triple-negative breast cancer patients through integrative analysis of cancer genomics data and protein interactome data. Oncotarget. 2016;7(44):71620–34.
6. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10(1):57–63.
7. Anders CK, Carey LA. Biology, metastatic patterns, and treatment of patients with triple-negative breast cancer. Clin Breast Cancer. 2009;9 Suppl 2:73–81.
8. Willis S, De P, Dey N, Long B, Young B, Sparano JA, Wang V, Davidson NE. B.R. Leyland-Jones, Enriched transcription factor signatures in triple negative breast cancer indicates possible targeted therapies with existing drugs. Meta Gene. 2015;4:129–41.
9. Ren B, Cui M, Yang G, Wang H, Feng M, You L, Zhao Y. Tumor microenvironment participates in metastasis of pancreatic cancer. Mol Cancer. 2018;17(1):108.
10. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li Y, Barthel F, Cho HJ, Lin YH, Satani N, Martinez-Ledesma E, Zheng S, Chang E, Gabriel Sauve CE, Olar A, Lan ZD, Finocchiaro G, Phillips JJ, Berger MS, Gabrusiewicz KR, Wang G, Eskilsson E, Hu J, Mikkelsen T, DePinho RA, Muller F, Heimberger AB, Sulman EP, Nam DH, Verhaak RGW. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell. 2018;33(1):152.

11. Xu M, Li Y, Li W, Zhao Q, Zhang Q, Le K, Huang Z, Yi P. Immune and Stroma Related Genes in Breast Cancer: A Comprehensive Analysis of Tumor Microenvironment Based on the Cancer Genome Atlas (TCGA) Database. Front Med (Lausanne). 2020;7:64.
12. Roma-Rodrigues C, Mendes R, Baptista PV, Fernandes AR. Targeting Tumor Microenvironment for Cancer Therapy, Int J Mol Sci 20(4) (2019).
13. Xiong Y, Wang K, Zhou H, Peng L, You W, Fu Z. Profiles of immune infiltration in colorectal cancer and their clinical significant: A gene expression-based study. Cancer Med. 2018;7(9):4496–508.
14. Mahmoud SM, Paish EC, Powe DG, Macmillan RD, Grainge MJ, Lee AH, Ellis IO, Green AR. Tumor-infiltrating CD8 + lymphocytes predict clinical outcome in breast cancer. J Clin Oncol. 2011;29(15):1949–55.
15. Castaneda CA, Mittendorf E, Casavilca S, Wu Y, Castillo M, Arboleda P, Nunez T, Guerra H, Barrionuevo C, Dolores-Cerna K, Belmar-Lopez C, Abugattas J, Calderon G, De La Cruz M, Cotrina M, Dunstan J, Gomez HL, Vidaurre T. Tumor infiltrating lymphocytes in triple negative breast cancer receiving neoadjuvant chemotherapy. World J Clin Oncol. 2016;7(5):387–94.
16. Bussard KM, Mutkus L, Stumpf K, Gomez-Manzano C, Marini FC. Tumor-associated stromal cells as key contributors to the tumor microenvironment. Breast Cancer Res. 2016;18(1):84.
17. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4:2612.
18. Shah N, Wang P, Wongvipat J, Karthaus WR, Abida W, Armenia J, Rockowitz S, Drier Y, Bernstein BE, Long HW, Freedman ML, Arora VK, Zheng D, Sawyers CL. Regulation of the glucocorticoid receptor via a BET-dependent enhancer drives antiandrogen resistance in prostate cancer, Elife 6 (2017).
19. Priedigkeit N, Watters RJ, Lucas PC, Basudan A, Bhargava R, Horne W, Kolls JK, Fang Z, Rosenzweig MQ, Brufsky AM, Weiss KR, Oesterreich S, Lee AV. Exome-capture RNA sequencing of decade-old breast cancers and matched decalcified bone metastases, JCI Insight 2(17) (2017).
20. Alonso MH, Ausso S, Lopez-Doriga A, Cordero D, Guino E, Sole X, Barenys M, de Oca J, Capella G, Salazar R, Sanz-Pamplona R, Moreno V. Comprehensive analysis of copy number aberrations in microsatellite stable colon cancer in view of stromal component. Br J Cancer. 2017;117(3):421–31.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W. G.K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
22. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ. C. von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43:D447-52. (Database issue).
23. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. cytoHubba: identifying hub objects and sub-networks from complex interactome. BMC Syst Biol. 2014;8 Suppl 4:11.
24. Metzger-Filho O, Tutt A, de Azambuja E, Saini KS, Viale G, Loi S, Bradbury I, Bliss JM, Azim HA Jr, Ellis P, Di Leo A, Baselga J, Sotiriou C. M. Piccart-Gebhart, Dissecting the heterogeneity of triple-negative

breast cancer. J Clin Oncol. 2012;30(15):1879–87.

