

Role of hypoxia-related genes in breast cancer based on a comprehensive analysis of scRNA-seq and bulk RNA-seq

Rui Sun

The Third Affiliated Hospital of Soochow University

Wenming Yin

The Third Affiliated Hospital of Soochow University

Yilin Zha

The Third Affiliated Hospital of Soochow University

Dan Xi

The Third Affiliated Hospital of Soochow University

Yingjie Shao

The Third Affiliated Hospital of Soochow University

Wendong Gu

The Third Affiliated Hospital of Soochow University

Jingting Jiang (✉ jiangjingting@suda.edu.cn)

The Third Affiliated Hospital of Soochow University

Research Article

Keywords: Single-cell RNA-seq, Bulk RNA-seq, Breast cancer, Hypoxia, Risk score

Posted Date: May 5th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1558418/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background:

The hypoxic state in tumor microenvironment of breast cancer favors the proliferation and metastasis of tumor cells thus affecting patient survival. In this study, we aimed to combine single-cell sequencing data and bulk sequencing data to construct hypoxia-related prognostic signature of breast cancer patients.

Methods:

Single-cell RNA transcriptome data of MCF7 cells subjected to hypoxia microenvironment, the bulk tumour transcriptome data and clinical data were loaded from Gene Expression Omnibus (GEO) database and The Cancer Genome Atlas (TCGA) database. Screening for differentially expressed genes (DEGs) in MCF7 cells under hypoxic microenvironment. Functional enrichment analysis of these DGEs was performed via Gene ontology annotation (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Set Variation Analysis (GSVA). Univariate, the least absolute shrinkage and selection operator (LASSO) regression algorithm and multivariate Cox regression were applied to determine the prognostic gene signature. Risk score and clinical features were used to construct a nomogram model. 22 immune cell infiltrates were calculated using the CIBERSORT algorithm and correlations between risk scores and tumor microenvironment were analyzed.

Results:

Total 329 DEGs between the hypoxia and normoxia state were identified from scRNA-seq data. GO, KEGG and GSVA analysis revealed that the most gene sets were related to hypoxia. Combined with information of 1099 breast cancer samples sourced from the TCGA database, we identified four genes (ERRFI1, HSPB8, PGK1, STC2) to be independent prognostic genes, and risk scores based on their gene expression were calculated for each patient. Kaplan-Meier survival analysis showed a negative correlation between risk score and patient survival time. The validation using GSE20685 dataset data also yielded similar results. Risk scores and clinical data were considered as independent prognostic factors for breast cancer patients for the construction of a nomogram model, and the model showed good prognostic power. CIBERSORT algorithms was used to calculate 22 immune cell infiltrates in BRCA patients, and the analysis demonstrated that risk scores were positively associated with immunosuppression.

Conclusions:

In summary, we used single-cell sequencing data from MCF-7 cells under hypoxic conditions to identify prognosis-related genes, and the constructed prognostic model displayed well predictive properties.

Background

The tumor microenvironment (TME) was generated by the interaction between tumor cells and the organism[1]. Most rapidly growing solid tumors have a hypoxic microenvironment, which is characterized

by lower-than-normal oxygen levels and pressures, and is particularly common in rapidly growing solid tumors. In the hypoxic microenvironment, tumor cells rely on glucose metabolism, enhance resistance to apoptosis, induce genomic instability, reduce immune surveillance, and induce angiogenesis and migration to hypoxic regions[2]. There has been an extensive review on the clinical significance of hypoxia in cancer treatment, which not only reduces the sensitivity to radiotherapy involved in chemoresistance but also can compromise immunotherapy efficacy[3–6].

Breast cancer is a malignant tumor common in women on a global scale, and approximately 276,480 women in the US are estimated to develop breast cancer in 2020[7]. Despite the development of comprehensive breast cancer treatment in recent years, it is still the second deadliest tumor in women[8]. Local recurrence and metastasis are the most important reasons for reducing breast cancer survival time[9]. Hypoxia-inducible factor 1 (HIF-1) is widely recognized as a major regulator of the body's response to hypoxia and involved in angiogenesis, invasion, metastasis and treatment resistance in breast cancer[10]. The exact mechanism by which hypoxia causes breast cancer recurrence and metastasis have been investigated comprehensively[11, 12]. Genomic analysis of bulk tumors had shed light on the biological pathways of tumor cells in the hypoxic microenvironment. Recent developments in single-cell analysis methods can reveal more precisely the extent, function and origin of differences between cells[13]. Such as Ana Miar et al. used single-cell sequencing of MCF7 cells to determine the hypoxic impacts on type I IFN pathway and the underlying mechanisms involved[14], however, the result was not consistent with the previously accepted view that the downregulation of IFN during hypoxia is mainly associated with HIF1 α and HIF2 α .

Herein, we used scRNA-seq data of MCF7 cells under hypoxia environment, wishing to provide a more valid indicator of the role between hypoxia and breast cancer progression.

Methods

1.Data collection

Single-cell transcriptome data of MCF7 cells subjected to 0.1% hypoxia or normoxia microenvironment and the data of 326 breast cancer samples were available from the GEO database under numbers GSE134038 and GSE20685, respectively. The bulk tumour transcriptome data and clinical data of BRCA patient was downloaded from TCGA database, total 1222 samples were collected. The clinical information included age, histological type, T stage, N stage, M stage, follow-up time, survival status.

2. scRNA-Seq Data Processing and screening differentially expressed genes under hypoxia microenvironment

The scRNA-seq data was analyzed by 'Seurat' package under R environment[15]. Genes detected in less than 3 cells, detected cells with less than 50 genes and the percentage of mitochondrial-derived gene

expression exceeds 5% were used as the data filtering criteria. Genes that differed significantly between cells were identified, and the top 1500 variable genes was plotted. Principal component analysis (PCA) was taken for linearly dimension reduction and identify significant available dimensions in the dataset[16]. The t-Distributed Stochastic Neighbor Embedding (tSNE) analysis was to visualize the cluster classification of all cells[17]. FindMarkers was utilized to screen the differentially expressed genes (DEGs) among the hypoxia and normoxia clusters, which were defined as hypoxia related DEGs. The cutoff value of differentially expressed genes was defined as the absolute log2 fold change of ≥ 1 and the adjusted P value of < 0.05 .

3. Functional analysis of hypoxia related DEGs

Gene ontology annotation (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) were done for the screened hypoxia related DEGs to clarify GO and KEGG function classifications annotation and functional enrichment. The false discovery rate (FDR) < 0.05 was used as a filtering criterion. These analyses were made in R environment with 'clusterProfiler' package[18] and then visualized with the 'ggplot2' package. Gene set variation analysis (GSVA) was used to analyze the differences in gene enrichment between the hypoxia and normoxia clusters, and was performed in R environment with 'GSVA' package. Gene sets of "h.all.v7.4.symbols" were downloaded from MSigDB database for GSVA analysis. p value < 0.01 was considered as statistically significance.

4. Construction and validation of hypoxia related signature in BRCA patients

First, a matrix of hypoxia-related DEGs in tumor patients were extracted from TCGA-BRCA transcriptome profiles. Prognostic genes that correlate with overall survival (OS) were then screened by univariate analysis. Those genes with prognostic significance were put into the least absolute shrinkage and selection operator (LASSO) regression algorithm to screen the candidate prognostic genes. Then, the result of multivariate Cox regression analysis was finally used to construct hypoxia related risk score model. Univariate and multivariable Cox regression analysis were conducted by the 'survival' package in R environment. Risk score was figured out as followed: Risk score = $\sum_{i=1}^n \text{Coef}_i \times x_i$, Coef is the coefficient analyzed in multivariable Cox regression analysis, x is the gene number of selected hypoxia-associated gene. Patients in the TCGA-BRCA database were the training cohort and patients in the GSE20685 database were the validation cohort. Use this equation to figure out the risk score, and classify patient into two groups of high and low risk based on the median value of their results. The expression levels of prognostic genes in the two groups were demonstrated using heatmaps, which were created in the R environment by the 'heatmap' package. The Kaplan-Meier method was used to evaluate the survival analysis for the two groups, which was created using the 'survival' and 'survminer' packages in the R environment. We evaluate the efficiency of our risk score for OS prediction by manipulating receiver

operating characteristic (ROC) curves, which are constructed with "survivalROC" package in R environment.

