Construction and Validation of Immune-Related LncRNA Signature to Predict the Prognosis and Therapeutic Efficacy of Breast Cancer

Shujing Wang
Anhui Medical University
Jiezhen Yang
Anhui Medical University
Louis Kwantwi
Anhui Medical University
Changhai Ding
University of Tasmania
Qiang Wu (wuqiang@ahmu.edu.cn)
Anhui Medical University

Research Article

Keywords: breast cancer, The Cancer Genome Atlas, immune-related gene, long non-coding RNA, prognosis.

Posted Date: February 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1338441/v1

Abstract

Breast cancer ranks first in morbidity and mortality among women worldwide. The immune microenvironment including immune-related long non-coding RNAs (irlncRNAs) are involved in various processes in tumors. We downloaded the RNA-sequencing data and corresponding clinical characteristics of breast samples from TCGA website and obtained the irlncRNAs and identified differentially expressed irlncRNAs (DEirlncRNAs) compared tumor with normal samples. Cox and Lasso regression analysis were used to establish a risk model after DEirlncRNAs were cyclically and separately paired. 15 DEirlncRNAs pairs were used to construct the risk model and to distinguish the high- or low-risk groups based on the median value of risk score. Patients in the high-risk group had a poorer prognosis, more aggressive clinicopathologic features, lower expression of immune checkpoint inhibitors, and higher drug sensitivity. The high-risk group was positively related to cancer-promoting immune cells like M2 macrophages, and negatively associated with anti-cancer immune cells like T cells. Gene set enrichment analysis were performed to investigate the potential molecular mechanisms on tumor progression. We constructed a risk model, which based on the relative expression values of the two irlncRNAs, could predict the prognosis of breast cancer patients, as well as the efficacy of immunotherapy and chemotherapy.

Introduction

According to the global report, 2.26 million new cases of breast cancer were diagnosed and 0.68 million died from the disease in 2020[1]. Breast cancer continues to threaten maternal health by ranking first in morbidity and mortality[2]. The disease is known for its complex clinical manifestations, morphological and molecular biological characteristics, and therapeutic resistance[3]. Therefore, finding a suitable detection method to predict the prognosis of breast cancer is imminent due to the disease's high mortality rate and heterogeneity.

Long non-coding RNAs (lncRNAs) are defined as RNAs longer than 200 nucleotides that do not encode proteins. LncRNAs, accounting for approximately 80% of the human transcriptome, perform many biological functions depending on their location. In the nucleus, lncRNAs perform diverse roles, including regulating gene expression in cis or trans, regulation of splicing, and nucleation of subnuclear domains. Their presence in the cytoplasm is associated with cytoplasmic functions such as miRNA sponging, interaction with signaling proteins, and modulation of translation of specific mRNAs[4, 5].

The tumor microenvironment is a crucial variable to breast cancer progression[6, 7]. This results from the infiltration of the tumor microenvironment by several immune cells such as T cells, B cells, and lymphocytes infiltrating around the breast cancer and tumor stroma[8]. Emerging evidence has suggested that the dysregulation of these immune cells correlates to immunosuppression and progression in several malignant tumors. Hence, the utility of immunotherapy in the treatment of tumors has been very critical in recent times[9, 10]. Accordingly, some of the biomarkers associated with T cells, known as immune checkpoint inhibitors (ICIs) like PD-1/PD-L1, have shown to be a novel strategy in breast cancer

treatment[11]. To further understand the benefit and risk associated with these therapies, biomarkers that can predict the treatment response are urgently needed.

Over the past years, growing evidence shows that lncRNAs can regulate tumor immunity in immune cells and the immune microenvironment[12, 13]. Accordingly, the relationship between lncRNAs and cancer immunity has received increasing attention. Some immune-related lncRNAs (irlncRNAs) are involved in tumor cell migration, invasion, epithelial-mesenchymal transformation, and metabolism. In breast cancer, lncRNA SNHG1 regulates the differentiation of Treg cells to promote the immune escape of breast cancer via regulating miR-448/IDO[14]. Furthermore, lncRNA HISLA from tumor-associated macrophages regulates aerobic glycolysis in breast cancer cells[15]. Moreover, lncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis[16]. In addition, lncRNA GATA3-AS1 facilitates tumor progression and immune escape in triple-negative breast cancer through destabilization of GATA3 and stabilization of PD-L1[17].

Signatures focusing on the tumor immune infiltration show promising predictive and prognostic value in the diagnosis, evaluation, and treatment of cancer. Individual genes are generally used to analyze immune-related signature to predict the prognosis of breast cancer. Here, we utilized a novel model of irlncRNAs pairing and iteration, which predicted breast cancer prognosis irrespective of the specific expression levels.

Results

Identification of Differentially Expressed irlncRNAs (DEirlncRNAs)

The flow chart of the study was shown in Fig. 1. Firstly, we downloaded transcriptome profiling consisting of 1053 breast cancer samples and 111 normal samples from the TCGA and then annotated according to gene transfer format (GTF) files from Ensembl. Next, we obtained the irlncRNAs by co-expression of immune-related genes from the ImmPort database. A total of 1041 irlncRNAs were identified (shown in Supplementary Table S1), and 55 were distinguished as DEirlncRNAs (Fig. 2a). Among the 55 distinguished DEirlncRNAs, 39 were upregulated while 16 were downregulated (Fig. 2b, Supplementary Table S2).

Construction Of Deirlncrna Pairs And A Prognostic Model

According to the iteration loop and the 0 or 1 matrix, we obtained 1172 valid DEirlncRNA pairs. Following, the DEirlncRNA pairs were merged with the survival information. Through Cox regression analysis, we obtained 20 DEirlncRNA pairs that were correlated with survival (Fig. 2c). Furthermore, 15 DEirlncRNA pairs were included in the model after Lasso regression analysis (Fig. 2d and e, Table 1).

Table 1
The risk model of 15 DEirlncRNAs pairs for breast cancer.

Pairs	Coef
AP005131.7|LINC00511	-0.01085
AP005131.7|AC009093.1	-0.25014
U62317.4|AP000251.1	-0.27626
AC011247.1|ATP2A1-AS1	-0.17886
C6orf99|LINC01087	0.104848
C6orf99|U62317.1	0.294303
C6orf99|AC020663.2	0.272957
C6orf99|LINC02544	0.411963
LINC01929|AC020663.2	0.083829
LINC01929|AP005131.2	0.213263
U62317.1|ZNF350-AS1	-0.05333
LINC00511|AP005233.2	0.010714
LINC00511|LINC01152	0.092366
AL356417.2|AP005131.2	0.337368
AC009093.1|AP005131.2	0.063

In order to verify the accuracy of the prognostic model, we drew the ROC curve and calculated the AUC value. We analyzed the model in the training set, and the 1-year AUC was 0.838, the 5-year AUC was 0.688, and the 10-year AUC was 0.788 (Fig. 3a). We validated the model in the validation set, and observed that all the AUC values were more than 0.680 (1-year AUC = 0.760, 5-year AUC = 0.719, 10-year AUC = 0.683, Fig. 3b). Similar trend was obtained for the total set (1-year AUC = 0.797, 5-year AUC = 0.700, 10-year AUC = 0.736, Fig. 3c).

According to the median value of risk score in the training set, all the samples were divided into the high-risk group and the low-risk group. We then performed a survival analysis of the training, the validation, and the total sets using Kaplan-Meier curves. As shown in Fig. 3d-f, patients in the high-risk group had worse overall survival than those in the low-risk group ($P < 0.001$).

Assessment of the Correlation between the Risk Model and Clinicopathological Characteristics

The risk curves and scatter plots were used to display the risk score and the survival outcome of each breast cancer patient in the training, the validation, and the total sets. The results showed that the

mortality in the low-risk group was lower than in the high-risk group (Fig. 4a-f). The heatmap was used to show the relationship between DEirlncRNAs pairs and risk score in the three sets. AP005131.7 | LINC00511, AP005131.7 | AC009093.1, U62317.4 | AP000251.1, AC011247.1 | ATP2A1-AS1, U62317.1 | ZNF350-AS1 pairs had more values of 1 in the high-risk group than in the low-risk group. On the contrary, C6orf99 | LINC01087, C6orf99 | U62317.1, C6orf99 | AC020663.2, C6orf99 | LINC02544, LINC01929 | AC020663.2, LINC01929 | AP005131.2, LINC00511 | AP005233.2, LINC00511 | LINC01152, AL356417.2 | AP005131.2, AC009093.1 | AP005131.2 pairs had more values of 1 in the low-risk group than in the high-risk group (Fig. 4g-i).