25. Mao Y, Keller ET, Garfield DH, Shen K, Wang J. Stromal cells in tumor microenvironment and breast cancer. Cancer Metastasis Rev. 2013;32(1–2):303–15.
26. Merlano MC, Abbona A, Denaro N, Garrone O. Knowing the tumour microenvironment to optimise immunotherapy. Acta Otorhinolaryngol Ital. 2019;39(1):2–8.
27. Dias K, Dvorkin-Gheva A, Hallett RM, Wu Y, Hassell J, Pond GR, Levine M, Whelan T, Bane AL, Claudin-Low Breast Cancer; Clinical & Pathological Characteristics, PLoS One 12(1) (2017) e0168669.
28. Liu J, Shen JX, Wu HT, Li XL, Wen XF, Du CW, Zhang GJ. Collagen 1A1 (COL1A1) promotes metastasis of breast cancer and is a potential therapeutic target. Discov Med. 2018;25(139):211–23.
29. Denkert C, Loibl S, Noske A, Roller M, Muller BM, Komor M, Budczies J, Darb-Esfahani S, Kronenwett R, Hanusch C, von Torne C, Weichert W, Engels K, Solbach C, Schrader I, Dietel M, von Minckwitz G. Tumor-associated lymphocytes as an independent predictor of response to neoadjuvant chemotherapy in breast cancer, J Clin Oncol 28(1) (2010) 105 – 13.
30. Lee HJ, Lee JJ, Song IH, Park IA, Kang J, Yu JH, Ahn JH, Gong G. Prognostic and predictive value of NanoString-based immune-related gene signatures in a neoadjuvant setting of triple-negative breast cancer: relationship to tumor-infiltrating lymphocytes. Breast Cancer Res Treat. 2015;151(3):619–27.
31. Hoffman JD, Graff RE, Emami NC, Tai CG, Passarelli MN, Hu D, Huntsman S, Hadley D, Leong L, Majumdar A, Zaitlen N, Ziv E, Witte JS. Cis-eQTL-based trans-ethnic meta-analysis reveals novel genes associated with breast cancer risk. PLoS Genet. 2017;13(3):e1006690.
32. Johnstone CN, Pattison AD, Gorringe KL, Harrison PF, Powell DR, Lock P, Baloyan D, Ernst M, Stewart AG, Beilharz TH, Anderson RL. Functional and genomic characterisation of a xenograft model system for the study of metastasis in triple-negative breast cancer, Dis Model Mech 11(5) (2018).
33. Li J, Chen Y, Yu H, Tian J, Yuan F, Fan J, Liu Y, Zhu L, Wang F, Zhao Y, Pang D. DUSP1 promoter methylation in peripheral blood leukocyte is associated with triple-negative breast cancer risk. Sci Rep. 2017;7:43011.
34. Tuglu MM, Bostanabad SY, Ozyon G, Dalkilic B, Gurdal H. The role of dualspecificity phosphatase 1 and protein phosphatase 1 in beta2adrenergic receptormediated inhibition of extracellular signal regulated kinase 1/2 in triple negative breast cancer cell lines. Mol Med Rep. 2018;17(1):2033–43.
35. Urru SAM, Gallus S, Bosetti C, Moi T, Medda R, Sollai E, Murgia A, Sanges F, Pira G, Manca A, Palmas D, Floris M, Asunis AM, Atzori F, Carru C, D'Incalci M, Ghiani M, Marras V, Onnis D, Santona MC, Sarobba G, Valle E, Canu L, Cossu S, Bulfone A, Rocca PC, De Miglio MR, Orru S. Clinical and pathological factors influencing survival in a large cohort of triple-negative breast cancer patients. BMC Cancer. 2018;18(1):56.
36. Lehmann BD, Pietenpol JA. Identification and use of biomarkers in treatment strategies for triple-negative breast cancer subtypes. J Pathol. 2014;232(2):142–50.
37. Runa F, Hamalian S, Meade K, Shisgal P, Gray PC, Kelber JA. Tumor microenvironment heterogeneity: challenges and opportunities. Curr Mol Biol Rep. 2017;3(4):218–29.

Figures

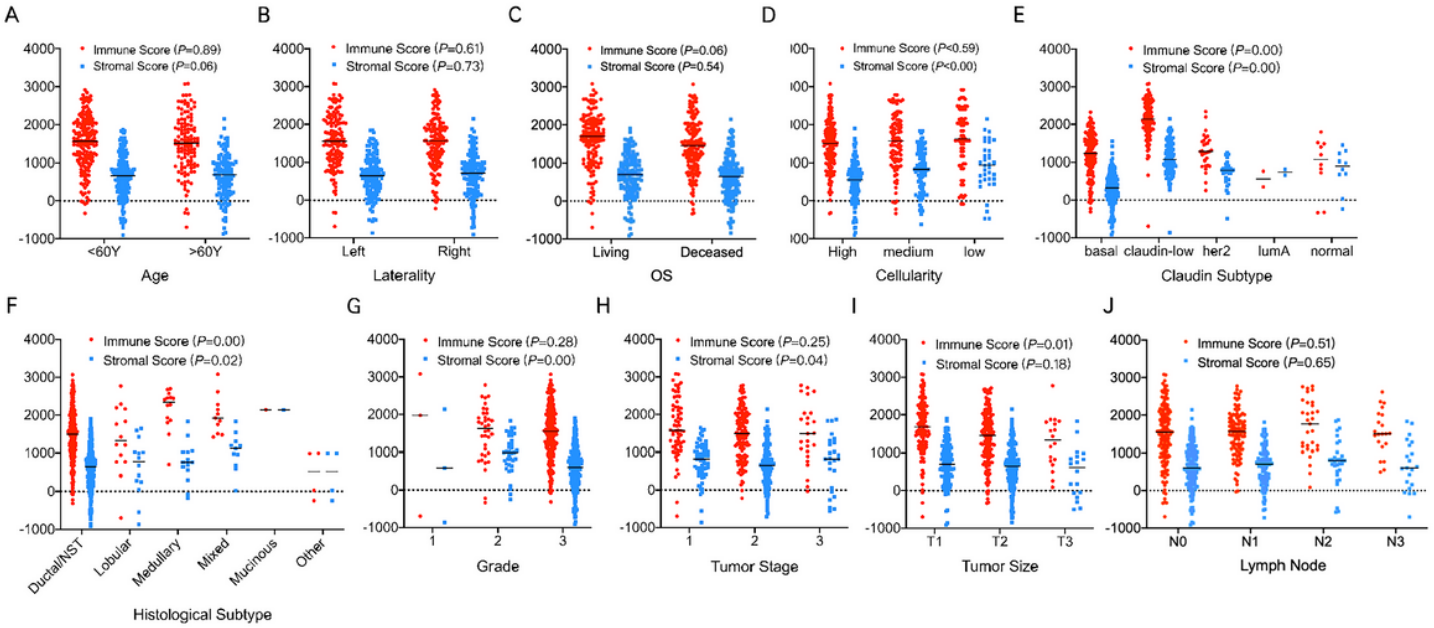

Figure 1

The scatter plot diagram demonstrated that the distribution of immune score and stromal score by clinical characteristics of the enrolled subjects including age (A), laterality (B), overall survival (OS) status (C), cellularity (D), claudin subtype (E), histological subtype (F), grade (G), tumor stage (H), tumor size (I), Lymph node (J).