5. Construction of Nomogram based on hypoxia related signature and clinical characteristics

Univariate and multivariate Cox regression analysis were done to investigate the correlation between risk score, clinical characteristics and OS in BRCA patients, and then construct the nomograms by the 'rms' and 'survival' package, which is used to estimate the probability of 3-, 5- and 10-years survival for BRCA patients. The calibration curve was obtained by plotting the observed rates against the probability predicted by nomogram via a Bootstrap method with 1,000 resamples for internal validation of the analyzed database.

6. Analysis of the relevance of risk scores and immune cell infiltration in TCGA-BRCA patients

The 'CIBERSORT' R package was used to calculate 22 immune cell infiltrates in TCGA-BRCA patients. Then, we calculate the correlation between risk score and immune infiltration using Pearson correlation.

7. Statistical analysis

All of the statistical tests including univariate and multivariate Cox regression, LASSO regression analysis, K-M survival analyses, Pearson correlation analysis were implemented by R version 4.0.2 in the RStudio environment. All analyses performed were two-sided, and P values < 0.05 considered statistically significant.

Results

1. Identification of DEGs of MCF7 cells among 0.1% hypoxia or normoxia microenvironment using scRNA-seq data

Quality control of scRNA data of MCF7 cells subjected to 0.1% hypoxia or normoxia microenvironment was shown in Fig. 1A, plotting the range of single-cell RNA amounts, total gene numbers, and the proportion of mitochondria sequenced per cell. Scatterplot of mitochondrial gene proportions in relation to sequencing data and scatterplot of genes in relation to sequencing data was shown in Fig. 1B. 13186 genes and 1,500 highly variable genes found in the scRNA data between cells were list in Fig. 1C, and the top 10 variable genes were labeled. And these variable genes were divided into 20 different components by PCA, of which statistically significant components were further used for tSNE dimensionality reduction

(Fig. 2A-D). After tSNE dimensionality reduction, MCF7 cells were divided into 4 clusters. According to the experimental design, cluster 0 and cluster 3 from 0.1% hypoxic microenvironment, cluster 1 and cluster 2 from normoxic microenvironment (Fig. 3A-B). Subsequently, 329 DEGs between hypoxia and normoxia clusters were identified, the violin plot shows the top 10 up-regulated and top 10 down-regulated differentially expressed genes between these two clusters. (Fig. 3C).

2. Functional analysis of hypoxia related DEGs

GO function and KEGG pathway analysis was performed on 329-hypoxia associated DEGs. In the biological processes, the hypoxia related DEGs were mainly enriched in response to oxygen levels, response to hypoxia, response to decreased oxygen levels et al. In the cellular component, chromosomal region, chromosome, telomeric region; nuclear chromosome, telomeric region et al. In the molecular functions, the hypoxia related DEGs were mainly enriched in ATPase activity, DNA-binding transcription factor binding, nuclear receptor binding et al. (Fig. 4A). In the KEGG pathways, the hypoxia related DEGs were mainly enriched in Pathways of neurodegeneration-multiple diseases, Parkinson disease, Alzheimer disease et al. (Fig. 4B). We used GSVA enrichment analysis to investigate the differences in biological behavior between hypoxic and normoxic clusters. As shown in Fig. 4C, in the hypoxic state, such as hypoxia, glycolytic pathways are activated, while oxidative phosphorylation and DNA repair pathways are inhibited, and this analysis is also consistent with experimental studies.

3. Construction and validation of hypoxia related signature in breast cancer patients

Total 1099 BRCA tumor samples downloaded from TCGA database were included in the analysis by removing the tissue information of 113 normal samples. A matrix of 329 depleted hypoxia-related DEGs was extracted for subsequent analysis. The results of univariate analysis indicated that 22 prognostic genes were statistically correlated with OS in BRCA patients, of which, 16 genes (BAMBI, BNIP3, CBX5, GRB10, HSPB8, KRT80, LSS, NDRG1, P4HA2, PGK1, SLC7A5, SRD5A3, ST3GAL1, TUBA1B, TXNRD1, XPOT) were considered as dangerous genes, 6 genes (CYBA, DUS1L, ERFFI1, FAU, PSME2, STC2) were thought as protective genes (Fig. 5A). LASSO regression analysis was done for the selection of the ideal genome, and 14 candidate prognostic genes were screened, including BAMBI, CBX5, CYBA, DUS1L, ERFFI1, GRB10, HSPB8, KRT80, LSS, P4HA2, PGK1, PSME2, SRD5A3, STC2 (Fig. 5B-C). Finally, according the result of multivariate Cox regression analysis, 4 genes (ERFFI1, HSPB8, PGK1, STC2) was thought as independent prognostic factors in BRCA patients, and prognostic models were constructed (Fig. 5D). The risk score was obtained below: Risk score = $(-0.3240 \times \text{ERFFI1}) + (0.1613 \times \text{HSPB8}) + (0.5816 \times \text{PGK1}) + (-0.1245 \times \text{STC2})$ (Fig. 4). Risk scores were used to classify TCGA-BRCA patients into high-risk and low-risk groups. Kaplan-Meier survival analysis showed a negative correlation between risk score and patient survival time (Fig. 6A). As the risk score increased, there were significantly more deaths in BRCA patients (Fig. 7A-C). And the area under the curves (AUCs) in ROC curves was 0.747, demonstrates positive

predictive accuracy of this prognostic model (Fig. 6B). We also validated the predictive power of this prognostic feature with the GSE20685 dataset containing 327 breast cancer samples. As well, the number of people in the high-risk group who were in the death state was significantly higher (Fig. 7D-F). The accuracy of this model was confirmed in validation set with an AUC of 0.674 in the ROC curve (Fig. 6C-D).

4. Construction of Nomogram based on risk score and clinical characteristics

Samples were excluded for lack of any clinical features, at last, total of 705 cases of clinical information data were included. Clinical characteristic of 705 BRCA tumor samples in the TCGA database were presented by Table 1. The clinicopathological factors such as age, histological type (Infiltrating Ductal Carcinoma, Infiltrating Lobular Carcinoma and other), T stage (T0, T1, T2 and T3), N stage (N0, N1, N2, and N3), M stage (M0 and M1) were included. The results of univariate Cox regression analysis indicated that age, T stage, N stage, and M stage were correlated with OS in BRCA patients (Fig. 8A). After multivariate Cox regression analysis, it was determined that risk score, age, N stage, and M stage in clinical characteristics, were significantly associated with OS in BRCA patients (Fig. 8B). Depending on the results of the multifactor Cox regression analysis, age, T stage, N stage, M stage and risk score were integrated into the OS competing nomograms of BRCA patients (Fig. 9A). The calibration plot for the probability of OS at 3-, 5-, and 10-years showed promising line of agreement between the prediction using nomograms and actual observed survival (Fig. 9B-D).

Table 1
Clinical characteristic of breast invasive carcinoma (BRCA) patients in the TCGA database.

Characteristics		Total	%
Age at diagnosis(y)	≥ 65	186	26.38
	< 65	519	73.62
Histological type	IDC	530	75.18
	ILC	112	15.89
	Other	63	8.94
T stage	T1	184	26.10
	T2	432	61.28
	T3	74	10.50
	T4	15	2.13
N stage	N0	360	51.06
	N1	224	31.77
	N2	83	11.77
	N3	38	5.39
M stage	M1	695	98.58
	M0	10	1.42

IDL, Infiltrating Ductal Carcinoma; ILC, Infiltrating Lobular Carcinoma.

5.The correlation of risk scores and immune cell infiltration in TCGA-BRCA database

Immune cell infiltration was computed using the CIBERSORT algorithm for 1099 BRCA tumor samples obtained from the TCGA database. As the risk score increased, B cell memory, B cell naive, Monocyte, T cell follicular helper, Macrophage M1, T cell CD8+, Myeloid dendritic cell resting was negatively activated, whereas NK cell resting, Macrophage M0, Macrophage M2 were positively activated (Fig. 10).