To evaluate whether the risk model of the DEirlncRNAs pairs was an independent prognostic factor for breast cancer, univariate and multivariate Cox regression analyses were conducted. In the training set, the hazard ratio (HR) of risk score and 95% CI were 1.737 and 1.419–2.116 in univariate Cox regression analysis ($P < 0.001$), and 1.714 and 1.338–2.197 in multivariate Cox regression analysis ($P < 0.001$) respectively (Fig. 5a and b). This suggested that the risk model of DEirlncRNAs pairs were independent prognostic factors in patients with breast cancer. In the validation set, the risk model also showed statistical differences by univariate Cox regression analysis ($P < 0.001$, HR = 1.728, 95% CI [1.334–2.239]) and multivariate Cox regression analysis ($P < 0.001$, HR = 1.816, 95% CI [1.369–2.408], Fig. 5c and d). Similar results were obtained in the total set by univariate Cox regression analysis ($P < 0.001$, HR = 1.722, 95% CI [1.471–2.016]) and multivariate Cox regression analysis ($P < 0.001$, HR = 1.697, 95% CI [1.416–2.035], Fig. 5e and f).

Furthermore, we compared the differences in clinicopathological characteristics between the low-and high-risk group of all the samples. As shown by the strip chart and scatter plots, ER, PR, age, clinical stage, T stage, M stage, and N stage were significantly related to the risk score (Fig. 6).

Estimation of Tumor-Infiltrating Immune Cells and ICIs with Risk Assessment Model

Because lncRNAs play an important role in tumor immunity, we investigated the relationship between the model and the tumor immune microenvironment using all the breast cancer samples. As shown in Fig. 7a and b the immune score and stroma score were higher in the low-risk group than in the high-risk group ($P < 0.05$). We compared various tumor-infiltrating immune cells between the high-risk and low-risk groups, the results are listed in Supplementary Fig. S1. In addition, we integrated the relationships between various immune cells within the risk group performed by Spearman correlation analysis into a bubble chart. The results showed that the risk score was more positively associated with M2 macrophages and cancer-associated fibroblasts, whereas they were negatively associated with T cells, CD8 + T cells, NK cells, B cells, and M1 macrophages (Fig. 7c, Supplementary Table S3).

Immunotherapy is a novel and effective treatment method for breast cancer. We analyzed differences in ICIs common in breast cancer between high- and low-risk groups. The expression of PCDC1 (PD-1), CD274 (PD-L1), CTLA4, and CDK4 were significantly lower in the high-risk group (Fig. 8a-d).

Analysis of the Correlation between the Risk Model and Chemotherapeutic Drugs

The prediction of how effective a chemotherapy drug can guide the selection of clinical drugs. We assessed the correlation between the risk model and the efficacy of common chemotherapeutics used in the treatment of breast cancer. A higher IC50 of chemotherapeutic agents, such as methotrexate, doxorubicin, and gemcitabine were associated with the high-risk group ($P < 0.05$, Fig. 8e-g). While paclitaxel had no significance with the model ($P > 0.05$, Fig. 8H). The data indicated that the model might predict the treatment response to chemotherapy agents.

Gene Set Enrichment Analysis

The GSEA results indicated that the high-risk score group had markedly positive correlations with 6 enrichment pathways and negative correlations with 23 enrichment pathways (Supplementary Table S4). As shown in Fig. 9, the enriched KEGG pathways contained “glycolysis gluconeogenesis”, “citrate cycle TCA cycle”, “fructose and mannose metabolism”, “pentose phosphate pathway”, “terpenoid backbone biosynthesis” and “steroid biosynthesis”.

Discussion

Breast cancer is the most commonly diagnosed cancer among women worldwide. Although the 5-year survival rate of stage II breast cancer is about 93% and 72% for stage III [3], breast cancer is still the leading cause of cancer death [1]. Therefore, besides the traditional clinical risk factors, additional biomarkers to predict the prognosis and treatment of breast cancer are needed as well.

In recent years, the utility of immunotherapy for the treatment of breast cancer patients has widely gained attention. Although previous studies have used immune-related lncRNAs to predict breast cancer prognosis, the signature in those studies was constructed using a single gene[18, 19]. Herein, our study presents the first report using a differentially expressed immune-related lncRNAs pairs model to predict the treatment and prognosis of breast cancer. The two-lncRNA pairs were superior to a single gene, that it is not dependent on the expression levels of each gene, but the relative expression of the two genes. For different detection systems, data correction can be done without.

To construct the model, we downloaded breast cancer transcription data from the TCGA database and obtained DEirlncRNAs. The DEirlncRNAs pairs were identified using an improved method of cyclically and singly pairing along with a 0 or 1 matrix. We further performed univariate regression analysis to select survival-related DEirlncRNAs pairs. To better verify the accuracy of the model, we randomly and equally divided the samples into the training set and the validation set. The Lasso regression analysis was used to screen the DEirlncRNAs pairs for the risk model construction in the training set. According to the median value of risk score in the training set, all the patients, whether in the training set or the validation set, were divided into high-risk and low-risk groups. Next, we calculated the AUC value from the ROC curve to validate the risk model. Finally, we evaluated the accuracy of the model by analyzing the differences of

various clinical factors including survival, clinicopathological features, tumor-infiltrating immune cells, and immune checkpoints.

In our signature, 15 DEirlncRNAs pairs consisting of 18 DEirlncRNAs were used to construct the model. Some of the DEirlncRNAs identified in our study have been reported to play an important role in malignant tumors. For example, LINC00511 may contribute to breast cancer tumorigenesis, proliferation, radioresistance, and stemness[20, 21]. The tumor-promoting functions of LINC00511 have also been reported in gastric cancer[22, 23], hepatocellular carcinoma[24, 25], colorectal cancer[26, 27] and bladder cancer[28, 29]. LINC01087 could represent a novel, specific and promising biomarker not only for the diagnosis and prognosis of luminal subtypes and triple-negative breast cancers but also as a predictive biomarker of pharmacological interventions[30]. It has also been reported that the overexpression of LINC01087 in breast cancer can promote the invasion and migration of breast cancer cells[31]. LINC02544 may promote proliferation, invasion, and migration of breast cancer cells after neoadjuvant chemotherapy[32]. LINC01152 could induce tumorigenesis in glioblastoma via the Notch signaling pathway[33] and promotes cell proliferation and survival in hepatocellular carcinoma[34]. Amelia et al. found that there were many lncRNAs dysregulation in non-small cell lung carcinoma, in which LINC01929 was upregulated[35]. LINC01929 functioned as a tumor-promoting lncRNA in oral squamous cell carcinoma via the miR-137-3p/FOXC1 axis[36]. In addition, some DEirlncRNAs such as AL356417.2, AP005233.2, ATP2A1-AS1, C6orf99, U62317.1, and U62317.4 appear only in bioassay and have not been experimentally confirmed. Other DEirlncRNAs, AC009093.1, AC011247.1, AC020663.2, AP000251.1, AP005131.2, AP005131.7, and ZNF350-AS1 were revealed for the first time.

Over the years, the role of immune cells in breast cancer has been increasingly discovered. To investigate the relationship between risk score and tumor-infiltrating immune cells, we used seven common acceptable methods including XCELL[37], TIMER[38, 39], QUANTISEQ[40, 41], MCPcounter[42, 43], EPIC[44, 45], CIBERSORT-ABS[46], and CIBERSORT[47, 48] to estimate the infiltration of immune cells in breast cancer samples. Due to the differences and complexity among the various algorithms, the results were not compared with each other. Our results showed that the risk score was more positively related to cancer-promoting immune cells like M2 macrophages and cancer-associated fibroblast, while the risk score was negatively related to T cells, CD8 + T cells, NK cells, B cells, and M1 macrophages, which were proved to be anti-cancer. In addition to immune cells, we also analyzed the relationship between risk score and ICIs common to breast cancer. ICIs, especially those associated with T cells, have been used in clinical treatment. The first immunotherapies to the immunomodulatory receptor CTLA4 and blockade of the immunoinhibitory receptor PD-1 in cancer immunotherapy have created a paradigm of cancer therapy[11, 49]. Based on our results, PCDC1 (PD-1), CD274 (PD-L1), and CTLA4 were highly expressed in the low-risk group. Our findings suggested that the risk model can predict the efficacy of immunotherapy in clinical settings.

Chemotherapy is a common treatment for breast cancer and hence we analyzed the relationship between risk score and IC50 of the 4 common chemotherapy drugs used in breast cancer treatment. It was

observed that methotrexate, doxorubicin, and gemcitabine were associated with the high-risk group. This finding showed that the risk model could indicate the chemotherapeutics sensitivity.

Further, we divided the samples into the training set and the validation set. The training set was used to establish and verify the model, while the validation and the total sets were used for further verification. The drawback of our study is that no external validation was conducted. Nonetheless, our model basing on the relative expression of the two genes reduced the errors caused by differences in different expression. Various methods were used to verify our model, and satisfactory results were obtained. We assumed that our model was acceptable. Of course, having external validation was even better.

Identification of the DEirlncRNAs signature could predict the prognosis, indicate the clinical outcome, and estimate the efficacy of immunotherapy and chemotherapy in breast cancer patients. According to the results of this present study, we hope to establish a more convenient clinical prognostic model.