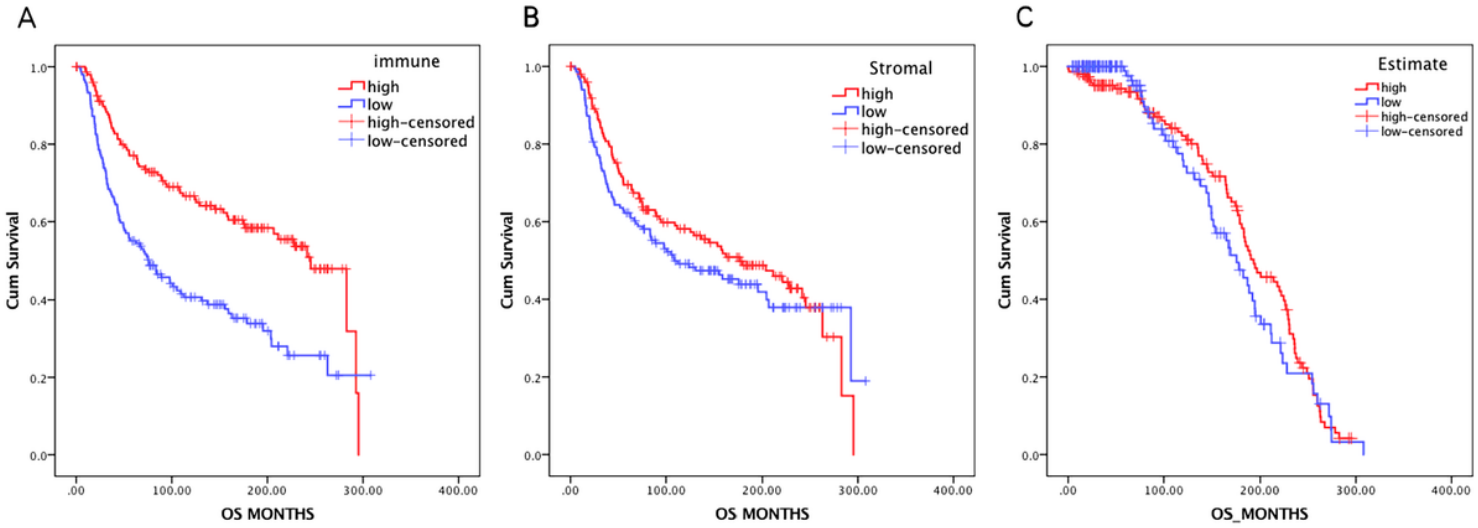

Figure 2

The overall survival curve obtained by the Kaplan-Meier method illustrated the prognosis of overall survival (OS) among high/low immune score (A) and stromal score (B), as well as the estimate score (C).

The horizontal and vertical axis indicated survival times and survival rates respectively. The red and blue curves represent samples with immune scores above and below the median value, respectively.