Discussion

Breast cancer is the highest incidence malignancy in the world and also the second most lethal malignancy in women. Hypoxia was involved in several biological processes in breast cancer such as

invasion, metastasis, and resistance to radiotherapy and chemotherapy. Hypoxia-related genes have been used to develop risk assessment models for predicting patient prognosis[19–21]. With recent advances in single-cell sequencing technology, it's now much more accessible to detect the heterogeneity of cells than traditional sequencing technology. In hence, in this study, we performed differential gene analysis using single-cell sequencing data of breast cancer cells under hypoxic and normoxia conditions, and combined conventional sequencing data and clinical data to construct a prognostic model of hypoxia-related genes in breast cancer patients.

At first, 329 DEGs were screened between the hypoxia and normoxia clusters. GO and KEGG functional analysis results indicated that DEGs identified were mainly related to oxygen metabolism, organism metabolic synthesis, and their functions were mainly enriched with degenerative disease-related pathways and oxygen metabolism-related pathways. GSVA enrichment analysis showed that breast cancer cells in hypoxic state, such as hypoxia, glycolysis and other pathways are upregulated, while oxidative phosphorylation, DNA repair and other pathways are downregulated. Results of the analysis are consistent with experimental conditioning and previous studies.

Based on the analysis results, four gene (ERRFI1, HSPB8, PGK1 and STC2) were thought as independent prognostic factors in BRCA patients. Some of these genes have already been studied and reported in other prognostic model[20, 22]. The biological functions of ERRFI1 mainly act on proliferation, apoptosis, senescence, migration, invasion, epithelial mesenchymal transition, DNA damage and glucose metabolism. ERRFI1 was involved in cancer treatment resistance, but the impact of ERRFI1 in breast cancer remains complex[23]. It has been shown that ERRFI1 is a protective gene in MCF10A, MDA-MB-468 cell lines, but promotes tumor growth by promoting proliferation and inhibiting apoptosis in MDA-MB-231 or MCF-7, respectively[24, 25]. Therefore, the biological function of ERRFI1 needs to be more intensively studied, and its protective role in this study is also consistent with other breast cancer prognostic model[20]. The Heat Shock Protein B8(HSPB8) is a small chaperone that participates in chaperone-assisted selective autophagy (CASA). The expression and function are tumor-specific[26]. The contribution of HSPB8 in breast cancer is still confused. The expression of HSPB8 was significantly elevated in radiation-resistant BC lines compared to parental MCF-7 and MDA-MB-231 cells[27], but the results of Sally Trent et al. showed that overexpression of HSPB8 significantly increased radiosensitivity[28]. HSPB8 silencing in MCF-7 cells inhibits the migration-promoting effect induced by 17 β -estradiol treatment[29], which is consistent with its role as a danger gene in this model. Phosphoglycerate kinase1 (PGK) was take part in a variety of biological activities, including regulation of glucose metabolism, initiation of autophagy, DNA replication and repair in the nucleus of mammalian cells[30]. According to the results of TCGA database analysis, PGK1 was high expressed in most tumor cells, and the expression level correlated with the T-stage and M-stage of BRCA patients[31]. Stanniocalcin 2 (STC2) is a kind of glycoprotein hormone, and its biological behavior in human cancer is tumor type dependent. High expression of STC2 promotes tumor invasion and metastasis in gastric, pancreatic and hepatocellular carcinomas, and negatively correlates with tumor patient survival[32]. However, in breast cancer cells, the expression of STC2 increased after estrogen induction and correlated with improved disease-free survival and longer overall survival[33–35].

In this study, the prognostic model was constructed using genes differentially expressed in MCF-7 cells under hypoxia, which can provide a good understanding of how hypoxia-related genes may act in the prognosis of breast cancer patients. Although the AUC values of this prognostic model showed to be effective in predicting patient survival, the state of the tumor microenvironment is complex, and hypoxia is only one part of it, and there is still room for further improvement of this assessment model.

The results of Pearson correlation analysis revealed that risk score was directly correlated with the expression of NK cell quiescent, macrophage M0, and macrophage M2 positive immune cells, which are weak tumor killers or even promote tumor cell growth. This result is consistent with the prior worse survival results for patients with high-risk scores obtained from the KM survival analysis. However, whether this risk score helps to predict the efficacy of immunotherapy in BRCA patients still needs further study.

Conclusions

In summary, we used single-cell sequencing data from MCF-7 cells under hypoxic conditions in combination with conventional sequencing data and clinical information to identify prognostically relevant genes, and the constructed prognostic model displayed well predictive properties.

Abbreviations

GEO: Gene Expression Omnibus; TCGA: The Cancer Genome Atlas; DEGs: differentially expressed genes; GO: Gene ontology annotation; KEGG: Kyoto Encyclopedia of Genes and Genomes; GSVA: Gene Set Variation Analysis; LASSO: the least absolute shrinkage and selection operator; TME: tumor microenvironment; HIF-1: Hypoxia-inducible factor 1; IFN: interferon; PCA: Principal component analysis; tSNE: t-Distributed Stochastic Neighbor Embedding; OS: overall survival; ROC: receiver operating characteristic; BRCA: Breast invasive carcinoma; ATP: Adenosine Triphosphate; DNA: Deoxyribo Nucleic Acid; AUCs: area under the curves; IDL: Infiltrating Ductal Carcinoma; ILC: Infiltrating Lobular Carcinoma; HSP8: Heat Shock Protein B8; CASA: chaperone-assisted selective autophagy; PGK: Phosphoglycerate kinase1; STC2: Stanniocalcin 2.

Declarations

Ethics approval and consent to participate

We confirm that all methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable

Availability of data and materials

Single-cell RNA transcriptome data of MCF7 cells subjected to 0.1% hypoxia, the bulk tumour transcriptome data and clinical information of breast patient were download from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE134038>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20685>, and <https://portal.gdc.cancer.gov/>, respectively.

Competing interests

The authors declare that they have no competing interests.

Funding

Project supported by the Naional Key R&D Program (2018YFC1313400), the Joint Research Fund for Overseas Chinese, Hong Kong and Macao Scholars (31729001), the National Natural Science Foundation of China (81972869, 81902386, 81903251), the Key R&D Project of Science and Technology Department of Jiangsu Province (BE2018645), Changzhou health and Family Planning Commission youth talent science and technology project (QN201804).

Authors' contributions

RS and JJ designed the study, RS wrote this article, YZ, and YS analyzed the data, WY, DX and WG reviewed and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors want to express gratitude to all participants.

Authors' information

¹Department of Oncology Radiotherapy, The Third Affiliated Hospital of Soochow University, Changzhou, People's Republic of China. ²Department of Tumor Biological Treatment, The Third Affiliated Hospital of Soochow University, Changzhou, People's Republic of China. ³Jiangsu Engineering Research Center for Tumor Immunotherapy, Changzhou, People's Republic of China. ⁴Institute of Cell Therapy, Soochow University, Changzhou, People's Republic of China.

References

1. Anderson NM, Simon MC. The tumor microenvironment. *Current biology* : CB. 2020;30(16):R921-r5.
2. Ruan K, Song G, Ouyang G. Role of hypoxia in the hallmarks of human cancer. *Journal of cellular biochemistry*. 2009;107(6):1053-62.
3. Harrison LB, Chadha M, Hill RJ, Hu K, Shasha D. Impact of tumor hypoxia and anemia on radiation therapy outcomes. *The oncologist*. 2002;7(6):492-508.