Methods

Data Source, Preprocessing and Differentially Expressed Analysis

The RNA-sequencing (RNA-Seq) data consisting of 1164 female breast samples were downloaded from The Cancer Genome Atlas (TCGA) website (https://portal.gdc.cancer.gov). The corresponding clinical characteristics such as age, survival information, and clinical stage, were also downloaded from TCGA. R4.1.0 software was used to normalize, process, and analyze the data. Perl (https ://www.perl.org) was used to convert the Ensembl ID of genes into a matrix of gene symbols and merge the RNA-Seq data files into a matrix file. Gene transfer annotation files were downloaded from Ensembl (http://asia.ensembl.org) to distinguish mRNAs from lncRNAs for further analysis. A list of recognized immune-related genes was downloaded from the ImmPort database (http://www.immport.org) and was used to screen irlncRNAs by a co-expression strategy. The immune-related genes with correlation coefficients more than 0.4 and p-value less than 0.001 were considered as irlncRNAs. All methods were performed in accordance with the relevant guidelines and regulations.

To identify the differentially expressed irlncRNAs (DEirlncRNA), we used the “limma” software package in R for differential expression analysis. The cutoff conditions were set as: |log2 fold change (logFC)| > 2.0, false discovery rate (FDR) < 0.05.

Pairing Deirlncrnas

The DEirlncRNAs were cyclically and separately paired, and a 0-or-1 matrix was constructed assuming Z is equal to lncRNA X plus lncRNA Y; Z is defined as 1 if the expression level of lncRNA X is higher than lncRNA Y, otherwise Z is defined as 0. Then, the constructed 0-or-1 matrix was further screened. When the

expression quantity of lncRNA pairs was 0 or 1, there was no relationship between pairs and prognosis. An effective match was considered when the pairs of lncRNAs with 0 or 1 expression exceeded 20%.

Construction Of The Prognostic Model

First, the survival information was combined with irlncRNAs pairs after which a univariate Cox regression analysis was conducted to select survival-related irlncRNAs pairs with $P < 0.01$ as filter criteria. Then we divided the samples randomly and equally into a training set (n = 518) and a validation set (n = 517), combining the two sets was the total set. In the training set, the Lasso regression analysis was performed with cross validation to select the irlncRNAs pairs most correlated with prognosis. These final irlncRNAs pairs were used for the construction of the prognostic model. The following formula was used to calculate the risk score with the constructed risk model for all the cases[50, 51]: RiskScore = Exp1*Coef1 + Exp2*Coef2 + …… + Exp*i**Coef*i* (Exp*i* represents the expression level of each irlncRNA pair, and Coef*i* represents the coefficient of each irlncRNA pair).

Application And Validation Of The Risk Model

The receiver operating characteristic (ROC) curve was used to evaluate the predicted values of the model and the "survival ROC" package of R was used to calculate the areas under the curve (AUC). The 1-, 5-, and 10-year ROC curves of the training set, the validation set, and the total set were plotted. According to the median risk score value of the training set, the breast cancer patients were divided into the high- and low-risk groups within the three sets. The Kaplan-Meier log-rank analysis was used to compare the differences in survival between the two groups among the three sets using the “survival” package. The univariate and multivariate Cox regression analyses were performed to validate the relationship between the model and clinicopathological characteristics. Risk curves, point maps, and heat maps were used to observe the survival of patients. The Wilcoxon signed-rank test was used to analyze the differences in the risk score among groups with different clinicopathological characteristics. In addition to the above, R packages also included “survminer”, “pHeatmap”, “ggupbr”, and “complexHeatmap”.

Exploration Of Tumor-infiltrating Immune Cells On Risk Score

Some well-known methods including XCELL, TIMER, QUANTISEQ, MCPcounter, EPIC, CIBERSORT-ABS, and CIBERSORT were utilized to calculate the content of tumor-infiltrating immune cells of the cases downloaded from the TCGA website. Spearman correlation analysis was performed to evaluate the relationship between risk score and tumor-infiltrating immune cells. Wilcoxon signed-rank test was used to compare the differences in tumor-infiltrating immune cells content between high- and low-risk groups of the model. The results were shown as boxplots. These results were analyzed using the R “ggplot2” package, and $P < 0.05$ was considered statistically significant.

Investigation Of Clinical Performance On Risk Score

To investigate the clinical performance of the model, we compared the differences in ICIs expressed between high- and low-risk groups. The results were shown in violin plots which were performed by “limma” and “ggpubr” packages of R. Furthermore, we calculated the half-maximal inhibitory concentration (IC50) of common chemotherapeutic drugs for breast cancer from the TCGA. According to the national comprehensive cancer network (NCCN), chemical drugs such as methotrexate, doxorubicin, gemcitabine, and paclitaxel are commonly used in the treatment of breast cancer. Wilcoxon signed-rank test was used to analyze the differences in the IC50 for the above-mentioned between the high-risk and low-risk groups. The results were shown as boxplots by using “pRRophetic” and “ggplot2” packages of R.

Gene Set Enrichment Analysis

Gene set enrichment analysis (GSEA) version 4.1.0 software was downloaded from the GSEA website (http://www.gsea-msigdb.org/gsea/index.jsp). KEGG pathway enrichment analysis were performed, compared high-risk score group vs. low-risk score group, to investigate the potential molecular mechanisms by which the risk score model might act on tumor progression. FDR < 0.25, NOM p-value < 0.05, and |NES|> 1 were considered significant enrichment.

Declarations

Conflict of Interest

The authors have no conflict of interest to declare.

Author Contributions

SW analyzed the data and wrote the manuscript. JY and LBK helped revise the description of the article. CD provided technical advice. QW provided overall research supervision, revised the manuscript, and funded the work.

Funding

This study was supported by The Academic and Technological Leaders and Reserves Foundation of Anhui Province of the 2019 academic year (2019H230) and The Natural Science Fund for Colleges and Universities in Anhui Province (KJ2021ZD0018).

Acknowledgments

We acknowledge the TCGA database for providing abundant data. We thank all the Perl and R programming package developers

Data Availability Statement

The datasets presented in this study can be found on the TCGA website (https://portal.gdc.cancer.gov).

References

1. SUNG, H.*, et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians.***71**: 209-249. https://doi.org/10.3322/caac.21660 (2021).
2. SIEGEL, R. L., MILLER, K. D., FUCHS, H. E. & JEMAL, A. Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians.***71**: 7-33. https://doi.org/10.3322/caac.21654 (2021).
3. HARBECK, N.*, et al.* Breast cancer. *Nat Rev Dis Primers.***5**: 66. https://doi.org/10.1038/s41572-019-0111-2 (2019).
4. GOODALL, G. J. & WICKRAMASINGHE, V. O. RNA in cancer. *Nature Reviews Cancer.***21**: 22-36. https://doi.org/10.1038/s41568-020-00306-0 (2020).
5. ENGREITZ, J. M., OLLIKAINEN, N. & GUTTMAN, M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol.***17**: 756-770. https://doi.org/10.1038/nrm.2016.126 (2016).
6. ZHENG, H., SIDDHARTH, S., PARIDA, S., WU, X. & SHARMA, D. Tumor Microenvironment: Key Players in Triple Negative Breast Cancer Immunomodulation. *Cancers.***13**: 3357. (2021).
7. EFTEKHARI, R.*, et al.* Study of the tumor microenvironment during breast cancer progression. *Cancer Cell International.***17**: 123. https://doi.org/10.1186/s12935-017-0492-9 (2017).
8. GOFF, S. L. & DANFORTH, D. N. The Role of Immune Cells in Breast Tissue and Immunotherapy for the Treatment of Breast Cancer. *Clin Breast Cancer.***21**: e63-e73. https://doi.org/10.1016/j.clbc.2020.06.011 (2021).
9. GENTLES, A. J.*, et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med.***21**: 938-945. https://doi.org/10.1038/nm.3909 (2015).
10. GANESH, K.*, et al.* Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol.***16**: 361-375. https://doi.org/10.1038/s41575-019-0126-x (2019).
11. PILIPOW, K., DARWICH, A. & LOSURDO, A. T-cell-based breast cancer immunotherapy. *Seminars in Cancer Biology.***72**: 90-101. https://doi.org/10.1016/j.semcancer.2020.05.019 (2021).
12. DENARO, N., MERLANO, M. C. & LO NIGRO, C. Long noncoding RNAs as regulators of cancer immunity. *Mol Oncol.***13**: 61-73. https://doi.org/10.1002/1878-0261.12413 (2019).
13. ATIANAND, M. K., CAFFREY, D. R. & FITZGERALD, K. A. Immunobiology of Long Noncoding RNAs. *Annu Rev Immunol.***35**: 177-198. https://doi.org/10.1146/annurev-immunol-041015-055459 (2017).
14. PEI, X., WANG, X. & LI, H. LncRNA SNHG1 regulates the differentiation of Treg cells and affects the immune escape of breast cancer via regulating miR-448/IDO. *Int J Biol Macromol.***118**: 24-30. https://doi.org/10.1016/j.ijbiomac.2018.06.033 (2018).
15. CHEN, F.*, et al.* Extracellular vesicle-packaged HIF-1alpha-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nat Cell Biol.***21**: 498-510.

https://doi.org/10.1038/s41556-019-0299-0 (2019).