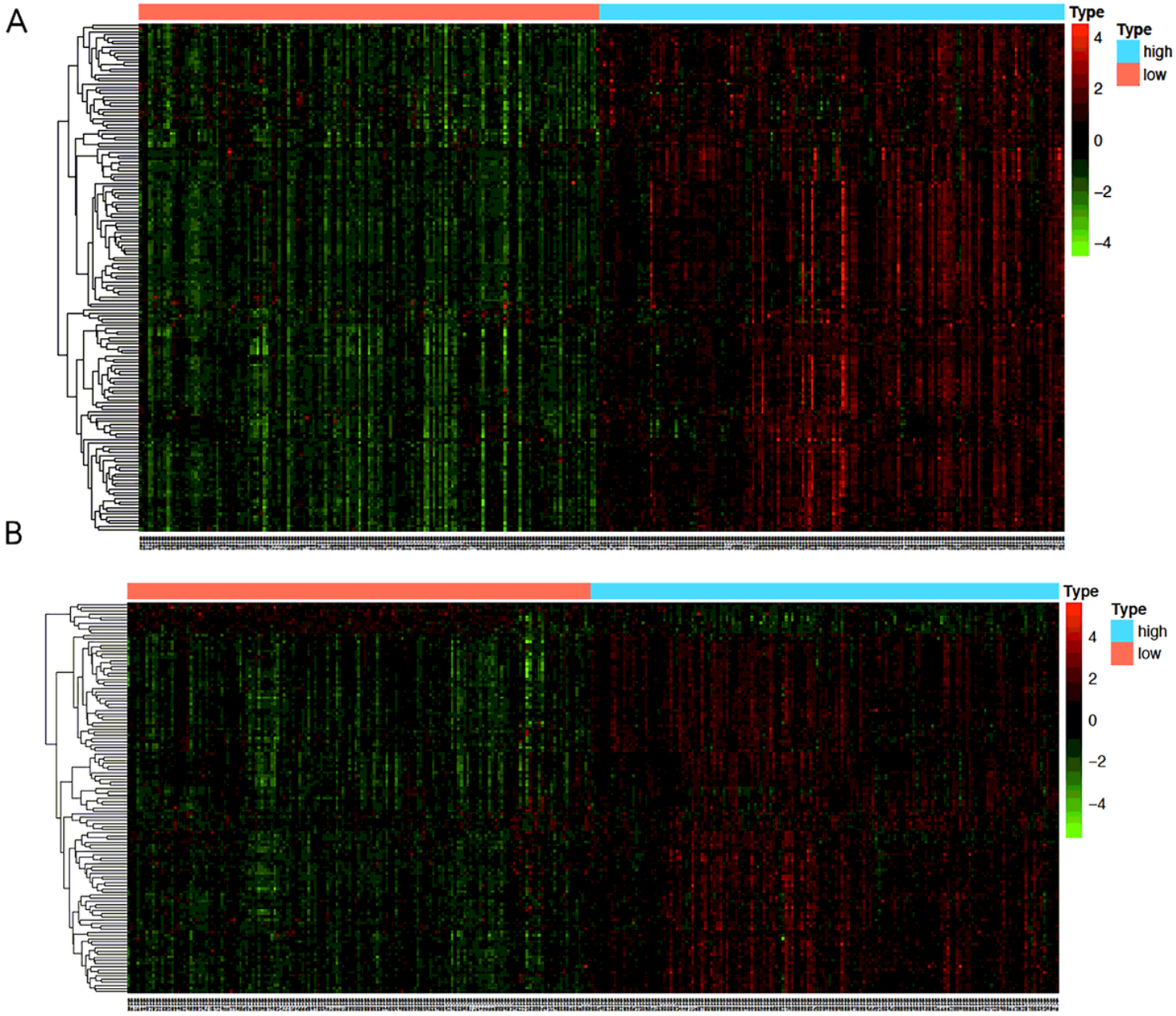

Figure 3

Gene expression profile of high/low immune and stromal score in triple negative breast cancer (TNBC). (A) Heat map of immune-related differentially expressed genes (DEG)s. (B) Heat map of stromal-related DEGs. The horizontal and vertical axis represent TNBC samples and genes expression level respectively. Genes with higher, lower or same expression levels are shown as red, green, and black correspondingly. The blue and pink strips located above heat map diagrams indicate samples with high/low scores.

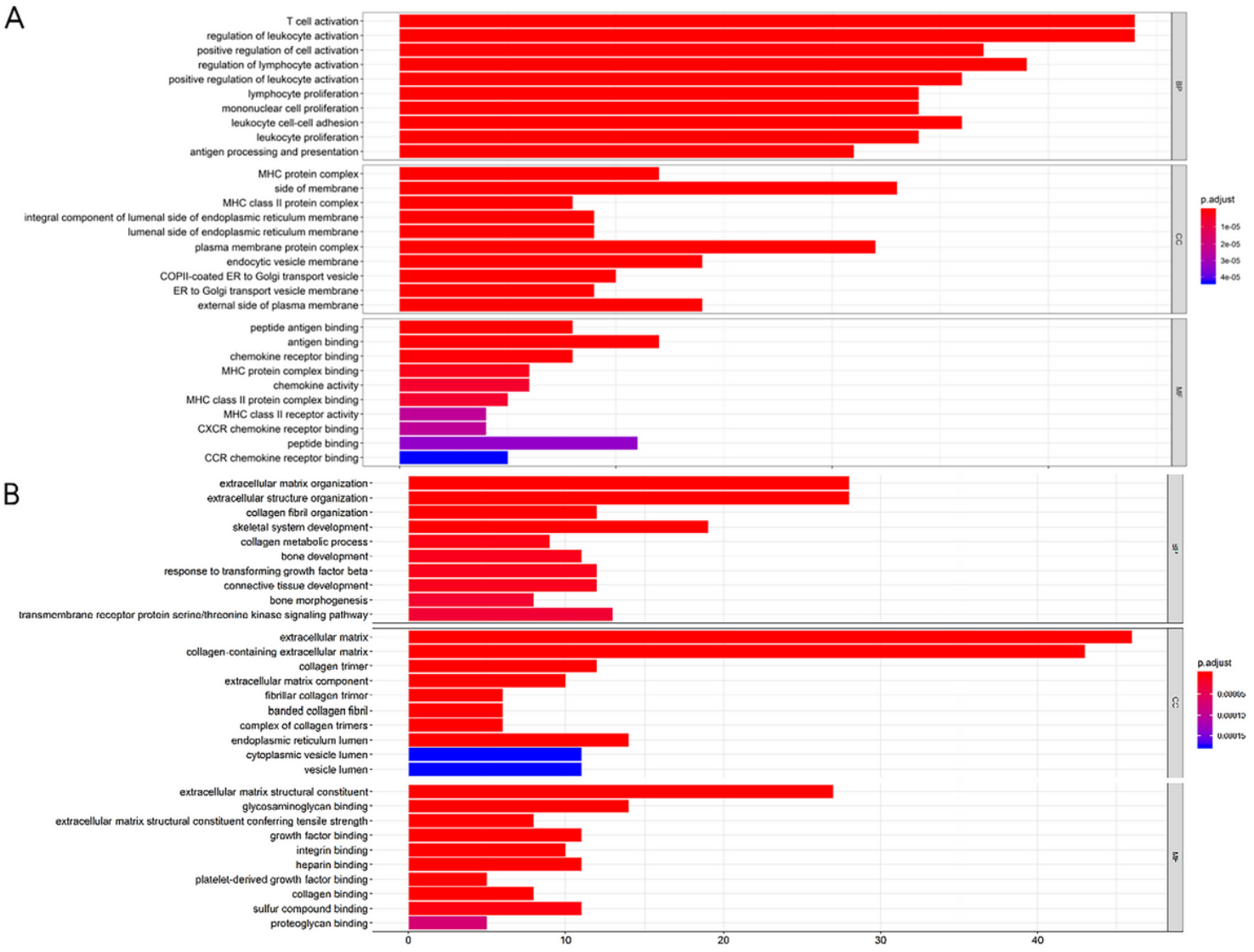

Figure 4

Gene Ontology (GO) term enrichment analysis. The bar plot diagram show the top ten items of Biological Process(BP)/Cellular Component(CC)/Molecular Function(MF) in immune score group (A) and stromal score group (B).