4. Vaupel P, Thews O, Hoeckel M. Treatment resistance of solid tumors: role of hypoxia and anemia. *Medical oncology (Northwood, London, England)*. 2001;18(4):243-59.
5. Jing X, Yang F, Shao C, Wei K, Xie M, Shen H et al. Role of hypoxia in cancer therapy by regulating the tumor microenvironment. *Molecular cancer*. 2019;18(1):157.
6. Multhoff G, Vaupel P. Hypoxia Compromises Anti-Cancer Immune Responses. *Advances in experimental medicine and biology*. 2020;1232:131-43.
7. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA: a cancer journal for clinicians*. 2020;70(1):7-30.
8. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA et al. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2007;13(15 Pt 1):4429-34.
9. Liang Y, Zhang H, Song X, Yang Q. Metastatic heterogeneity of breast cancer: Molecular mechanism and potential therapeutic targets. *Seminars in cancer biology*. 2020;60:14-27.
10. Liu ZJ, Semenza GL, Zhang HF. Hypoxia-inducible factor 1 and breast cancer metastasis. *Journal of Zhejiang University Science B*. 2015;16(1):32-43.
11. Ren S, Liu J, Feng Y, Li Z, He L, Li L et al. Knockdown of circDENND4C inhibits glycolysis, migration and invasion by up-regulating miR-200b/c in breast cancer under hypoxia. *Journal of experimental & clinical cancer research : CR*. 2019;38(1):388.
12. Liang Y, Song X, Li Y, Chen B, Zhao W, Wang L et al. LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. *Molecular cancer*. 2020;19(1):85.
13. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nature protocols*. 2014;9(1):171-81.
14. Miar A, Arnaiz E, Bridges E, Beedie S, Cribbs AP, Downes DJ et al. Hypoxia Induces Transcriptional and Translational Downregulation of the Type I IFN Pathway in Multiple Cancer Cell Types. *Cancer research*. 2020;80(23):5245-56.
15. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology*. 2018;36(5):411-20.
16. Lall S, Sinha D, Bandyopadhyay S, Sengupta D. Structure-Aware Principal Component Analysis for Single-Cell RNA-seq Data. *Journal of computational biology : a journal of computational molecular cell biology*. 2018.
17. Pont F, Tosolini M, Fournié JJ. Single-Cell Signature Explorer for comprehensive visualization of single cell signatures across scRNA-seq datasets. *Nucleic acids research*. 2019;47(21):e133.
18. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*. 2012;16(5):284-7.
19. Gong PJ, Shao YC, Huang SR, Zeng YF, Yuan XN, Xu JJ et al. Hypoxia-Associated Prognostic Markers and Competing Endogenous RNA Co-Expression Networks in Breast Cancer. *Frontiers in oncology*. 2020;10:579868.

20. Wang J, Wang Y, Xing P, Liu Q, Zhang C, Sui Y et al. Development and validation of a hypoxia-related prognostic signature for breast cancer. *Oncology letters*. 2020;20(2):1906-14.
21. Gu P, Zhang L, Wang R, Ding W, Wang W, Liu Y et al. Development and Validation of a Novel Hypoxia-Related Long Noncoding RNA Model With Regard to Prognosis and Immune Features in Breast Cancer. *Frontiers in cell and developmental biology*. 2021;9:796729.
22. Li Y, Zhao X, Liu Q, Liu Y. Bioinformatics reveal macrophages marker genes signature in breast cancer to predict prognosis. *Annals of medicine*. 2021;53(1):1019-31.
23. Xu D, Li C. Gene 33/Mig6/ERRFI1, an Adapter Protein with Complex Functions in Cell Biology and Human Diseases. *Cells*. 2021;10(7).
24. Xu J, Keeton AB, Wu L, Franklin JL, Cao X, Messina JL. Gene 33 inhibits apoptosis of breast cancer cells and increases poly(ADP-ribose) polymerase expression. *Breast cancer research and treatment*. 2005;91(3):207-15.
25. Mojica CAR, Ybañez WS, Olarte KCV, Poblete ABC, Bagamasbad PD. Differential Glucocorticoid-Dependent Regulation and Function of the ERRFI1 Gene in Triple-Negative Breast Cancer. *Endocrinology*. 2020;161(7).
26. Cristofani R, Piccolella M, Crippa V, Tedesco B, Montagnani Marelli M, Poletti A et al. The Role of HSPB8, a Component of the Chaperone-Assisted Selective Autophagy Machinery, in Cancer. *Cells*. 2021;10(2).
27. Miao W, Fan M, Huang M, Li JJ, Wang Y. Targeted Profiling of Heat Shock Proteome in Radioresistant Breast Cancer Cells. *Chemical research in toxicology*. 2019;32(2):326-32.
28. Trent S, Yang C, Li C, Lynch M, Schmidt EV. Heat shock protein B8, a cyclin-dependent kinase-independent cyclin D1 target gene, contributes to its effects on radiation sensitivity. *Cancer research*. 2007;67(22):10774-81.
29. Piccolella M, Crippa V, Cristofani R, Rusmini P, Galbiati M, Cicardi ME et al. The small heat shock protein B8 (HSPB8) modulates proliferation and migration of breast cancer cells. *Oncotarget*. 2017;8(6):10400-15.
30. He Y, Luo Y, Zhang D, Wang X, Zhang P, Li H et al. PGK1-mediated cancer progression and drug resistance. *American journal of cancer research*. 2019;9(11):2280-302.
31. Shao F, Yang X, Wang W, Wang J, Guo W, Feng X et al. Associations of PGK1 promoter hypomethylation and PGK1-mediated PDHK1 phosphorylation with cancer stage and prognosis: a TCGA pan-cancer analysis. *Cancer communications (London, England)*. 2019;39(1):54.
32. Joshi AD. New Insights Into Physiological and Pathophysiological Functions of Stanniocalcin 2. *Frontiers in endocrinology*. 2020;11:172.
33. Chang AC, Jellinek DA, Reddel RR. Mammalian stanniocalcins and cancer. *Endocrine-related cancer*. 2003;10(3):359-73.
34. Jansen MP, Sas L, Sieuwerts AM, Van Cauwenberghe C, Ramirez-Ardila D, Look M et al. Decreased expression of ABAT and STC2 hallmarks ER-positive inflammatory breast cancer and endocrine therapy resistance in advanced disease. *Molecular oncology*. 2015;9(6):1218-33.

35. Essegir S, Kennedy A, Seedhar P, Nerurkar A, Poulson R, Reis-Filho JS et al. Identification of NTN4, TRA1, and STC2 as prognostic markers in breast cancer in a screen for signal sequence encoding proteins. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2007;13(11):3164-73.