16. LIANG, Y.*, et al.* LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. *Mol Cancer.***19**: 85. https://doi.org/10.1186/s12943-020-01206-5 (2020).
17. ZHANG, M.*, et al.* LncRNA GATA3-AS1 facilitates tumour progression and immune escape in triple-negative breast cancer through destabilization of GATA3 but stabilization of PD-L1. *Cell Prolif.***53**: e12855. https://doi.org/10.1111/cpr.12855 (2020).
18. SHEN, Y., PENG, X. & SHEN, C. Identification and validation of immune-related lncRNA prognostic signature for breast cancer. *Genomics.***112**: 2640-2646. https://doi.org/10.1016/j.ygeno.2020.02.015 (2020).
19. ZHANG, D.*, et al.* Identification of a Novel Glycolysis-Related Gene Signature for Predicting Breast Cancer Survival. *Frontiers in Oncology.***10**: https://doi.org/10.3389/fonc.2020.596087 (2021).
20. LU, G.*, et al.* Long noncoding RNA LINC00511 contributes to breast cancer tumourigenesis and stemness by inducing the miR-185-3p/E2F1/Nanog axis. *Journal of Experimental & Clinical Cancer Research.***37**: https://doi.org/10.1186/s13046-018-0945-6 (2018).
21. SHI, G.*, et al.* Long non-coding RNA LINC00511/miR-150/MMP13 axis promotes breast cancer proliferation, migration and invasion. *Biochim Biophys Acta Mol Basis Dis.***1867**: 165957. https://doi.org/10.1016/j.bbadis.2020.165957 (2021).
22. CHEN, Z., WU, H., ZHANG, Z., LI, G. & LIU, B. LINC00511 accelerated the process of gastric cancer by targeting miR-625-5p/NFIX axis. *Cancer Cell Int.***19**: 351. https://doi.org/10.1186/s12935-019-1070-0 (2019).
23. SUN, C. B.*, et al.* LINC00511 promotes gastric cancer cell growth by acting as a ceRNA. *World J Gastrointest Oncol.***12**: 394-404. https://doi.org/10.4251/wjgo.v12.i4.394 (2020).
24. PENG, X.*, et al.* LINC00511 drives invasive behavior in hepatocellular carcinoma by regulating exosome secretion and invadopodia formation. *J Exp Clin Cancer Res.***40**: 183. https://doi.org/10.1186/s13046-021-01990-y (2021).
25. HU, W. Y.*, et al.* LINC00511 as a ceRNA promotes cell malignant behaviors and correlates with prognosis of hepatocellular carcinoma patients by modulating miR-195/EYA1 axis. *Biomed Pharmacother.***121**: 109642. https://doi.org/10.1016/j.biopha.2019.109642 (2020).
26. HU, Y., ZHANG, Y., DING, M. & XU, R. LncRNA LINC00511 Acts as an Oncogene in Colorectal Cancer via Sponging miR-29c-3p to Upregulate NFIA. *OncoTargets and Therapy.***Volume 13**: 13413-13424. https://doi.org/10.2147/ott.s250377 (2021).
27. SUN, S., XIA, C. & XU, Y. HIF-1α induced lncRNA LINC00511 accelerates the colorectal cancer proliferation through positive feedback loop. *Biomed Pharmacother.***125**: https://doi.org/10.1016/j.biopha.2020.110014 (2020).
28. DONG, L. M.*, et al.* LINC00511/miRNA-143-3p Modulates Apoptosis and Malignant Phenotype of Bladder Carcinoma Cells via PCMT1. *Front Cell Dev Biol.***9**: 650999. https://doi.org/10.3389/fcell.2021.650999 (2021).

29. LI, J., LI, Y., MENG, F., FU, L. & KONG, C. Knockdown of long non-coding RNA linc00511 suppresses proliferation and promotes apoptosis of bladder cancer cells via suppressing Wnt/beta-catenin signaling pathway. *Biosci Rep.***38**: https://doi.org/10.1042/BSR20171701 (2018).
30. DE PALMA, F. D. E.*, et al.* The abundance of the long intergenic non-coding RNA 01087 differentiates between luminal and triple-negative breast cancers and predicts patient outcome. *Pharmacol Res.***161**: 105249. https://doi.org/10.1016/j.phrs.2020.105249 (2020).
31. SHE, J. K., FU, D. N., ZHEN, D., GONG, G. H. & ZHANG, B. LINC01087 is Highly Expressed in Breast Cancer and Regulates the Malignant Behavior of Cancer Cells Through miR-335-5p/Rock1. *Onco Targets Ther.***13**: 9771-9783. https://doi.org/10.2147/OTT.S255994 (2020).
32. GUO, Z., YAO, L. & GUO, A. Clinical and biological impact of LINC02544 expression in breast cancer after neoadjuvant chemotherapy. *Eur Rev Med Pharmacol Sci.***24**: 10573-10585. (2020).
33. WU, J.*, et al.* LINC01152 upregulates MAML2 expression to modulate the progression of glioblastoma multiforme via Notch signaling pathway. *Cell Death Dis.***12**: 115. https://doi.org/10.1038/s41419-020-03163-9 (2021).
34. CHEN, T.*, et al.* HBx-related long non-coding RNA 01152 promotes cell proliferation and survival by IL-23 in hepatocellular carcinoma. *Biomed Pharmacother.***115**: 108877. https://doi.org/10.1016/j.biopha.2019.108877 (2019).
35. ACHA-SAGREDO, A.*, et al.* Long non-coding RNA dysregulation is a frequent event in non-small cell lung carcinoma pathogenesis. *Br J Cancer.***122**: 1050-1058. https://doi.org/10.1038/s41416-020-0742-9 (2020).
36. CHE, H., CHE, Y., ZHANG, Z. & LU, Q. Long Non-Coding RNA LINC01929 Accelerates Progression of Oral Squamous Cell Carcinoma by Targeting the miR-137-3p/FOXC1 Axis. *Front Oncol.***11**: 657876. https://doi.org/10.3389/fonc.2021.657876 (2021).
37. ARAN, D., HU, Z. & BUTTE, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.***18**: 220. https://doi.org/10.1186/s13059-017-1349-1 (2017).
38. LI, T.*, et al.* TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res.***48**: W509-W514. https://doi.org/10.1093/nar/gkaa407 (2020).
39. LI, T.*, et al.* TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.***77**: e108-e110. https://doi.org/10.1158/0008-5472.CAN-17-0307 (2017).
40. FINOTELLO, F.*, et al.* Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.***11**: 34. https://doi.org/10.1186/s13073-019-0638-6 (2019).
41. PLATTNER, C., FINOTELLO, F. & RIEDER, D. Deconvoluting tumor-infiltrating immune cells from RNA-seq data using quanTIseq. *Methods Enzymol.***636**: 261-285. https://doi.org/10.1016/bs.mie.2019.05.056 (2020).
42. BECHT, E.*, et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.***17**: 218. (2016).

43. DIENSTMANN, R.*, et al.* Relative contribution of clinicopathological variables, genomic markers, transcriptomic subtyping and microenvironment features for outcome prediction in stage II/III colorectal cancer. *Ann Oncol.***30**: 1622-1629. (2019).
44. ROSENTHAL, R.*, et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature.***567**: 479-485. https://doi.org/10.1038/s41586-019-1032-7 (2019).
45. RACLE, J. & GFELLER, D. EPIC: A Tool to Estimate the Proportions of Different Cell Types from Bulk Gene Expression Data. *Methods Mol Biol.***2120**: 233-248. (2020).
46. TAMMINGA, M.*, et al.* Immune microenvironment composition in non-small cell lung cancer and its association with survival. *Clin Transl Immunology.***9**: e1142. https://doi.org/10.1002/cti2.1142 (2020).
47. CHEN, B., KHODADOUST, M. S., LIU, C. L., NEWMAN, A. M. & ALIZADEH, A. A. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol.***1711**: 243-259. https://doi.org/10.1007/978-1-4939-7493-1_12 (2018).
48. KIM, Y.*, et al.* Novel deep learning-based survival prediction for oral cancer by analyzing tumor-infiltrating lymphocyte profiles through CIBERSORT. *Oncoimmunology.***10**: 1904573. https://doi.org/10.1080/2162402X.2021.1904573 (2021).
49. PARDOLL, D. M. Immunology beats cancer: a blueprint for successful translation. *Nature Immunology.***13**: 1129-1132. https://doi.org/10.1038/ni.2392 (2012).
50. HONG, W.*, et al.* Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. *Mol Ther Nucleic Acids.***22**: 937-947. https://doi.org/10.1016/j.omtn.2020.10.002 (2020).
51. PING, S., WANG, S., HE, J. & CHEN, J. Identification and Validation of Immune-Related lncRNA Signature as a Prognostic Model for Skin Cutaneous Melanoma. *Pharmgenomics Pers Med.***14**: 667-681. https://doi.org/10.2147/PGPM.S310299 (2021).