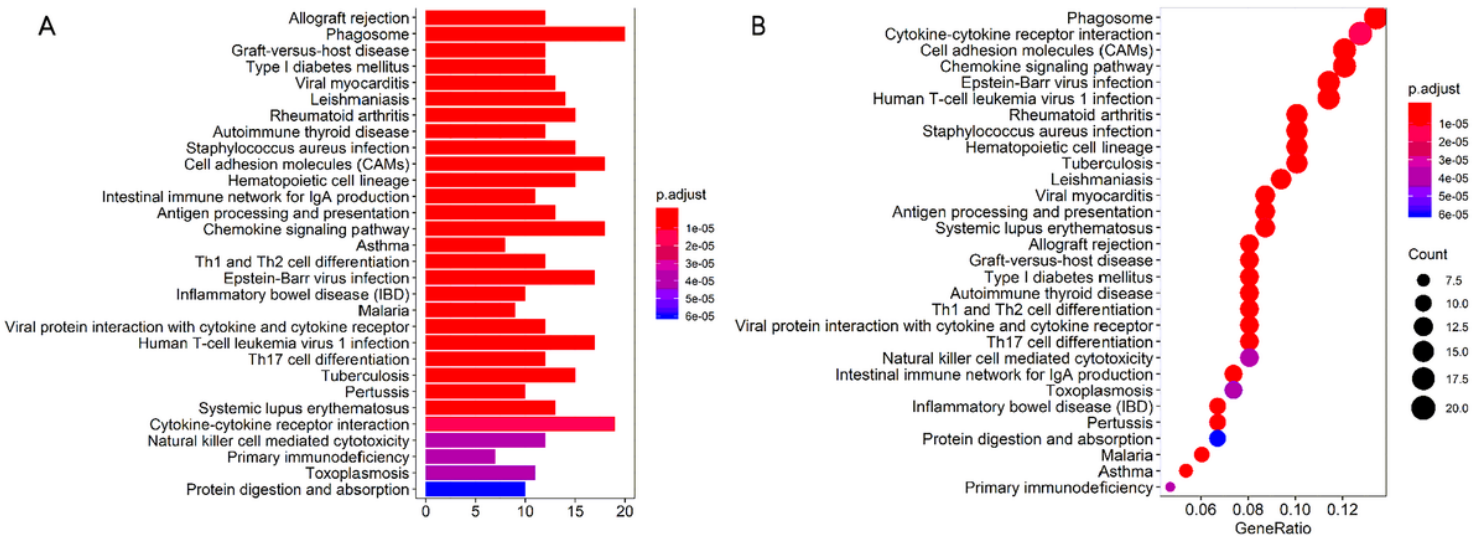

Figure 5

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis results. Barplot(A) and dotplot(B) respectively show the main participating pathways (top30, p<0.05).