Figures

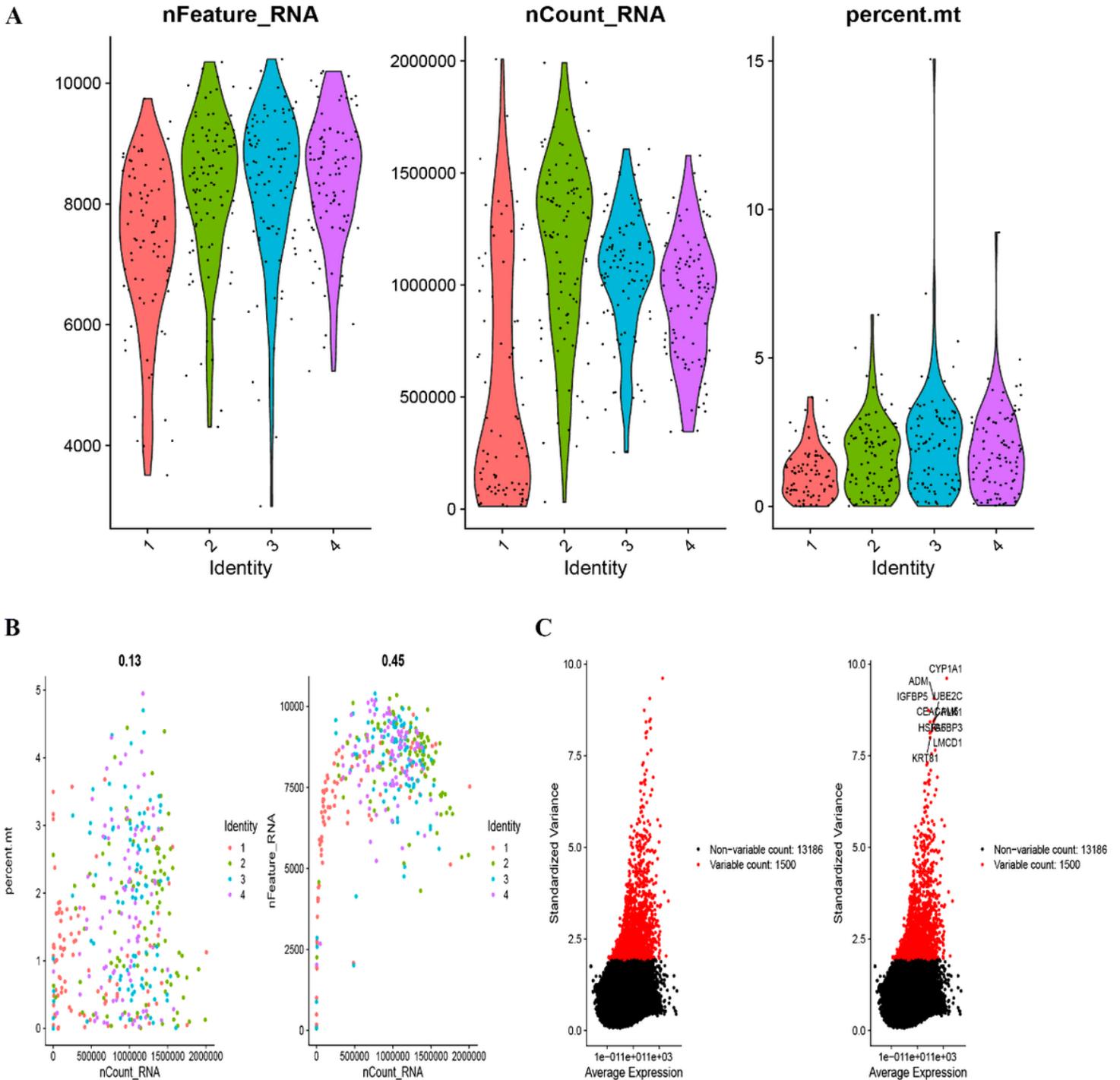


Figure 1

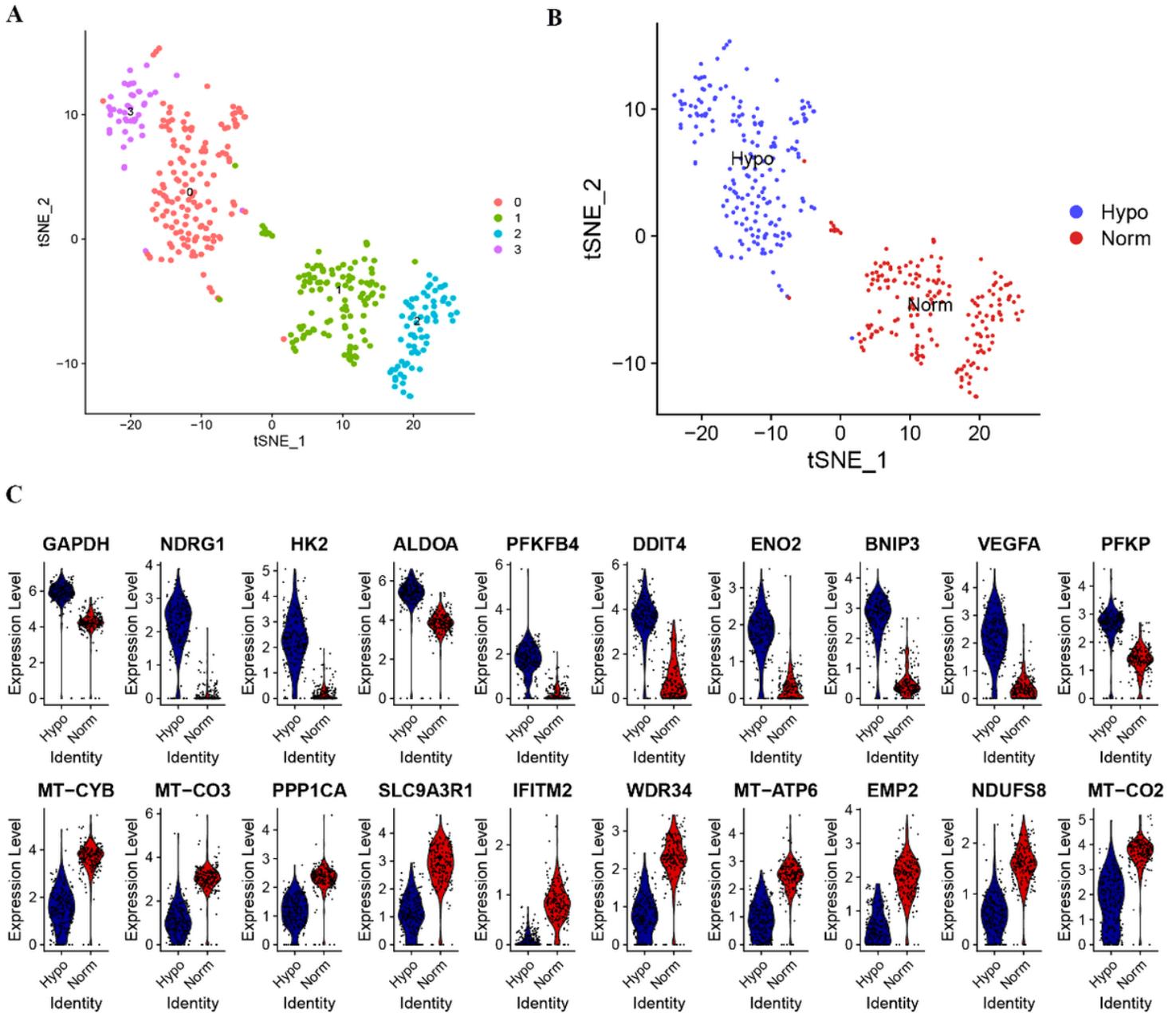


Figure 3

Clustering and identification of DEGs of MCF7 cells among 0.1% hypoxia or normoxia microenvironment. (A) Basing on the t-SNE algorithm, scRNA sequencing profiles were classified into 4 cell clusters. (B) Based on experimental design, cluster 0 and cluster 3 from 0.1% hypoxic microenvironment, cluster 1 and cluster 2 from normoxic microenvironment. (C) Violin plot shows the top 10 up-regulated genes and the top 10 down-regulated genes in hypoxia microenvironment.