Figures

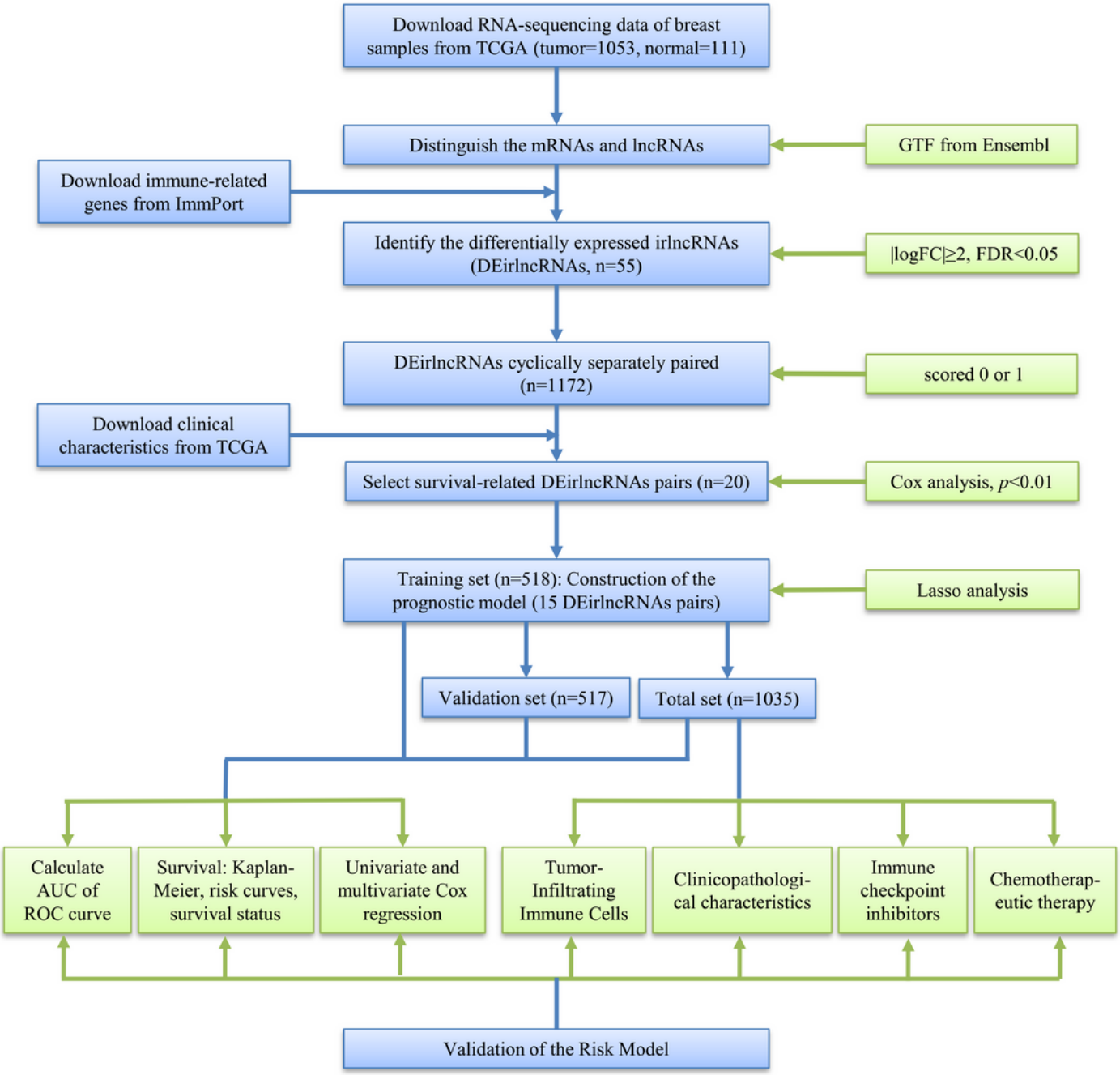

Figure 1

The flow chart of this study.

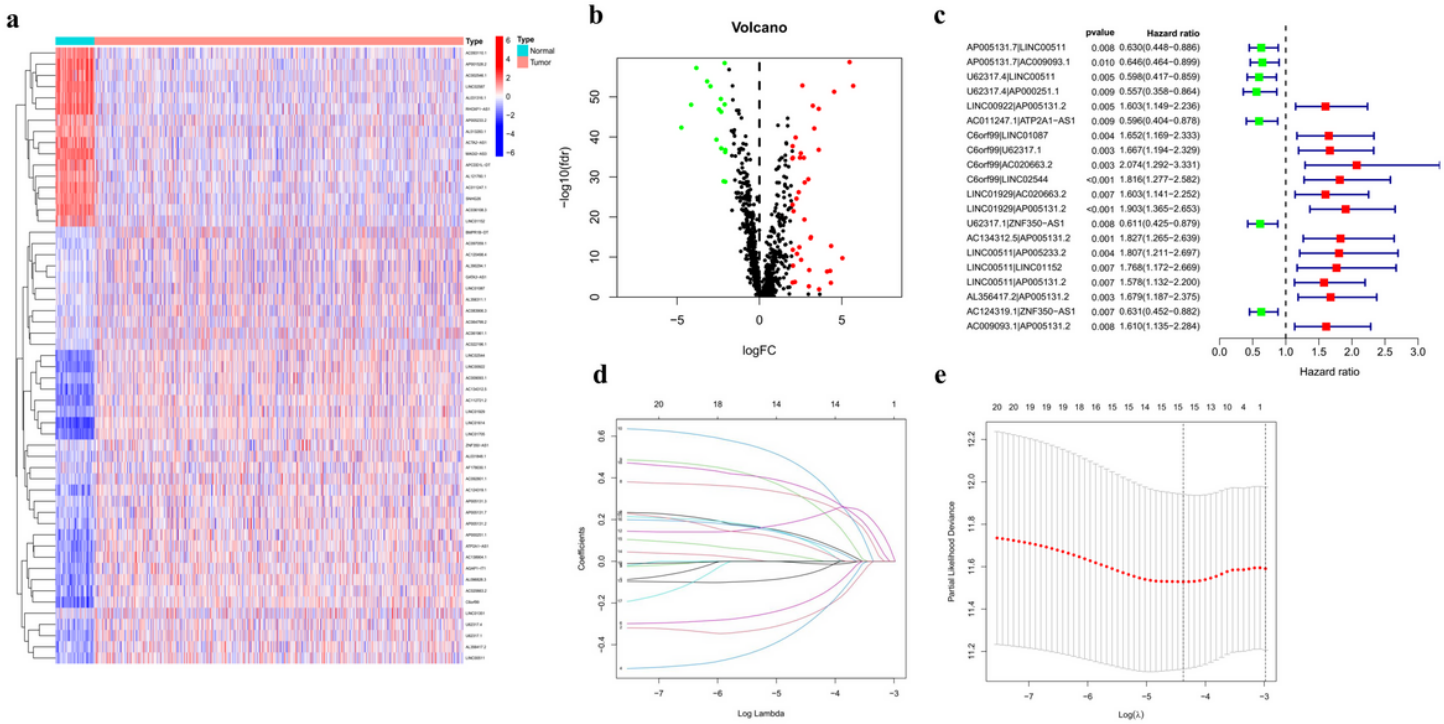

Figure 2

Construction of a risk model using DEirlncRNA pairs. **a** The DEirlncRNAs showedby heatmap. **b** The DEirlncRNAs showedby volcano, red dots represented the upregulated while green dots represented the downregulated DEirlncRNAs. **c** Forest plots showing HR ofDEirlncRNAspairs obtained by multivariate Cox regression analysis.D EirlncRNAs: the differentially expressed immune-related lncRNAs.