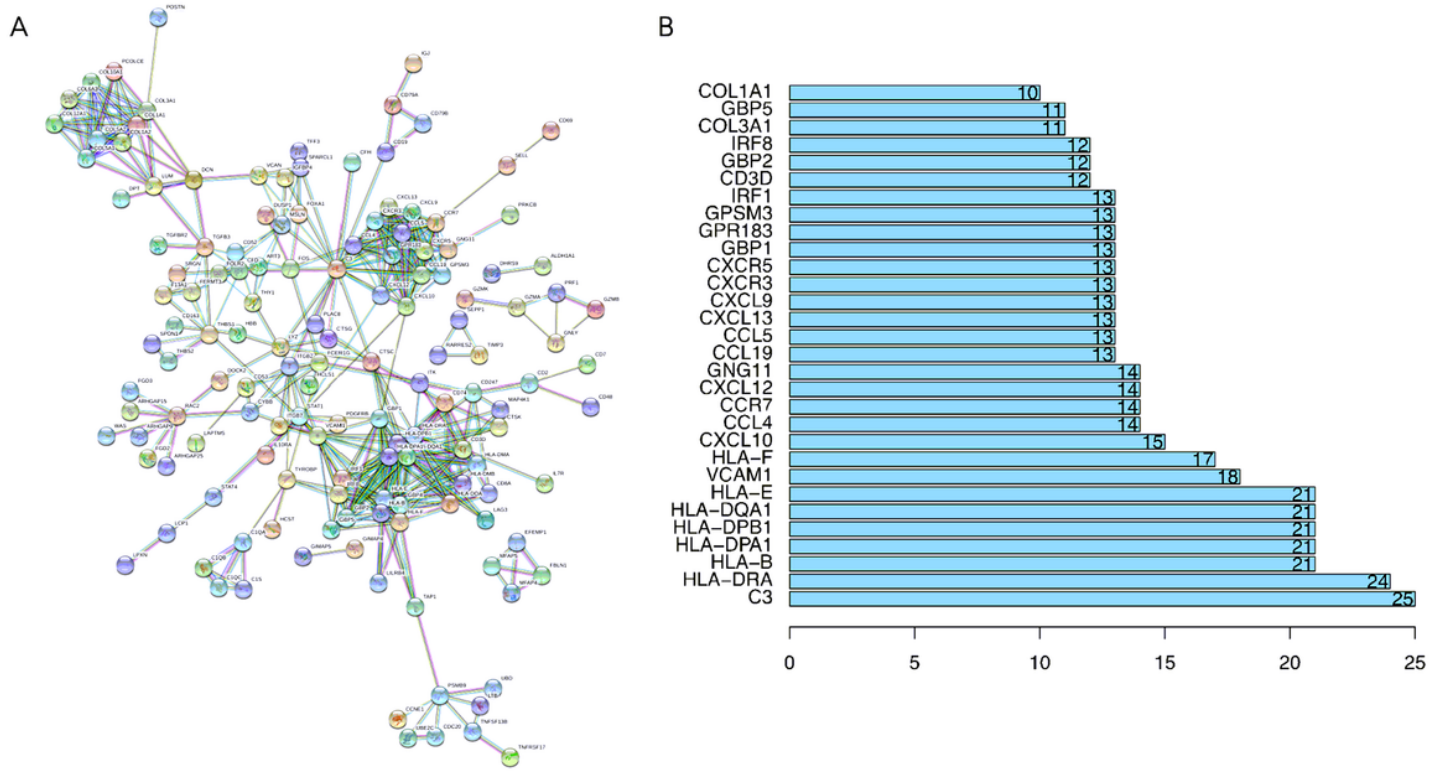

Figure 6

Protein-Protein Interaction (PPI) network in TNBC. Bioinformatics prediction of essential proteins in databases are showed in Module (A) and Hub genes (B).

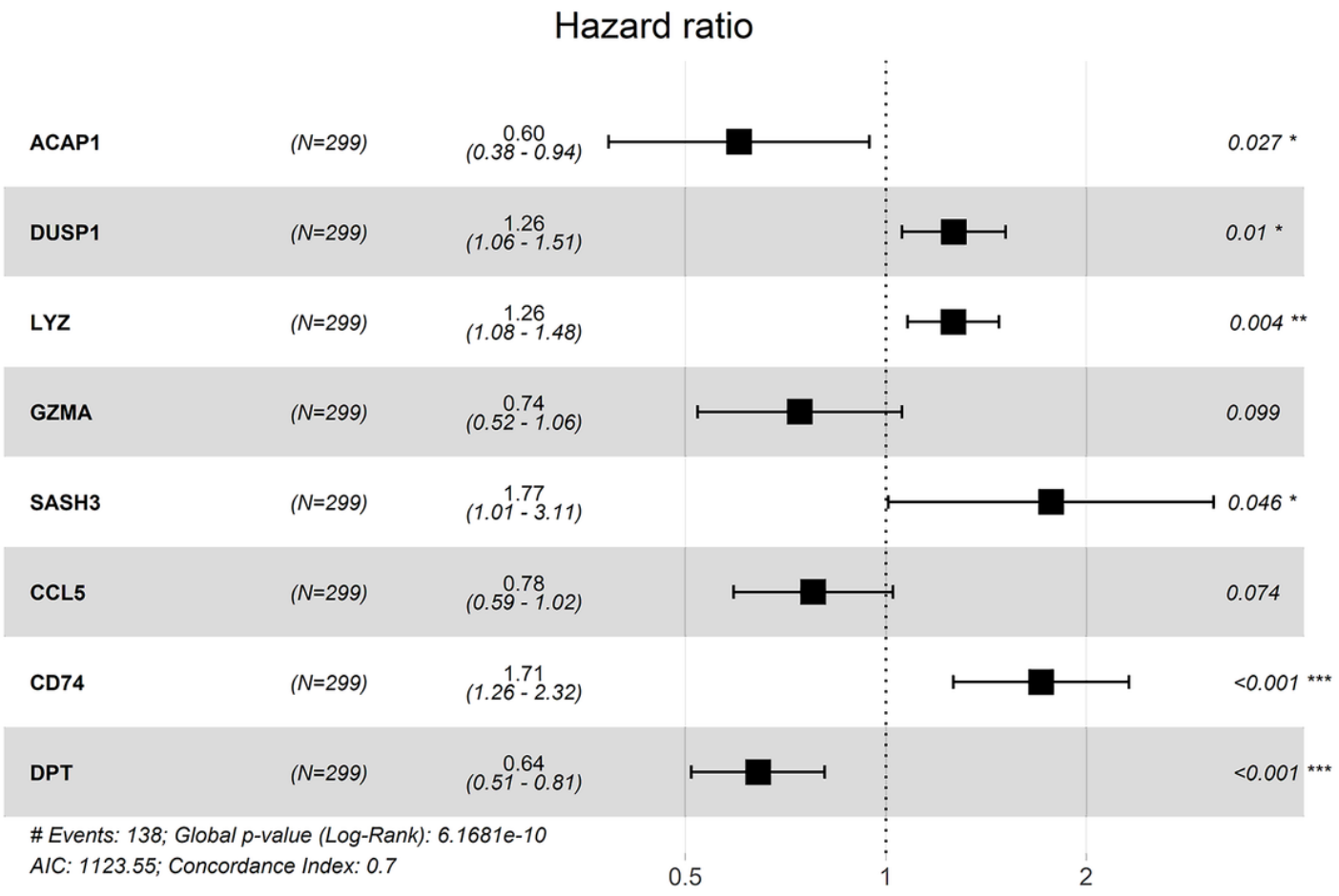

Figure 7

Cox regression indicates the hazard ratio (HR) of OS in the cBioPortal data portal. The confidence interval is shown as the length of the line. *p < 0.05; ***p < 0.01; ***p < 0.001.