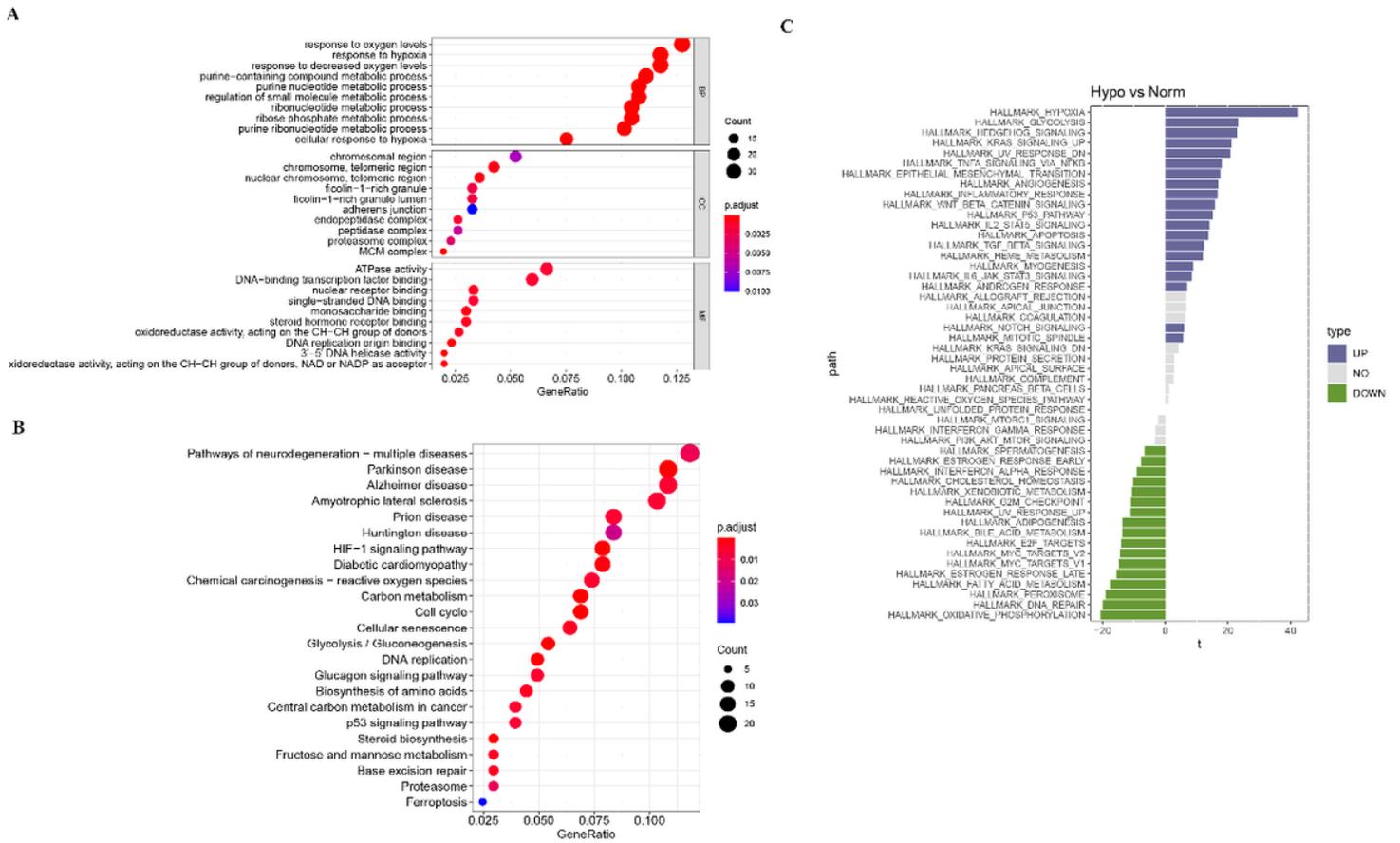


Figure 4

Functional analysis of hypoxia related DEGs. (A) Gene Ontology (GO) function enrichment results. (B) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis results. (C) Gene set variation analysis (GSVA) results of the differences in gene enrichment between the hypoxia and normoxia clusters.

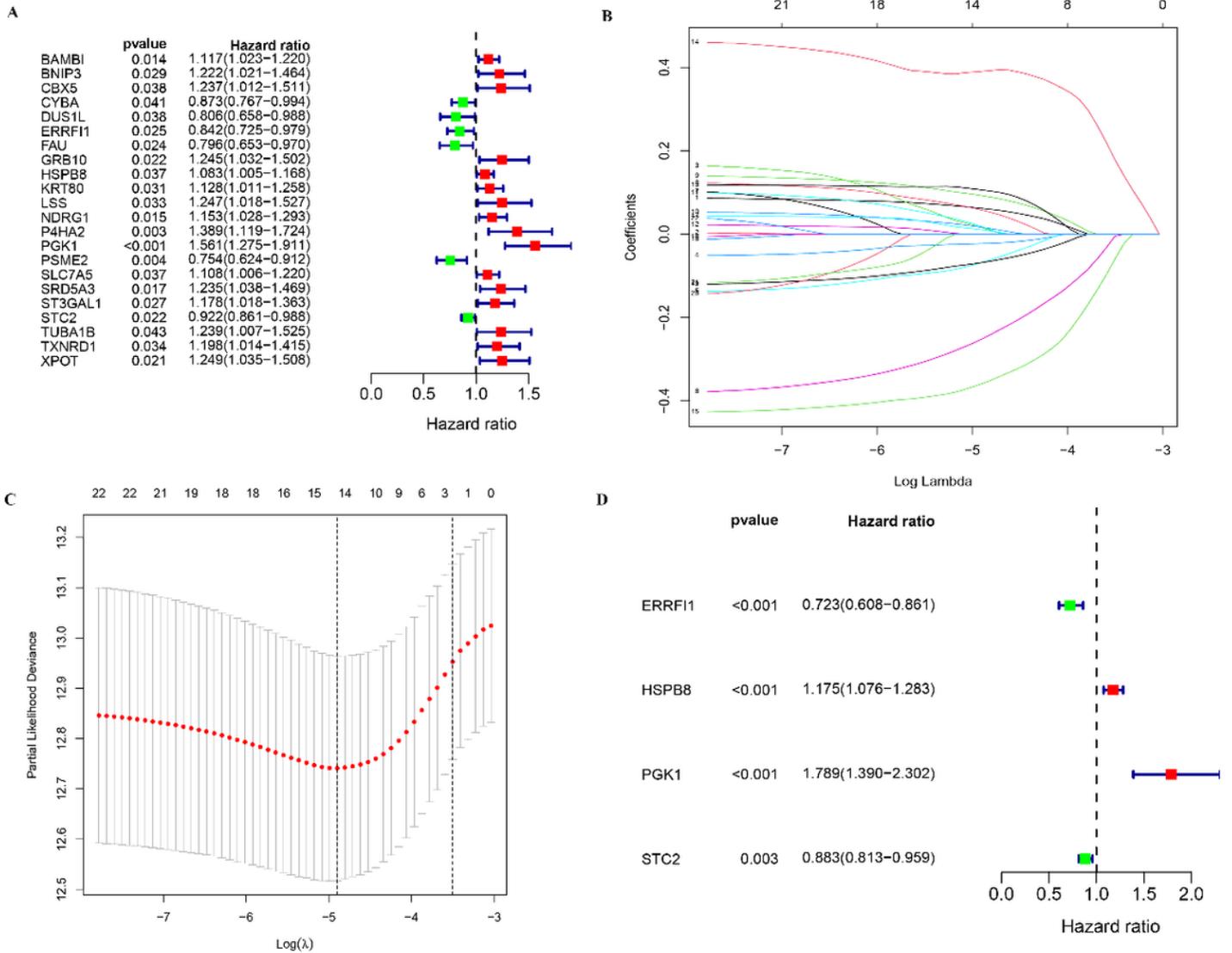


Figure 5

Identification of hypoxia-associated prognostic genes for survival prediction. (A) Univariate analysis results of 22 hypoxia-related DEGs that were associated with OS in BRCA patients. (B) Lasso coefficient profile of the 22 hypoxia-related DEGs. (C) Selection the optimal parameter (λ) in the LASSO model. (D) Multivariate Cox analysis showed that four hypoxia-related DEGs were the independent prognostic factors identified for prognosis of survival prediction.