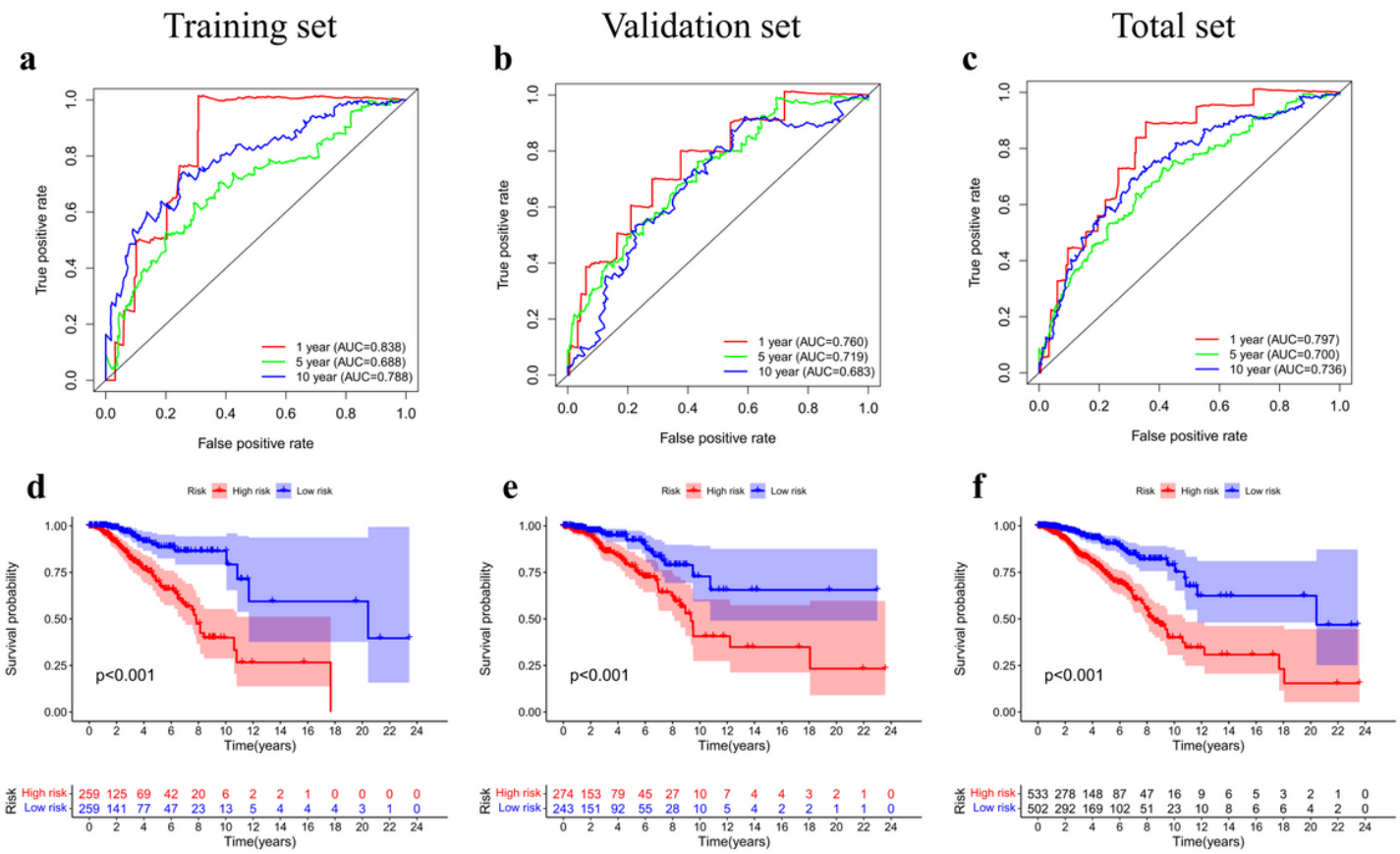

Figure 3

Validation of the risk model. **a-c** The 1-, 5-, and 10-year survival receiver operating characteristic (ROC) curves of the training, validation, and total sets. **d-f** Survival curves showed that patients in the high-risk group had worse overall survival than those in the low-risk group

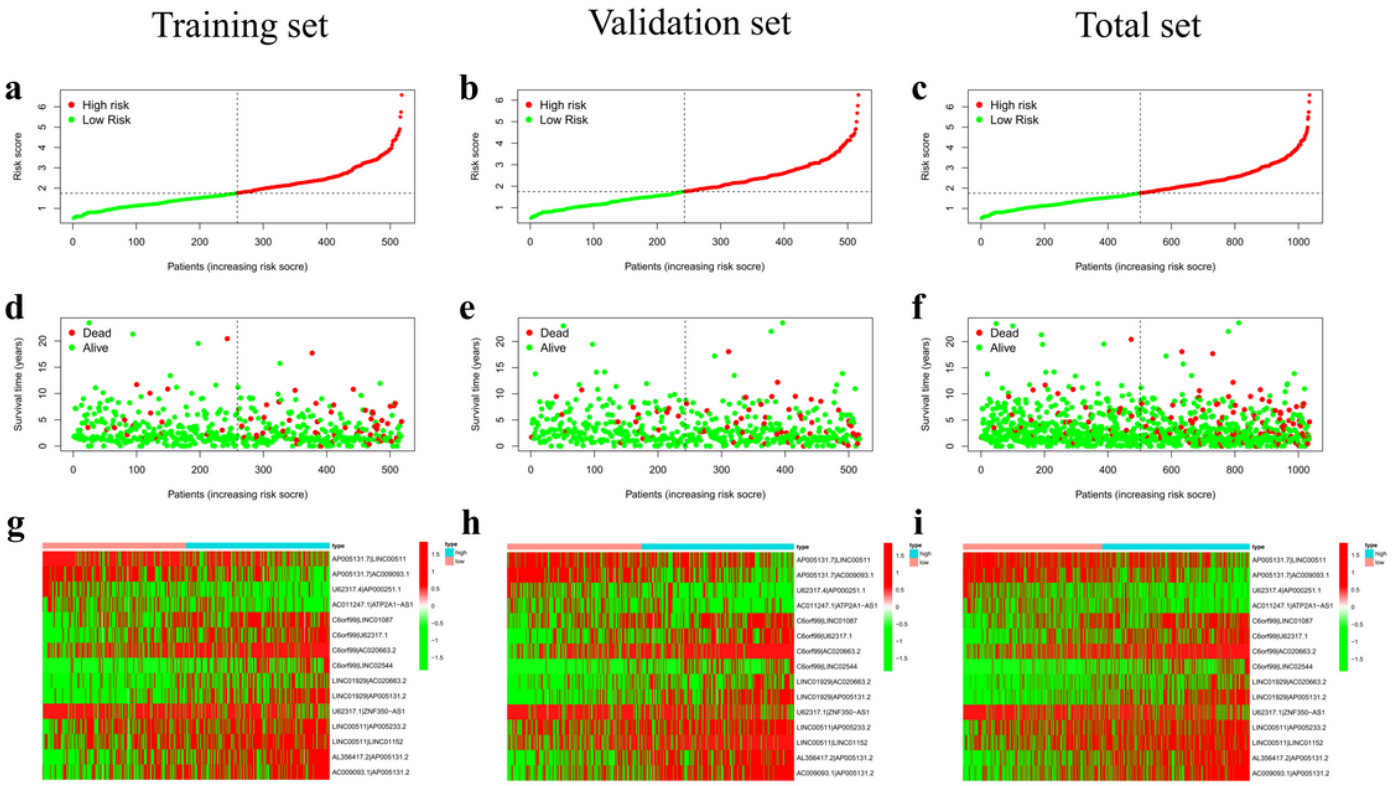

Figure 4

Risk curves showed the survival outcome of each breast cancer patient. The signature risk score distribution in the training (**a**), validation (**b**) and total set (**c**). The scatter plot of the sample survival overview in the training (**d**), validation (**e**) and total set (**f**), the green and red dots respectively represent survival and death. Heatmap showed the DEirlncRNAs pairs value of each patient in the training (**g**), validation (**h**) and total set (**i**).