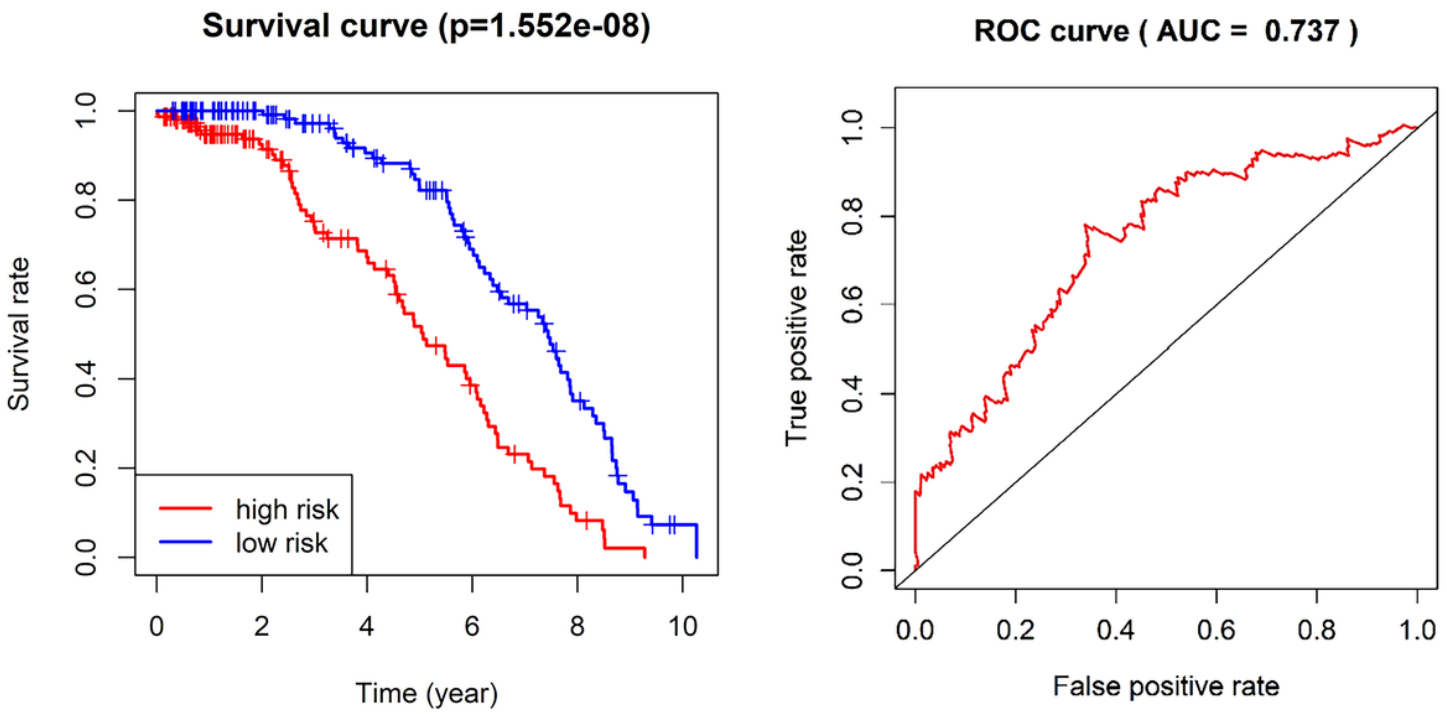

Figure 8

Kaplan-Meier survival analysis. Overall survival curves indicate that the risk score is significantly associated with triple negative breast cancer (TNBC) prognosis. Horizontal and vertical axes represent survival times and rates, respectively. Red and blue curves are samples with risk score higher and lower than the median value, respectively. P<0.05 indicates that the difference is statistically significant. (B) Time-dependent receiver operating characteristic (ROC) curves for the prognostic model. The area under the curve (AUC) values for the risk model for 5-year survival was 0.737.

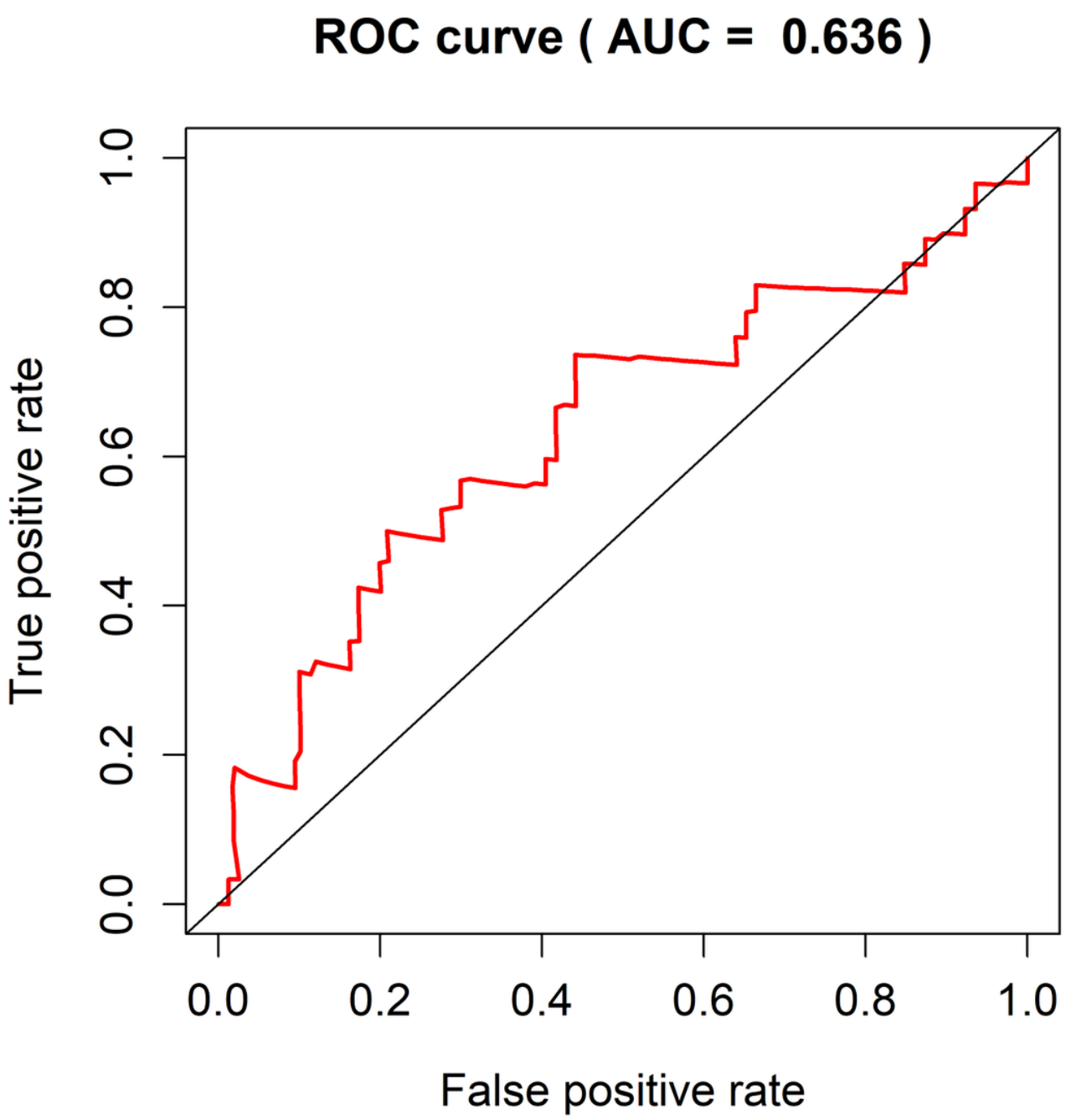

Figure 9

Model validation in GEO data set. Use the above risk scoring formula to calculate the risk value of each patient and predict the five-year survival of the patient. The area under the curve (AUC) value is 0.636.

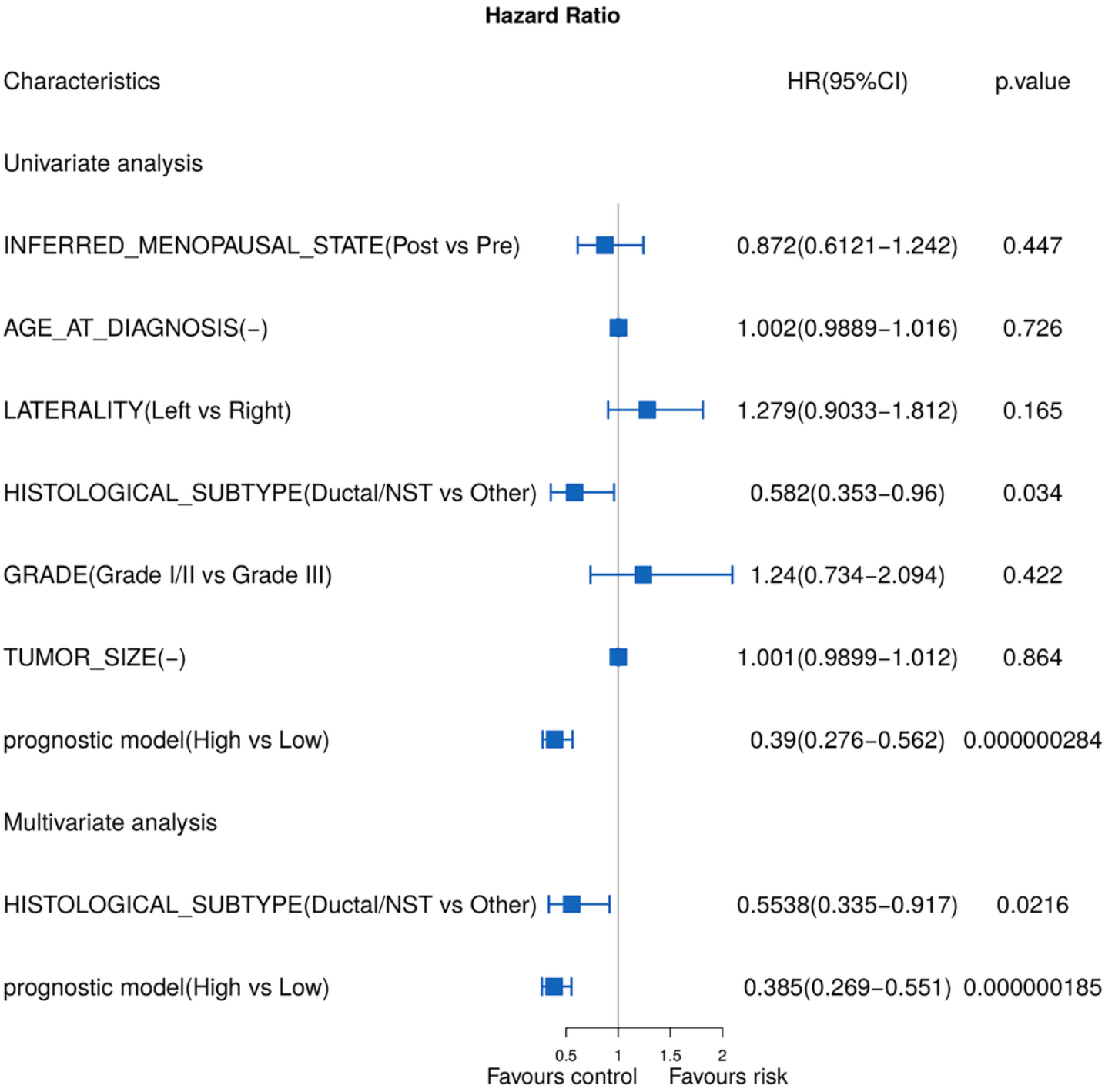

Figure 10

Univariate and multivariate association of the prognostic model and clinical characteristics with overall survival (OS). The confidence interval is shown as the length of the line.