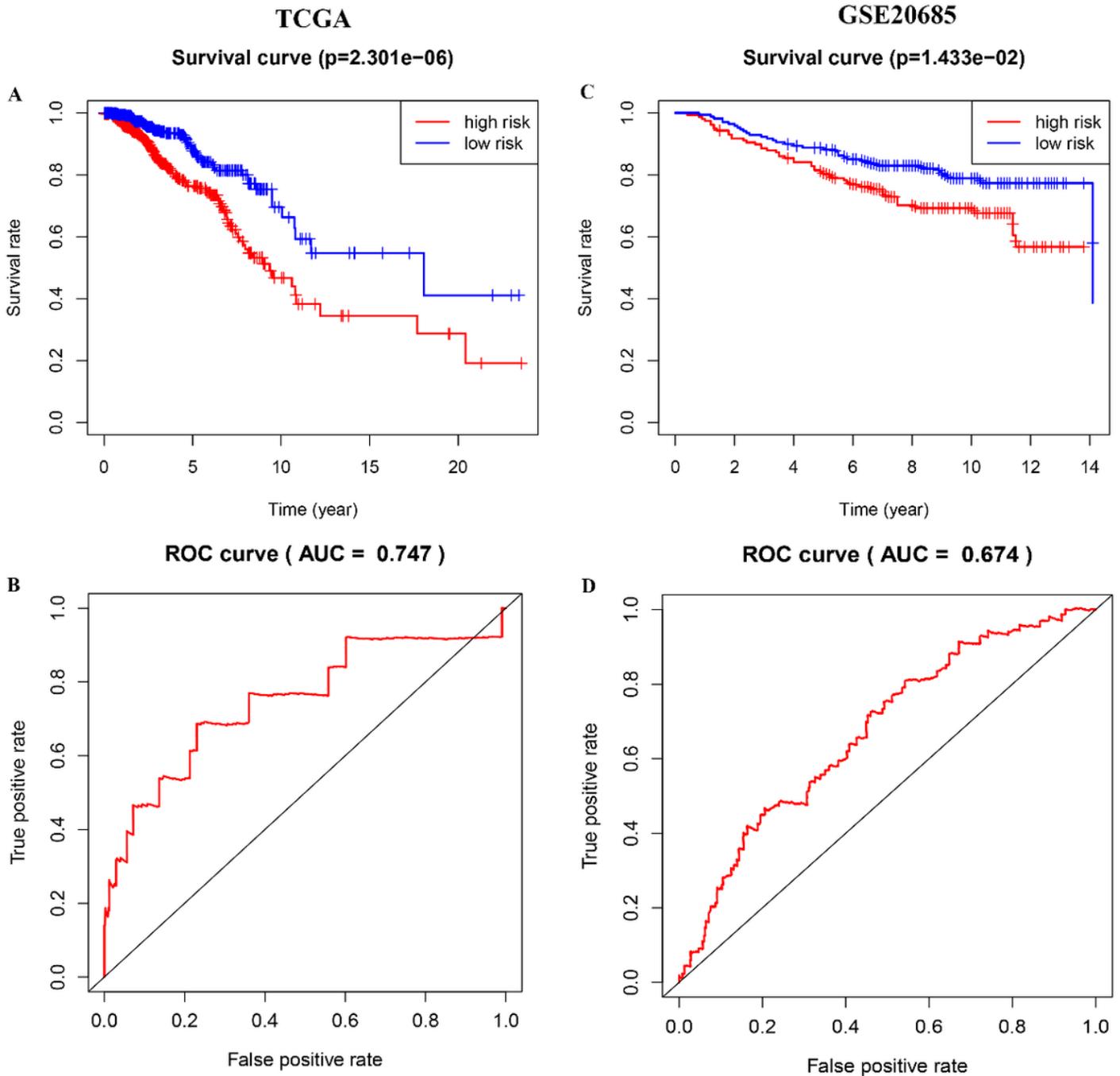


Figure 6

Kaplan-Meier curves and the receiver operating characteristic (ROC) curve of hypoxia related prognostic model for overall survival of breast invasive carcinoma (BRCA) patients. The KM survival curves in the Cancer Genome Atlas (TCGA) dataset (A) and the GSE20685 dataset(C) showed that the low-risk group survived longer than the high-risk group. The receiver operating characteristic (ROC) curve of hypoxia related prognostic for overall survival of BRCA patients in the Cancer Genome Atlas (TCGA) dataset(B)and GSE20685 dataset(D).

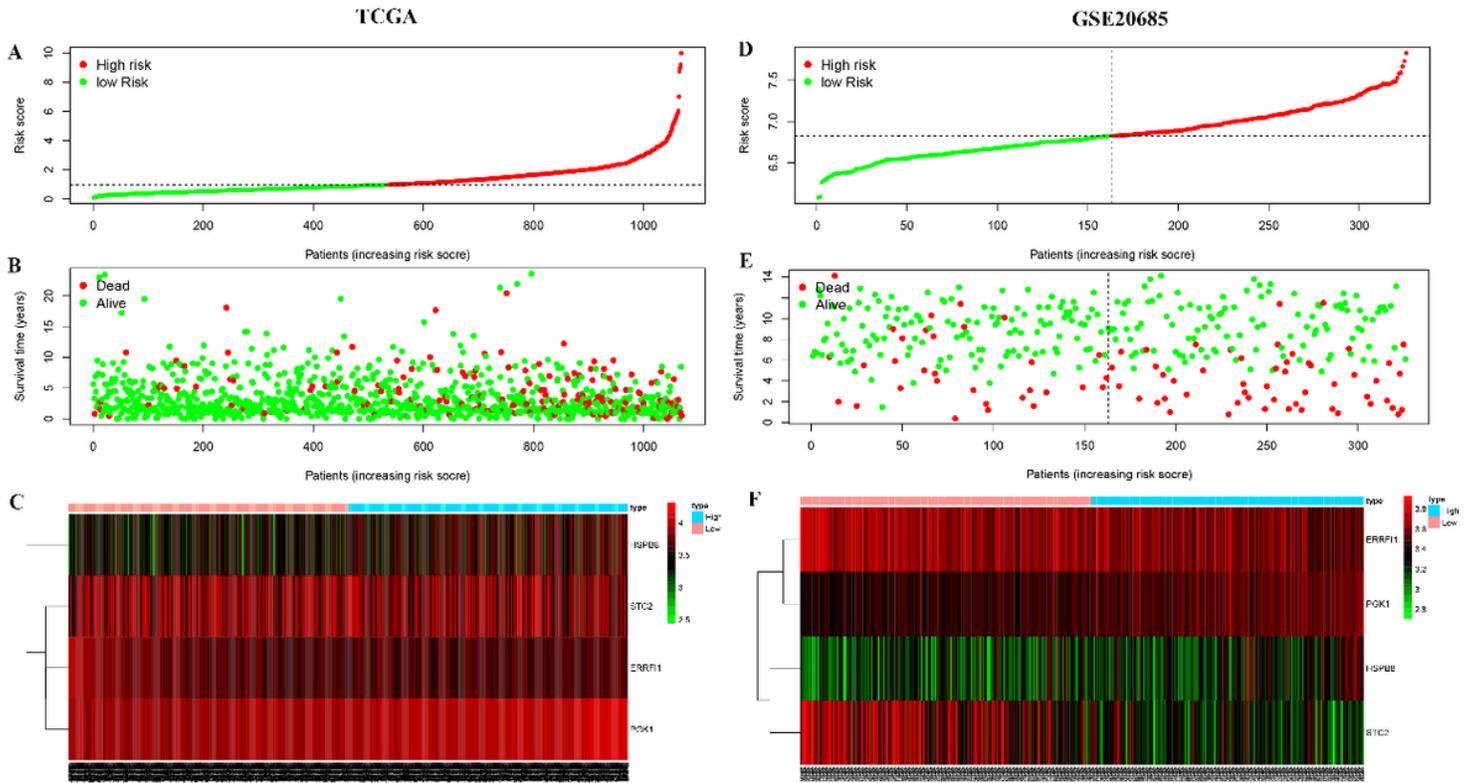


Figure 7

Construction and validation of hypoxia related prognostic in breast invasive carcinoma (BRCA) patients. The risk scores of BRCA patient distribution in TCGA dataset(A)and GSE20685 dataset(D). Survival and death of patients in different risk scores subgroups in TCGA dataset(B)and GSE20685 dataset(E). Heatmap of differentially expressed hypoxia related genes in TCGA dataset(C)and GSE20685 dataset(F).

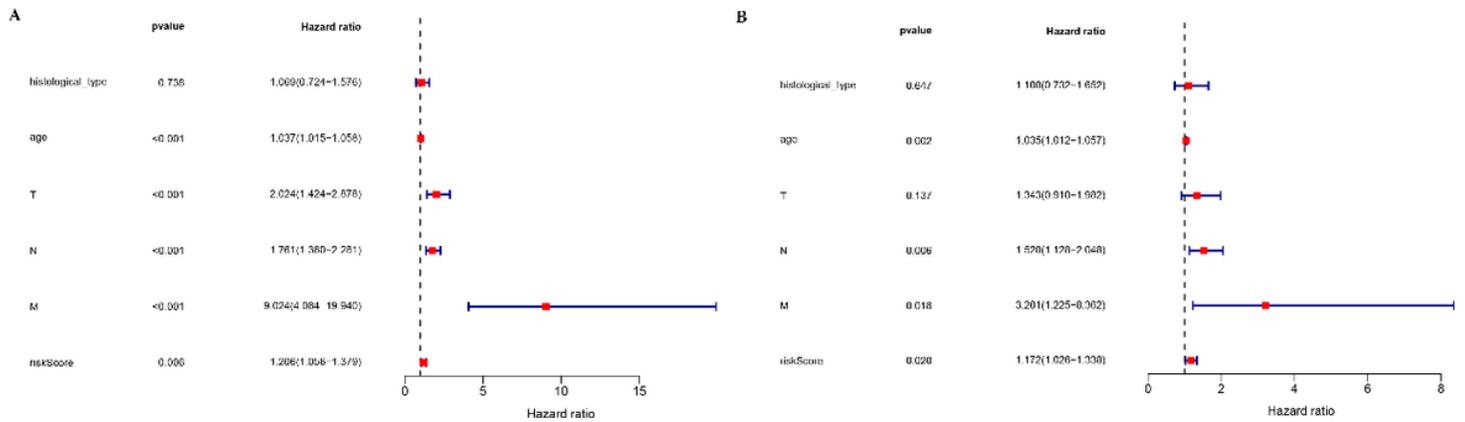


Figure 8

Univariate (A) and multivariate Cox regression analysis (B) of the correlation results between risk scores, clinical parameters and overall survival in patients with breast invasive carcinoma (BRCA).