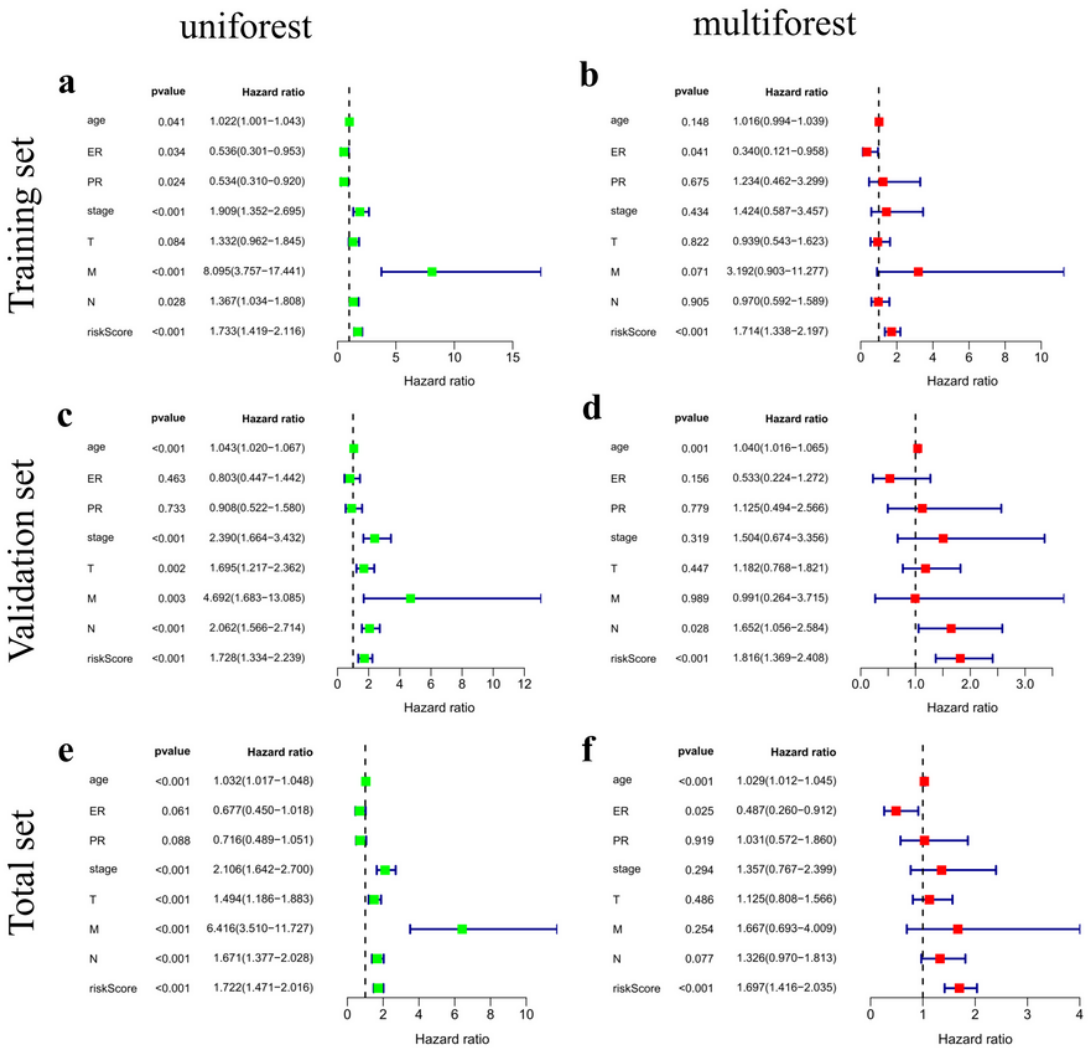

Figure 5

The univariate and multivariate Cox regression analysis for evaluating the independent prognostic value of the risk model and clinicopathologic parameters in the training (**a** and**b**), validation (**c**and**d**) and total set (**e** and **f**). *P*< 0.05 was considered statistically significant.

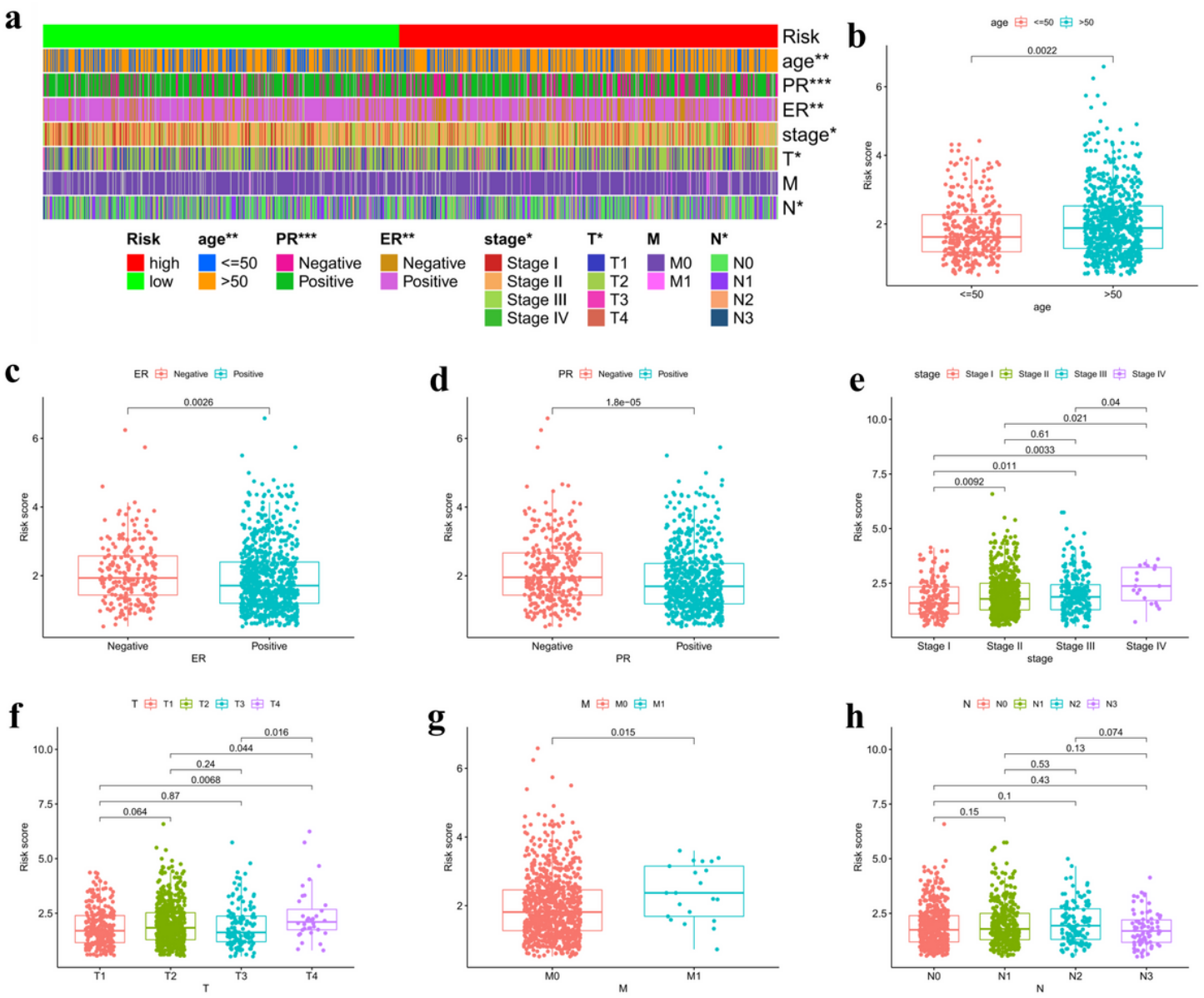

Figure 6

Evaluation of clinicopathological characteristics by the risk model. **a** The strip chart showed the differences in clinicopathological characteristicsbetween high- and low-risk groups. **b-h**Scatter plots showed the differences in risk scores among different clinical feature groupings. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

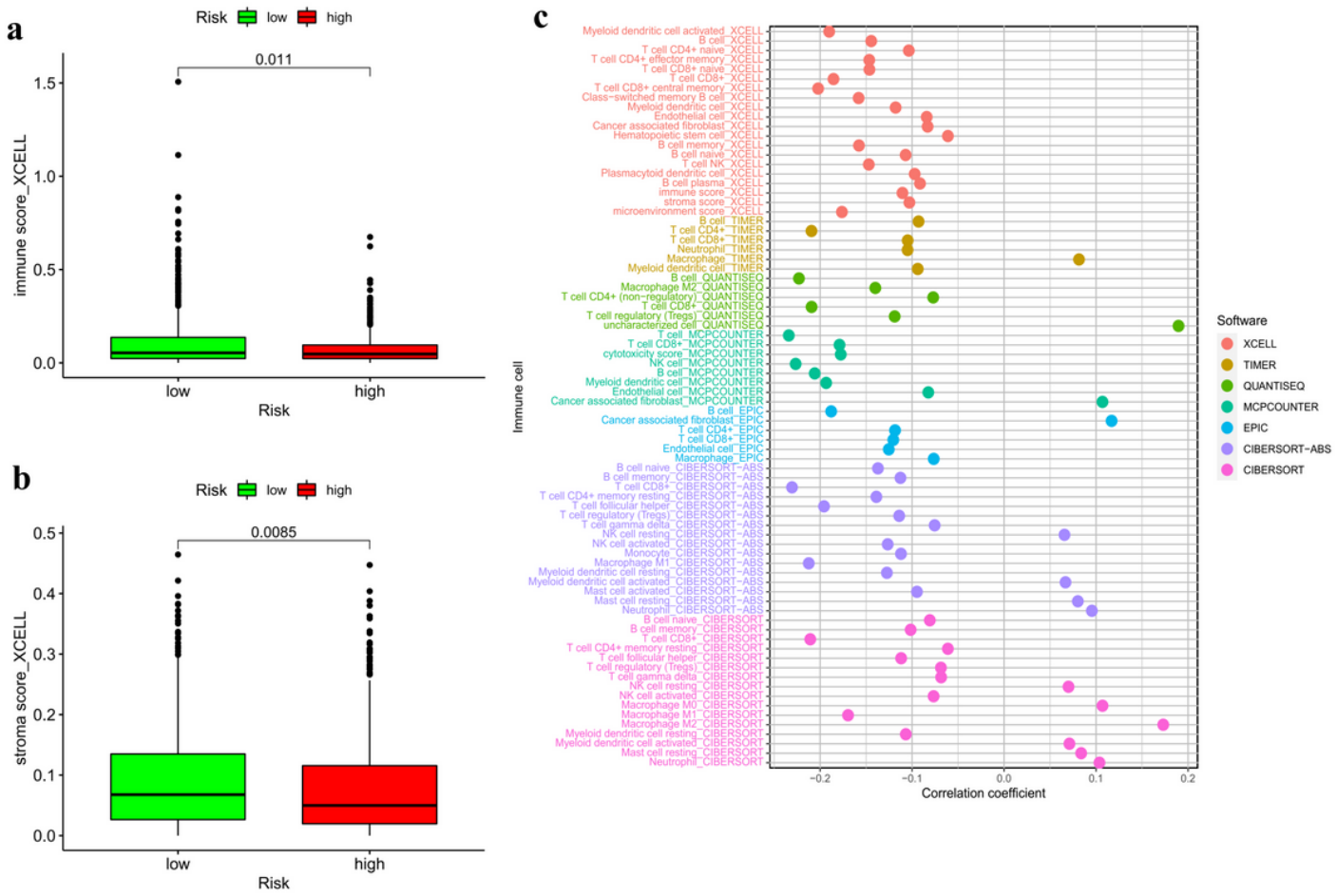

Figure 7

Estimation of tumor-infiltrating immune cells by the risk model. Patients in the high-risk group were more negatively associated with tumor-infiltrating immune cells (**a**) and stroma score (**b**). The bubble chart (**c**) showed the detailed correlation between different tumor-infiltrating immune cells.