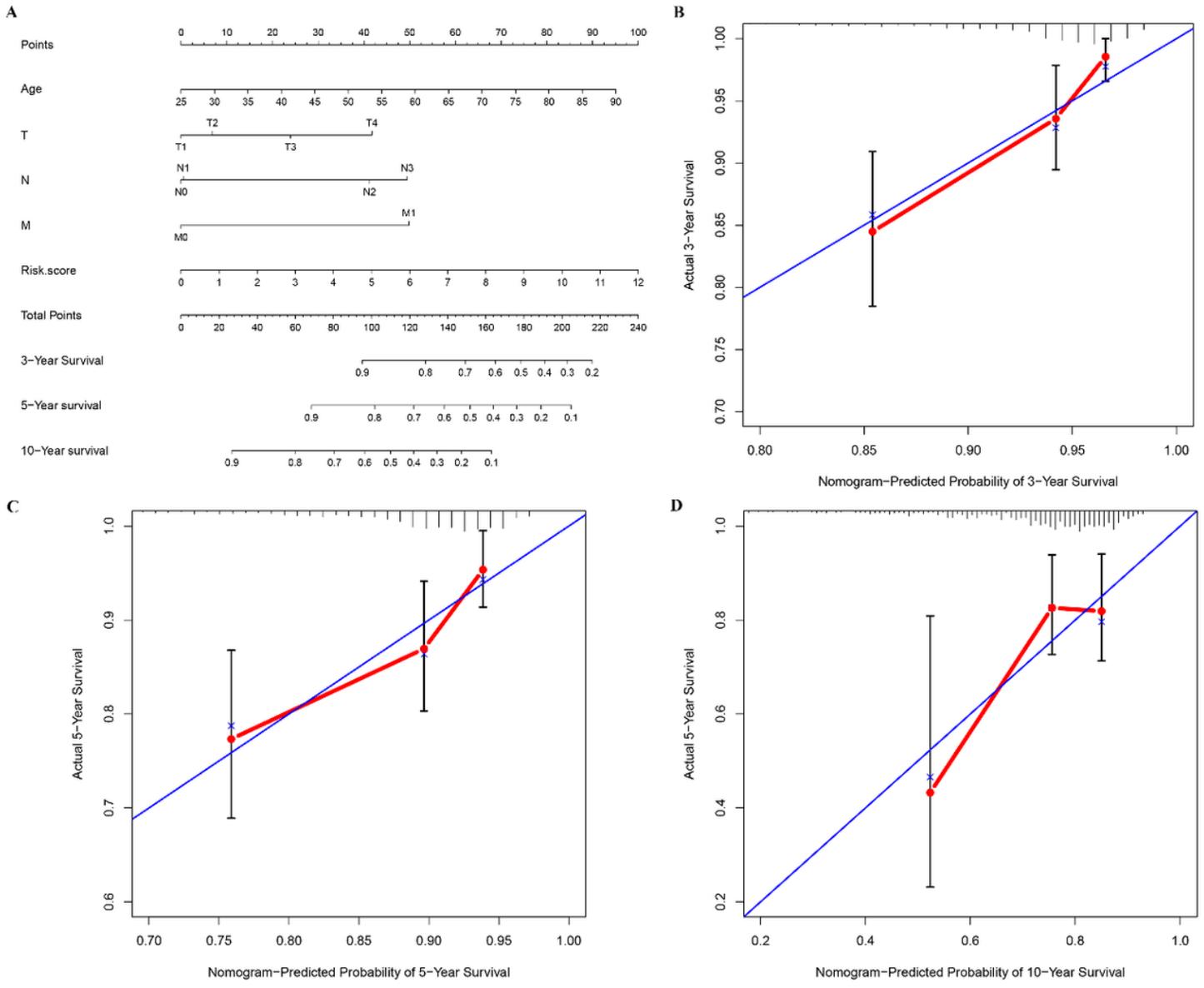


Figure 9

Nomogram for predicting survival of breast invasive carcinoma (BRCA) patients. (A) The nomogram to predict the overall survival of BRCA patients. (B-D) The calibration curves for the nomogram when predicting 3-, 5-, and 10-years overall survival for BRCA patients.

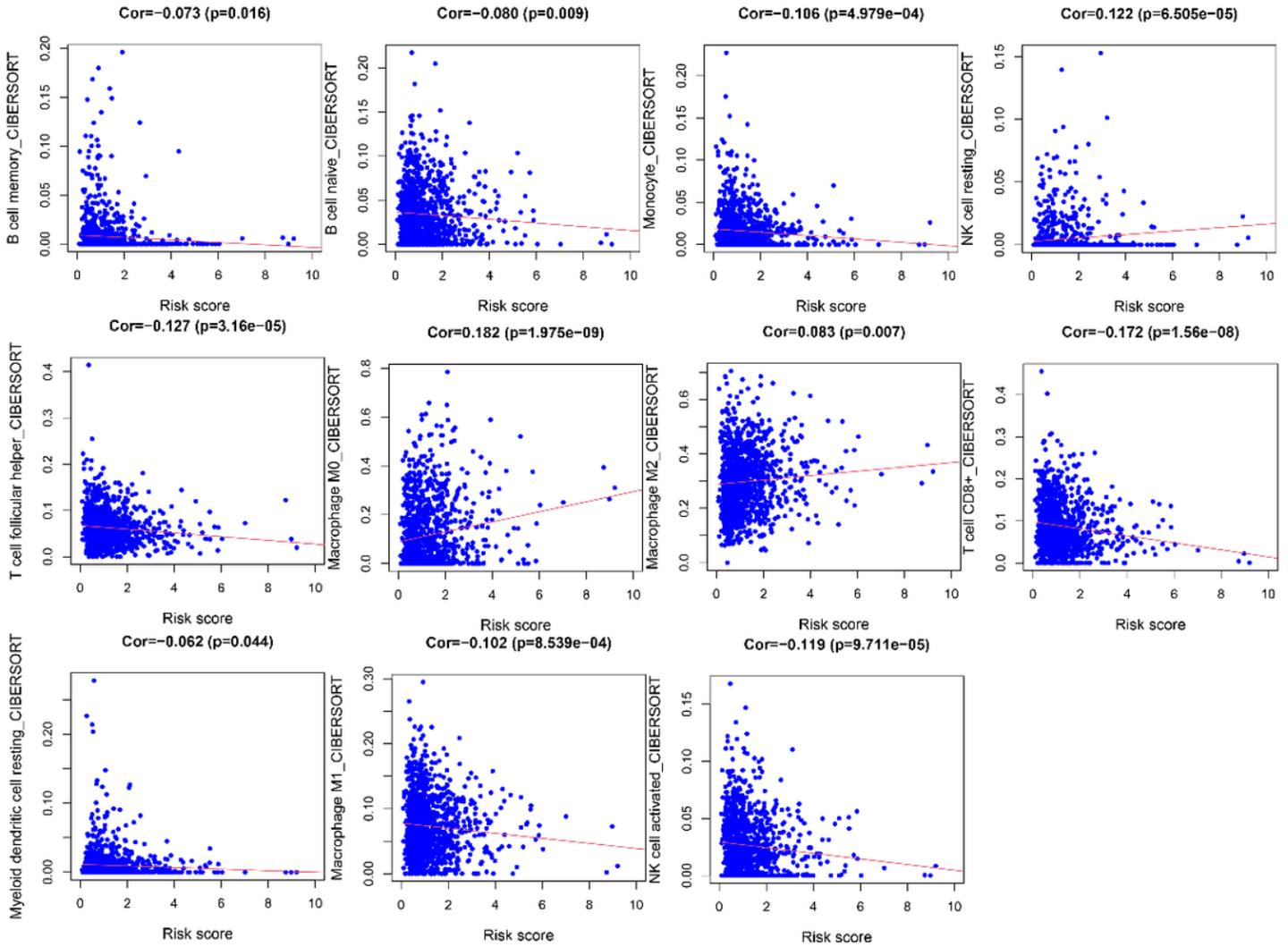


Figure 10

Relationships between risk scores and immune cell infiltration in TCGA-BRCA database.