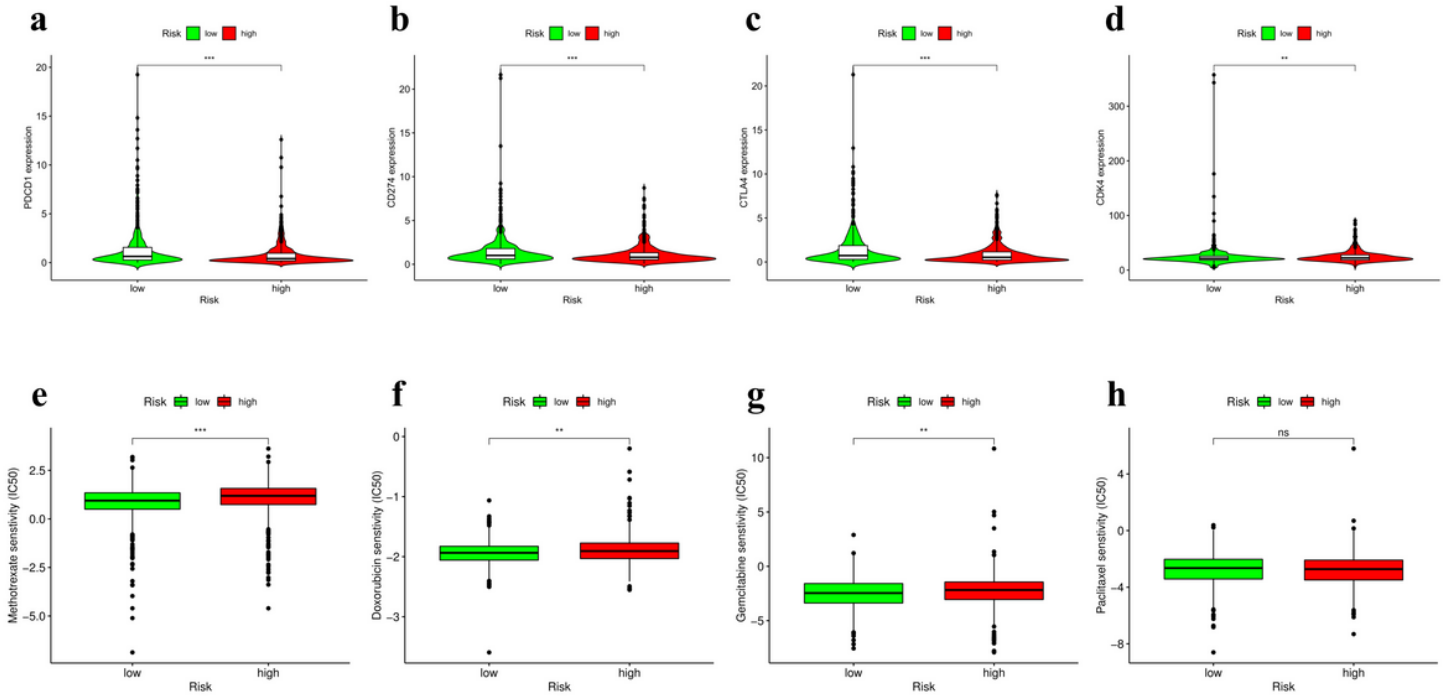

Figure 8

Verification of the correlation between the risk model and therapeutic targets in breast cancer. Violin plots showed that the expression of ICIs biomarkers including PCDC1 (**a**), CD274 (**b**), CTLA4 (**c**), and CDK4 (**d**) were significantly lower in the high-risk group. Box plots showed that chemotherapeutic agents, such as methotrexate (**e**), doxorubicin (**f**) and gemcitabine (**g**) were associated with the high-risk group while paclitaxel (**h**) showed no significant difference between the two groups. *P< 0.05, **P< 0.01, ***P< 0.001, ns represents no significance.

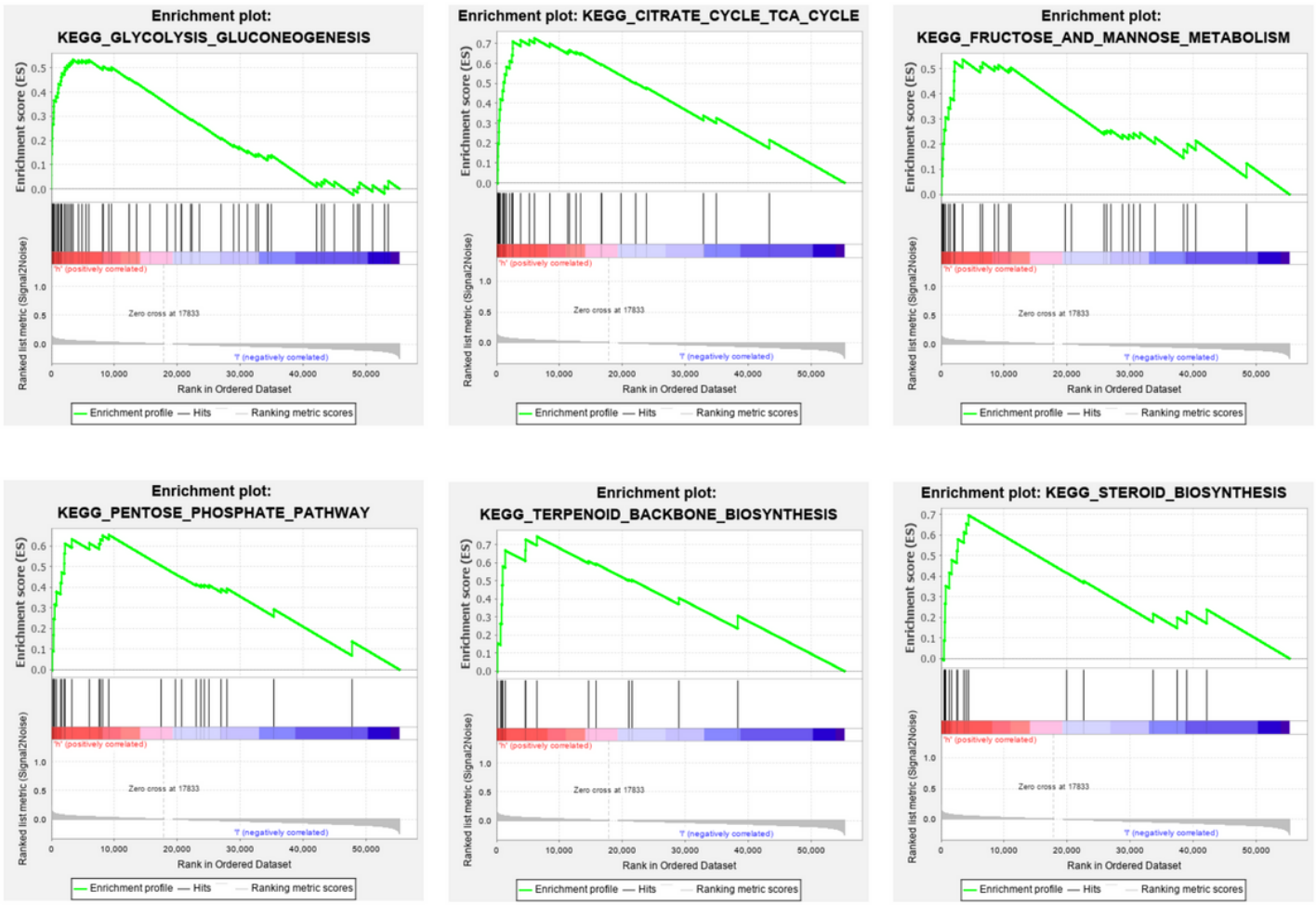

Figure 9

Gene set enrichment analysis of enriched signaling pathways.

Supplementary Files

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

- FigureS1.pdf
- SupplementaryFigureS1legend.docx
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx