Am J Cancer Res 2023;13(9):3875-3897
www.ajcr.us /ISSN:2156-6976/ajcr0151514

Original Article

A prognostic model related to necrotizing apoptosis of breast cancer based on biorthogonal constrained depth semi-supervised nonnegative matrix decomposition and single-cell sequencing analysis

Yuan Gao[1*], Qinghua Huang[2*], Yuling Qin[3], Xianhui Bao[4], You Pan[5], Jianlan Mo[6], Shipeng Ning[5]

[1]Department of Head and Neck Radiotherapy, Harbin Medical University Cancer Hospital, Harbin 150000, Heilongjiang, China; [2]Department of Breast Surgery, Wuzhou Red Cross Hospital, Wuzhou 543000, Guangxi, China; [3]Department of Clinical Laboratory, Guangxi Medical University Cancer Hospital, Nanning 530000, Guangxi, China; [4]Department of Neurology, Harbin The First Hospital, Harbin 150000, Heilongjiang, China; [5]Department of Breast Surgery, Guangxi Medical University Cancer Hospital, Nanning 530000, Guangxi, China; [6]Department of Anesthesiology, The Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, Nanning 530000, Guangxi, China. [*]Equal contributors.

Received May 31, 2023; Accepted July 31, 2023; Epub September 15, 2023; Published September 30, 2023

Abstract: Breast cancer (BC) is one of the most common malignant tumours in women, and its prognosis is poor. The prognosis of BC patients can be improved by immunotherapy. However, due to the heterogeneity of BC, the identification of new biomarkers is urgently needed to improve the prognosis of BC patients. Necrotic apoptosis has been shown to play an essential role in many cancers. First, this study proposed a novel clustering algorithm called biorthogonal constrained depth semisupervised nonnegative matrix factorization (DO-DSNMF). The DO-DSNMF algorithm added multilayer nonlinear transformation to the coefficient matrix obtained after decomposition, which was used to mine the nonlinear relationship between samples. In addition, we also added orthogonal constraints on the basis matrix and coefficient matrix to reduce the influence of redundant features and samples on the results. We applied the DO-DSNMF algorithm and analysed the differences in survival and immunity between the subtypes. Then, we used prognosis analysis to construct the prognosis model. Finally, we analysed single cells using single-cell sequencing (scRNA-seq) data from the GSE75688 dataset in the GEO database. We identified two BC subtypes based on the BC transcriptome data in the TCGA database. Immune infiltration analysis showed that the necrotizing apoptosis-related genes of BC were related to various immune cells and immune functions. Necrotizing apoptosis was found to play a role in BC progression and immunity. The role of prognosis-related NRGs in BC was also verified by cell experiments. This study proposed a novel clustering algorithm to analyse BC subtypes and constructed an NRG prognostic model for BC. The prognosis and immune landscape of BC patients were evaluated by this model. The cell experiment supported its role in BC, which provides a potential therapeutic target for the treatment of BC.

Keywords: Breast cancer, scRNA-seq, transcriptome, nonnegative matrix decomposition, prognostic model

Introduction

Breast cancer (BC) is a heterogeneous tumour with high incidence and rapid progression that mostly occurs in women [1]. BC has become the fifth leading cause of cancer deaths worldwide. In recent years, the prognosis of BC patients has been improved through surgery, chemotherapy, radiotherapy, and immunotherapy. However, due to the heterogeneity of BC, the prognosis of patients is quite different [2]. Therefore, it is still necessary to explore more markers for different BC subtypes to improve the prognosis of BC patients.

Apoptosis is a normal cell self-death process, also known as programmed cell death. It plays an important role in maintaining tissue struc-

ture and functional balance. Apoptosis has been confirmed to be related to the occurrence and development of many diseases. Nathalie Droin et al. elaborated that apoptotic mechanisms play a key role in haematopoietic cell homeostasis [3]. Tumour necrosis factor is a pleiotropic cytokine that triggers proinflammatory effects through NF-κB-related pathways or apoptosis through the activation of caspase-8 [4]. The activation of the mitochondria-dependent apoptotic pathway contributes to the loss of substantia nigra dense dopaminergic neurons in experimental mouse models of Parkinson's disease [5]. Necroptosis is a kind of apoptosis [6]. It is an essential mechanism of cell death that plays a vital role in the occurrence and development of immune system diseases and tumour diseases [7]. Necrosis can be regarded as a substitute for apoptosis, which is mainly mediated by ROP1, RIP3, and MLKL. Cell necrosis activates the human immune response [8]. As an alternative process of apoptosis, necrotizing apoptosis induces the immune system and kills tumour cells when apoptosis is blocked [9]. In addition, RIPK3, the critical mediator of the necrotizing apoptosis pathway, is downregulated in many cancer types, and this downregulation is related to an increase in tumour invasion and chemical resistance [10]. Necrotic apoptosis is also closely associated with immunity. Immunotherapy for BC is a therapeutic method to control and eliminate tumours by restoring the normal antitumour immune response of the body. Therefore, it is necessary to explore the important role of necrotizing apoptosis-related genes in the immune system during the occurrence and development of BC.

Different subtypes of BC may have significant differences in immune landscapes and treatments, so this study aims to examine the typing of BC. The NMF algorithm is a well-known, conventional data dimensionality reduction technique that is frequently utilized in bioinformatics research and analysis [11, 12]. NMF is particularly widely used in disease. Liu et al. analysed the RNA methylome using the NMF algorithm and revealed the comethylation pattern induced by potential enzyme regulators of the epitranscriptome [13]. Amy L Sherborne et al. used NMF to identify mutational signatures in tumours that were not separated by genetic background or histology [14]. Cassio P de Campos et al. found accurate subgroups with individual molecular and clinical characteristics in multiple datasets [15]. Renaud Gaujoux et al. found that the use of marker genes improved the accuracy of gene expression deconvolution using NMF [16]. Mara L Hartsperger proposed a novel fuzzy k-partite graph partitioning algorithm that can effectively explore biological networks [17]. NMF is typically used to categorize diseases. Wang et al. divided patients into two categories using NMF and data from RNA, single nucleotide polymorphisms (SNPs), and copy number variations (CNVs) associated with N6-methyladenosine (m6A) in thyroid cancer (PTC) samples in the cancer genome atlas (TCGA). The characteristics of m6A are significant in predicting the disease-free survival (DFS) of PTC patients [18]. Some scholars also applied the NMF algorithm to the clustering of hepatocellular carcinoma (HCC) samples. These scholars found three subtypes of HCC, and used weighted gene coexpression network analysis and single-cell regulatory network reasoning and clustering to determine a transcription factor that can upregulate immunosuppressive genes in the subtype group with poor prognosis [19]. Jiang et al. employed the NMF method to distinguish between several lung adenocarcinoma (LUAD) subtypes using the RNA data of lung adenocarcinoma (LUAD) samples in the TCGA dataset [20]. However, for transcriptome data with more highly correlated properties, the conventional NMF technique might only be partially appropriate. Trigeorgis and others proposed a deep semisupervised joint nonnegative matrix factorization (deep semi-NMF) algorithm, which is further innovated on the basis of a semi-NMF algorithm. Specifically, the features of the potential space can be fully captured by decomposing the base matrix or coefficient matrix in multiple layers, which is beneficial to the feature selection of the algorithm [21]. However, the deep semi-NMF algorithm does not consider the influence of redundant samples or features on the results. In addition, the nonlinear transformation of features is not considered in the process of layer-by-layer decomposition. Therefore, this study proposes a biorthogonal constrained depth semisupervised nonnegative matrix factorization (DODSNMF) algorithm, which adds orthogonal constraints to the base matrix and coefficient matrix on the basis of the deep semi-NMF algo-

rithm and can effectively reduce the influence of redundant samples or features on the results. Moreover, the DO-DSNMF algorithm uses an activation function to nonlinearly transform the coefficient matrix. We used 1094 breast cancer samples in the TCGA database to carry out the experiment. The results show that the contour coefficient of the DO-DSNMF algorithm is better than that of other competitive algorithms.

The function of necrotizing apoptosis-related genes (NRGs) in BC patients was thoroughly examined in this research from a bioinformatics standpoint. To be more precise, we first investigated the variations in the prognosis of various subtypes of samples using the DO-DSNMF algorithm to categorize the NRGs in the transcriptome data of BC samples in the TCGA database (https://portal.gdc.cancer.gov/). By using Cox and Lasso analysis to separate the samples into high-risk and low-risk groups, prognosis-related NRGs were discovered. A variety of immune analyses indicated that there were great differences between the two groups in the infiltration abundance and immune function of immune cells. Immunity is closely related to necrotizing apoptosis, so we further analysed the important role of genes related to necrotizing apoptosis in BC single-cell sequencing. To explore the role of necrotic apoptosis in breast cancer and its correlation with the immune microenvironment, BC single-cell sequencing (scRNA-seq) data in the GSE75688 dataset in the GEO database (https://www.ncbi.nlm.nih.gov/geo/) were analysed. Specifically, the dataset is clustered and annotated, and all cell groups are divided into high-risk cells and low-risk cells by NRGs. Enrichment analysis and cell communication analysis of high-risk cell groups were carried out. The important role of significant pathways in BC samples in necrotizing apoptosis and immunity was analysed in detail. Finally, the NRGs related to prognosis were verified by experiments.

Method

Differential expression analysis

In this study, 130 NRGs were collected from previous studies. Based on the Limma algorithm, the differences between BC in the TCGA database and NRGs in transcriptome data of the healthy control group were analysed, and 129 differentially expressed NRGs (DENRGs) with *P* values less than 0.05 were retained.

CNV mutation analysis

We explored the CNV incidence of DENRGs and used the "RCircos" package to map their positions on 23 chromosomes.

Deep semisupervised nonnegative matrix factorization (deep semi-NMF)

The deep semi-NMF algorithm is a multilayer feature extraction method that can effectively capture the multilayer feature transformation of datasets. Its objective function is as follows.

$$\begin{aligned} &\min_{Z_m, H_M} \left\| X - W_1(W_2(\cdots(W_M H_M))) \right\|_F^2 \\ &\text{s.t. } H_m \geq 0, m = 1, \ldots, M, \end{aligned} \tag{1}$$

Where $X \in R^{p \times n}$ represents the characteristic matrix of the data, p stands for the number of features, n represents the number of samples, W_1, W_2 ..., W_M represent the m fundamental matrices obtained by layer-by-layer decomposition, and H_M is the coefficient matrix of the m-th layer obtained by decomposition. In the decomposition process, the algorithm ensures the nonnegativity of H_M.

Biorthogonal constrained depth semisupervised nonnegative matrix factorization (DO-DSNMF)

The deep semi-NMF algorithm does not consider the influence of redundant features or samples on the results and cannot capture nonlinear characteristics. In addition, the proposed algorithm has better anti-noise performance. The DO-DSNMF algorithm can effectively solve this problem, and its objective function is as follows.

$$\begin{aligned} &\min \left\| X - Wg(Z_1 g(Z_2 \ldots g(Z_N H_N))) \right\|_F^2 + \lambda_1 \left\| S \right\|_F + \lambda_2 \left\| WW^T - I \right\|_F + \lambda_3 \left\| H_0 H_0^T - I \right\|_F \\ &\text{s.t. } H_{n-1} = g(Z_n H_n), H_0 \ldots H_{N-1} \geq 0, \\ &S \in \{W, Z_1, \ldots, Z_N, H_N\}, n = 1, \ldots, N \end{aligned} \tag{2}$$

Where $g(\cdot)$ represents the activation function of "$\cdot$", $Z_1, \ldots, Z_N$ are N potential connection matrices introduced in this study, $W \in R^{p \times k}$ is the base matrix obtained by decomposition, $H_0 \in R^{k \times n}$ is the coefficient matrix of the first layer, $Z_1 \in R^{k \times k-1}$ is the potential connection matrix of the first layer, and $k < min\{p, n\}$ needs to be guaranteed in the decomposition process. Composing the

coefficient matrix $\overline{H_0}$ layer by layer can help the algorithm perform multiple nonlinear transformations on the samples. The $\overline{H_0}$ obtained by the final iteration of the algorithm stores the low-dimensional representation of all sample features. In this study, the spectral clustering algorithm (implemented by python's scikit-learn package) was used to set the corresponding cluster number to obtain the final result.

Construction of the DENRG-related prognostic risk model

In this study, based on the TCGA cohort of BC patients, an NRG-related prognostic model was constructed. First, univariate Cox regression was used to screen prognosis-related DENRGs. Next, we used the glmnet package for Lasso-Cox regression to construct the prognostic model of BC. In addition, GSE20685 of the GEO cohort was used as the validation set for the prognosis model.

Verification of the DENRG-related prognostic model

To verify that the prognostic model constructed in this study can effectively distinguish samples with different prognosis outcomes, this was confirmed in the TCGA and GEO queues. Specifically, we divided BC patients into a high-risk group and a low-risk group according to the median sample risk value. We analysed the KM survival with the "Survival" package and "survminer" package and obtained the difference in overall survival (OS) between the two risk groups. Finally, the area under the ROC (AUROC) of the 1-year, 3-year, and 5-year survival was calculated to evaluate the prediction performance of the prognosis model.

Construction of nomogram model

To evaluate the survival status of BC patients through prognosis-related DENRGs and clinical data, a nomogram was constructed by clinical factors and prognosis-related DENRGs. In addition, this study evaluated whether the predicted and actual survival times of patients were consistent through the calibration curve.

GSEA and ssGSEA

For the purpose of exploring the KEGG pathways of BC patients in the high-risk and low-risk groups, the KEGG gene set from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb/) was downloaded, and GSEA was implemented with the clusterProfiler package. In addition, to explore the difference between high and low immune cell functions, the GSVA package was used to realize ssGSEA, which can quantify the infiltration level of 16 immune cell types and 13 immune-related functions in each sample.

Immune landscape of BC samples

In this study, the CIBERSORT algorithm was used to evaluate the difference in immune cell infiltration abundance between BC samples in high-risk and low-risk groups.

Data analysis of scRNA-seq

We downloaded the scRNA-seq data of BC from the GSE75688 dataset of the GEO database. The dataset contains a total of 563 cells with single-cell RNA sequencing data for primary and metastatic breast cancer. For the single-cell sequencing data, the data for 549 single cells were retained. Of these, 441 cells were primary breast cancer cells, and 108 cells were metastatic breast cancer cells. First, we controlled the quality of the data and kept the cells with less than 20% mitochondrial genes and the cells with more than 200 genes and an expression between 200 and 15000. These samples were processed by the Harmony package to remove the batch effect. In this study, the percentage of necrotic genes in each cell was obtained by inputting NRGs with the "PercentageFeatureSet" function, and the cells were divided into two groups according to the proportion of median necrotic genes: high and low expression of NRGs. Cell clustering was carried out by the PCA dimension reduction algorithm and Umap dimension reduction algorithm of the seurat package (resolusion was set to 0.5). The cell types were annotated by the singleR package. Finally, the percentage of NRGs in each cell was obtained by importing NRGs through the "PercentageFeatureSet" function. We also examined the communication between different cell groups using the cellchat algorithm.

Results

Expression and mutation analysis of NRGs in BC transcriptome data

Figure 1 shows the overall flow chart of this study. The NRGs of the BC group and healthy

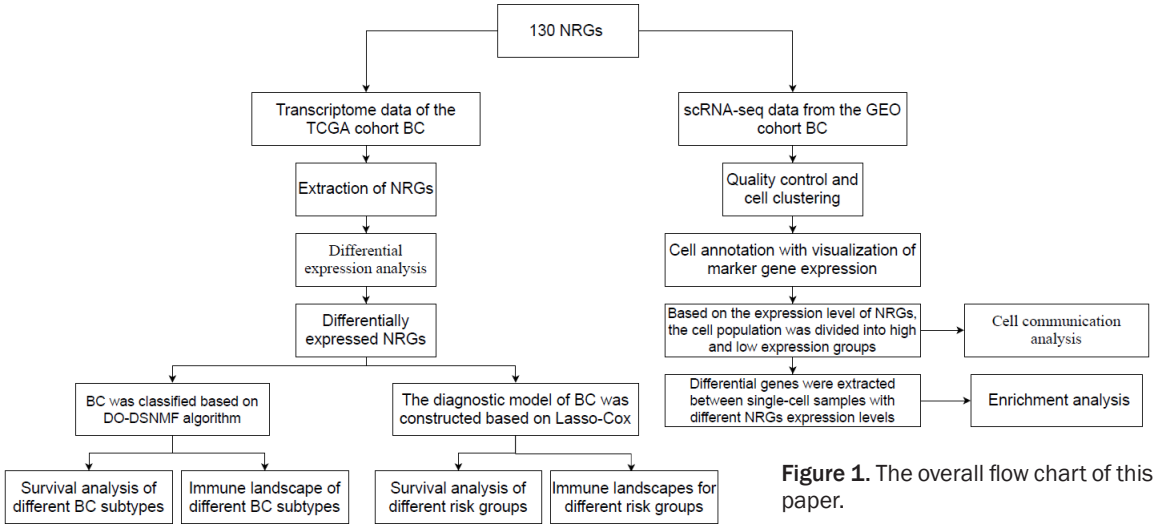


Figure 1. The overall flow chart of this paper.

control group in the TCGA database were first extracted, and then the differentially expressed NRGs (DENRGs) were obtained by the limma algorithm. We drew a differential expression heatmap of DENRGs between the BC group and the healthy control group (**Figure 2A**). DENRGs related to prognosis were obtained by univariate Cox analysis, and their copy number variations (CNVs) were analysed (**Figure 2B**). **Figure 2C** shows the correlation network diagram between the expression of DENRGs and the prognosis of breast cancer.

Parameter selection of the DO-DSNMF algorithm

First, the maximum number of iterations of the algorithm was set to 20000. Before selecting parameters, we chose the number of layers for dimensionality reduction. Specifically, the more layers of dimensionality reduction there are, the better it is at capturing the nonlinear feature changes of data. According to a previous paper [13], the number of layers of dimensionality reduction was set to 4. The hyperparameters involved in this study include λ_1, λ_2 and λ_3. These super parameters were selected in the range of [0.001, 0.01, 0.1, 1]. In addition, to determine the appropriate activation function, we selected three activation functions, namely, the sigmoid activation function, hyperbolic tangent activation function (tanh activation function), and rectifier linear unit activation function (ReLU activation function) (**Figure 3**).

In this study, the smallest first set of parameter combinations corresponding to the tanh activation function that minimized loss for subsequent analysis was selected. After deciding the parameters, we chose the number of layers and the dimension reduction of each layer. Specifically, the more layers of dimensionality reduction there are, the better it is at capturing the nonlinear feature changes of data. According to previous papers [11], the number of layers of dimensionality reduction was set to 4. Next, the best parameters were fixed, and the loss of the algorithm under different dimensionality reduction conditions was compared. Because the minimum dimension in this study was the number of genes (129), the dimension of the first layer was less than 129. The dimension of the subsequent layers was reduced to 1/2, 1/3 and 1/4 of the original (**Table 1**).

Clustering results of the DO-DSNMF algorithm

In this study, the contour coefficient was used to evaluate the results of different combinations of dimensionality reduction and clustering layers. The profile factor is defined as follows.

$$s(i) = \frac{y(i) - x(i)}{max\{x(i), y(i)\}} \tag{3}$$

Where $x(i)$ is the average distance from sample i to other samples in the same cluster. The smaller $x(i)$ is, the more sample i should be clustered into this cluster. $x(i)$ is called the

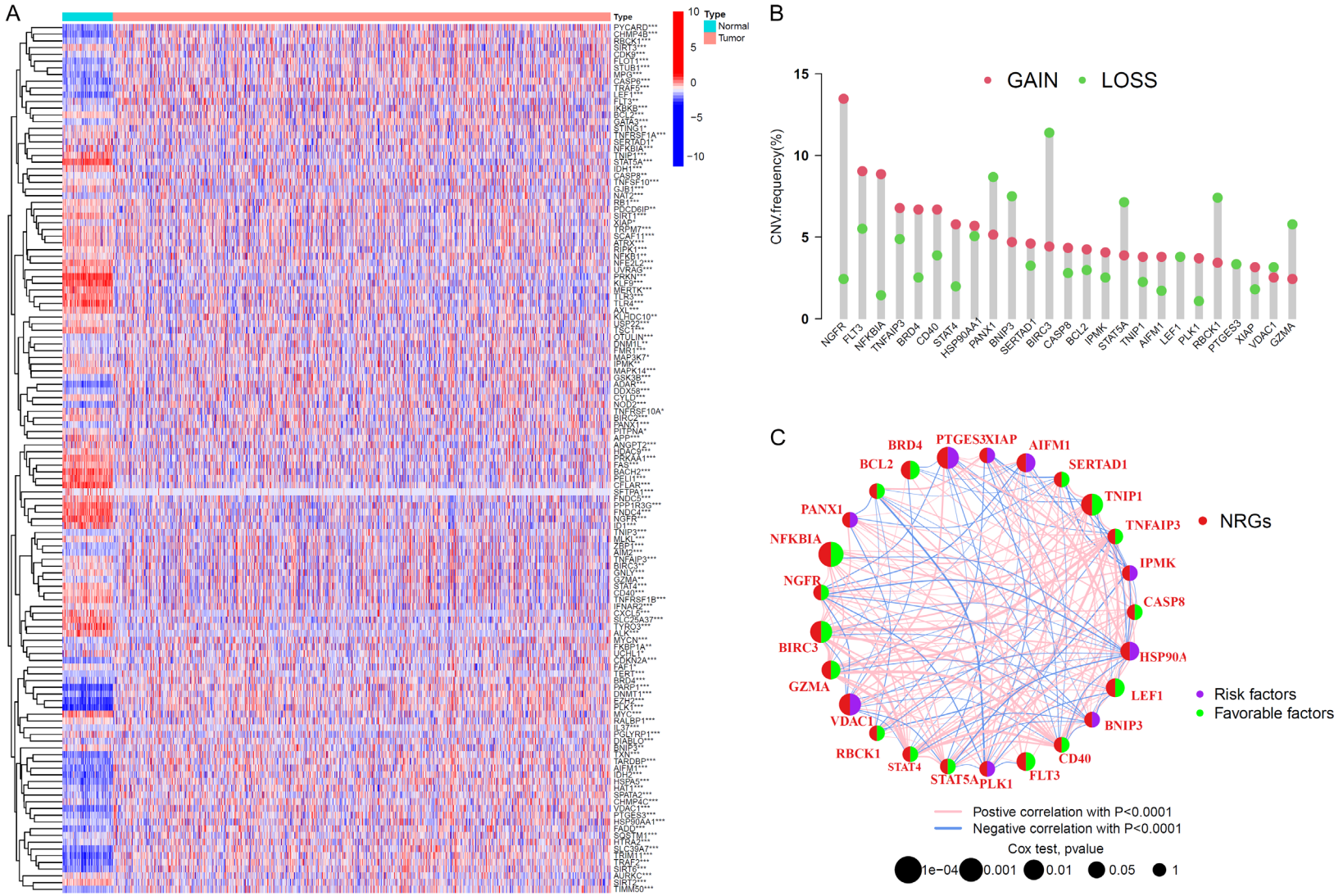


Figure 2. NRGs-related transcriptome data analysis. A. Differential expression heat map of NRGs. B. CNV mutation analysis. C. Correlation network diagram between the expression of DENRGs and breast cancer prognosis.

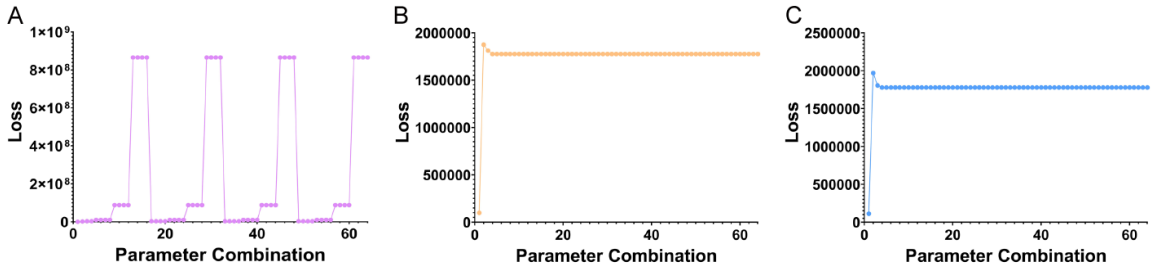


Figure 3. Parameter selection results. A. The line chart of parameter selection using the sigmoid activation function. B. The line chart of parameter selection using the tanh activation function. C. The line chart of parameter selection using the ReLU activation function.

Table 1. Loss corresponding to different dimensionality reduction

Index	Dimension reduction of the first layer	Dimension reduction of the second layer	Dimension reduction of the third layer	Dimension reduction of the fourth layer	Loss
1	120	60	20	5	113120
2	110	55	18	5	98489
3	100	50	17	4	107253
4	90	45	15	4	103316
5	80	40	13	3	102336
6	70	35	12	3	95819
7	60	30	10	3	91270
8	50	25	8	2	88430
9	40	20	7	2	81583

Table 2. Number of clusters and corresponding contour coefficients under different conditions

Index	Cluster number was 2	Cluster number was 3	Cluster number was 4	Cluster number was 5
1	0.1407	0.1337	0.1574	0.1102
2	0.8281	0.0238	-0.0479	0.1608
3	0.0819	-0.1187	-0.1129	-0.0880
4	0.1333	-0.1313	-0.0724	-0.1970
5	0.2187	-0.0792	-0.1216	-0.1572
6	0.0888	-0.1839	-0.2023	-0.2262
7	0.2704	-0.0788	-0.1230	-0.1255
8	0.1446	-0.1607	-0.0931	-0.1248
9	-0.1813	-0.3342	-0.2854	-0.1903

intracluster dissimilarity of sample i. $y(i)$ is the average distance from sample i to all samples in some other cluster. The value of the contour coefficient is between -1 and 1. The closer the value is to 1, the better the clustering effect will be.

In this study, the contour coefficients of nine cases of dimension reduction were calculated. Specifically, we compared the contour coefficients with the number of clusters between 2 and 5 in nine cases.

Table 2 shows that the highest and second highest profile coefficients were obtained in cases 2 and 7, respectively. However, case 2 determined that most sample points belong to the same category, which could lead to a higher contour coefficient. Therefore, we ruled out this situation and selected the second-highest result of case 7 for subsequent analysis. The change in the objective function in loss and Equation (2) in this case was determined (**Figure 4**). It was obvious that the algorithm could reach the convergence state at a faster

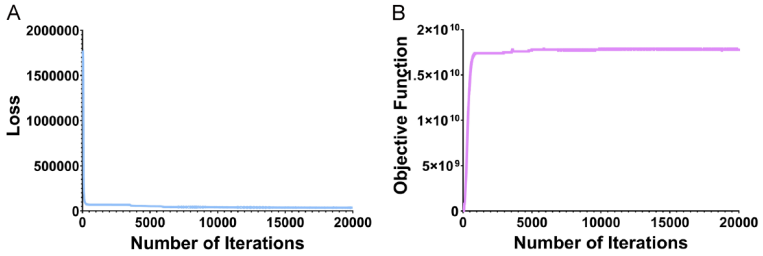


Figure 4. The line chart of the changing trend of loss and objective function under the optimal parameter combination. A. The change of Loss as the number of iterations increases. B. The change of the value of the objective function with the increase of iteration times.

speed. To confirm the superiority of the algorithm, the algorithm performance was compared with deep semi-NMF and NMF algorithms on real datasets and simulated datasets (Supplementary Materials).

Immune landscape among BC subtypes

To explore the immune landscape of different subtypes of BC samples, this study analysed the differences in immune cell infiltration abundance (**Figure 5A**), immune function (**Figure 5B**), immune-related gene expression (**Figure 5D**) and immune inspection sites (**Figure 5E**) between the two subtypes. In addition, heatmaps showed immune cells and immune functions that differed significantly between the two subtypes (**Figure 5C**). The ssGSEA results showed significant differences in immune cells and immune function between subtype B and subtype A, and the immune function and immune cell scores of subtype B were generally higher than those of subtype A. This indicated that necrotic apoptosis-related genes are related to the immune microenvironment in breast cancer. In addition, we also conducted GSVA on the two subtypes to explore the differences in their enrichment pathways (**Figure 5F**). As shown in **Figure 5F**, the differentially expressed genes between subtype A and subtype B were mainly involved in the chemokine signalling pathway, natural killer cell-mediated cytotoxicity, glycosylphosphatidylinositol (GPI) anchor biosynthesis, vasopressin-regulated water reabsorption and other pathways.

BC prognostic model construction

To explore the prognostic significance of NRGs in breast cancer, we used Lasso-Cox regression to analyse the prognostic genes obtained by univariate Cox regression. Specifically, we present the results of the Lasso-Cox regression in **Figure 6A**, **6B**. The results of the KM survival analysis between the two subtypes of the TCGA and GEO queues are provided in **Figure 6C** and **6E**, respectively. ROC analysis was used to evaluate the prediction accuracy of the prognostic models. **Figure 6D** shows the AUC curves of 1-year, 3-year, and 5-year survival in the TCGA queue, in which the AUCs of 1-year, 3-year, and 5-year survival were 0.637, 0.701, and 0.695, respectively. **Figure 6F** shows the AUC curves for 1-year, 3-year, and 5-year survival in the GEO queue, in which the AUCs of 1-year, 3-year, and 5-year survival were 0.693, 0.664, and 0.681, respectively. For the TCGA cohort, **Figure 7A-C** show the distribution of risk score, the scatter plot between survival status and risk score and the expression heatmap of prognosis-related NRGs between the two groups. For the GEO cohort, **Figure 7D-F** show the distribution of risk score, the scatter plot between survival status and risk score and the expression heatmap of prognosis-related NRGs between the two groups. In addition, **Figure 7G** shows the expression heatmap for different clinical factors. **Figure 7H** shows the correlations among the risk score, two subtypes, and survival status.

Finally, univariate and multivariate Cox regression analyses were constructed on the basis of clinical factors and risk scores. **Figure 8A** and **8B** show the forest maps of single-factor Cox and multifactor Cox regression, respectively. **Figure 8C** and **8F** show the nomogram models and DCA curves constructed by multivariate Cox regression. The calibration curve of **Figure 8D** shows the difference between the predicted and actual values. The results showed that the accuracy of the nomogram in predicting the prognosis of breast cancer was similar to that of the actual model. **Figure 8E** shows the prediction results of the risk model for 1-year, 3-year, and 5-year survival, with AUCs of 0.813, 0.765, and 0.760, respectively. We performed univariate Cox regression analysis of NRGs and identified multiple NRGs that were significantly associated with prognosis. Figure S1 shows the

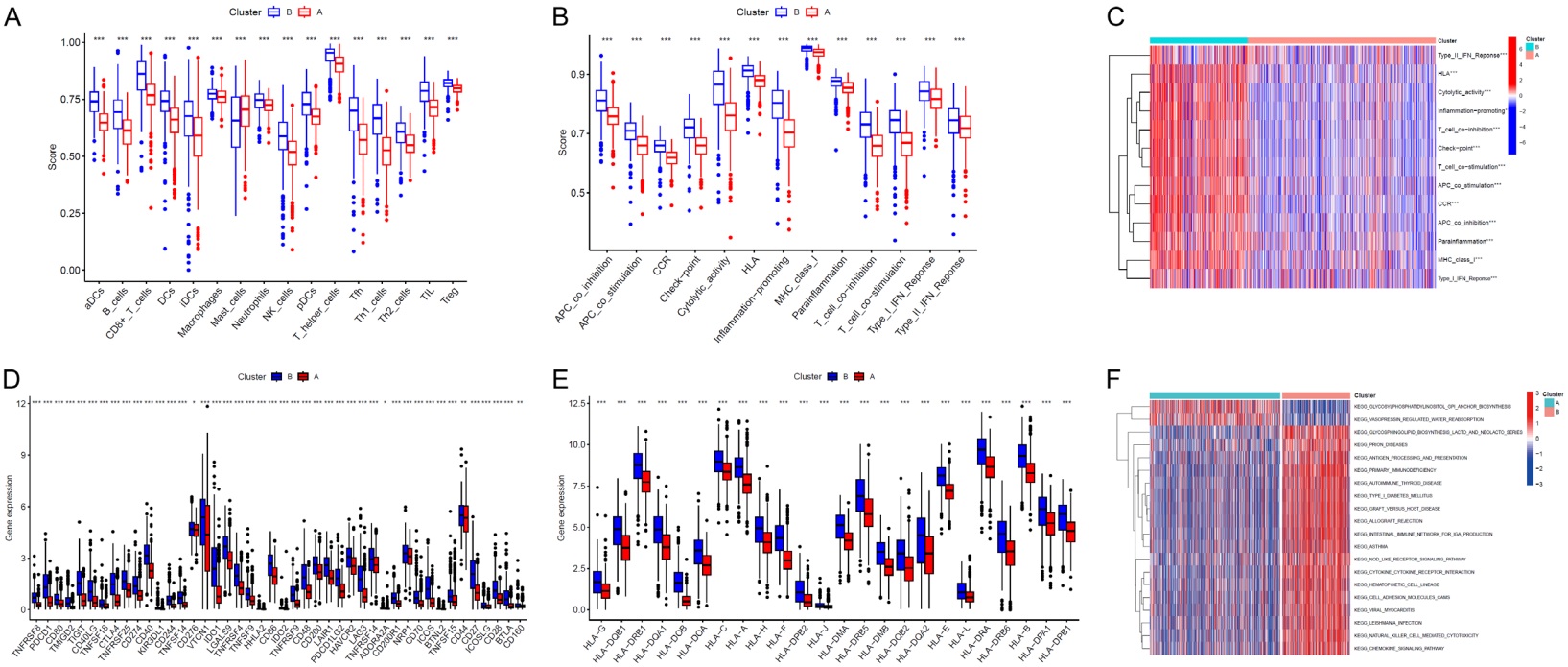


Figure 5. Immune landscape of different subtype samples. A. Box diagram of infiltration abundance of immune cells in different subtypes of samples. B. Box diagram of different subtypes of samples with different immune functions. C. Abundance thermogram of immune cells and immune functions between two subtypes. D. Box-diagram of the difference in immune gene expression in different subtypes of samples. E. Difference box diagram of HLA-related genes between subtypes. F. Thermogram of GSVA analysis between subtypes.

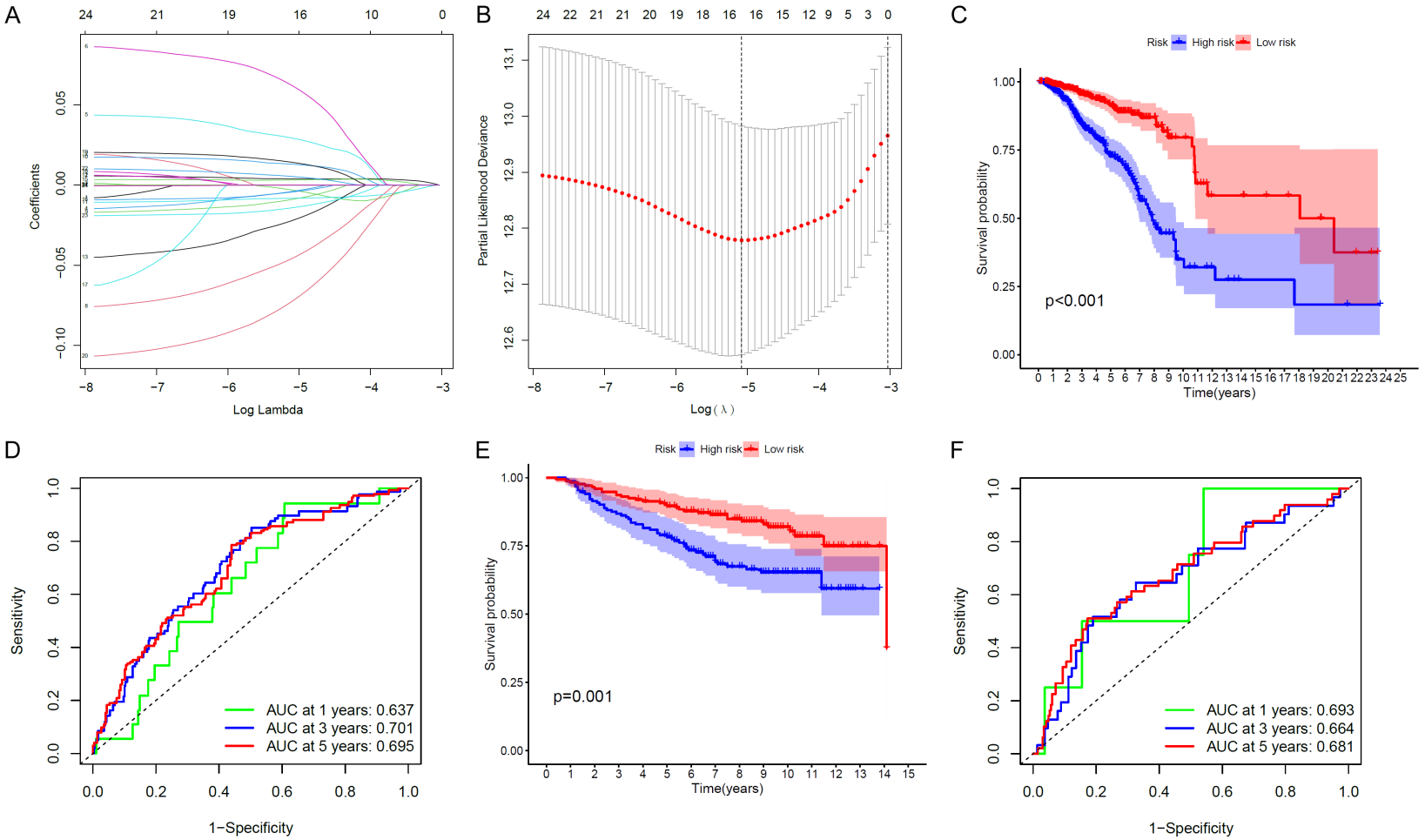


Figure 6. Construction of prognostic models. A. LASSO regression of prognosis-related DENRGs. B. Cross-validation results of LASSO regression of prognosis-related DENRGs. C. KM survival curves obtained from KM survival analysis of two subtypes of TCGA cohort. D. ROC curves drawn by 1-year, 3-year, and 5-year OS risk models in TCGA queues. E. KM survival curves obtained from KM survival analysis of two subtypes of GEO cohort. F. ROC curves drawn by 1-year, 3-year, and 5-year OS risk models in GEO queues.

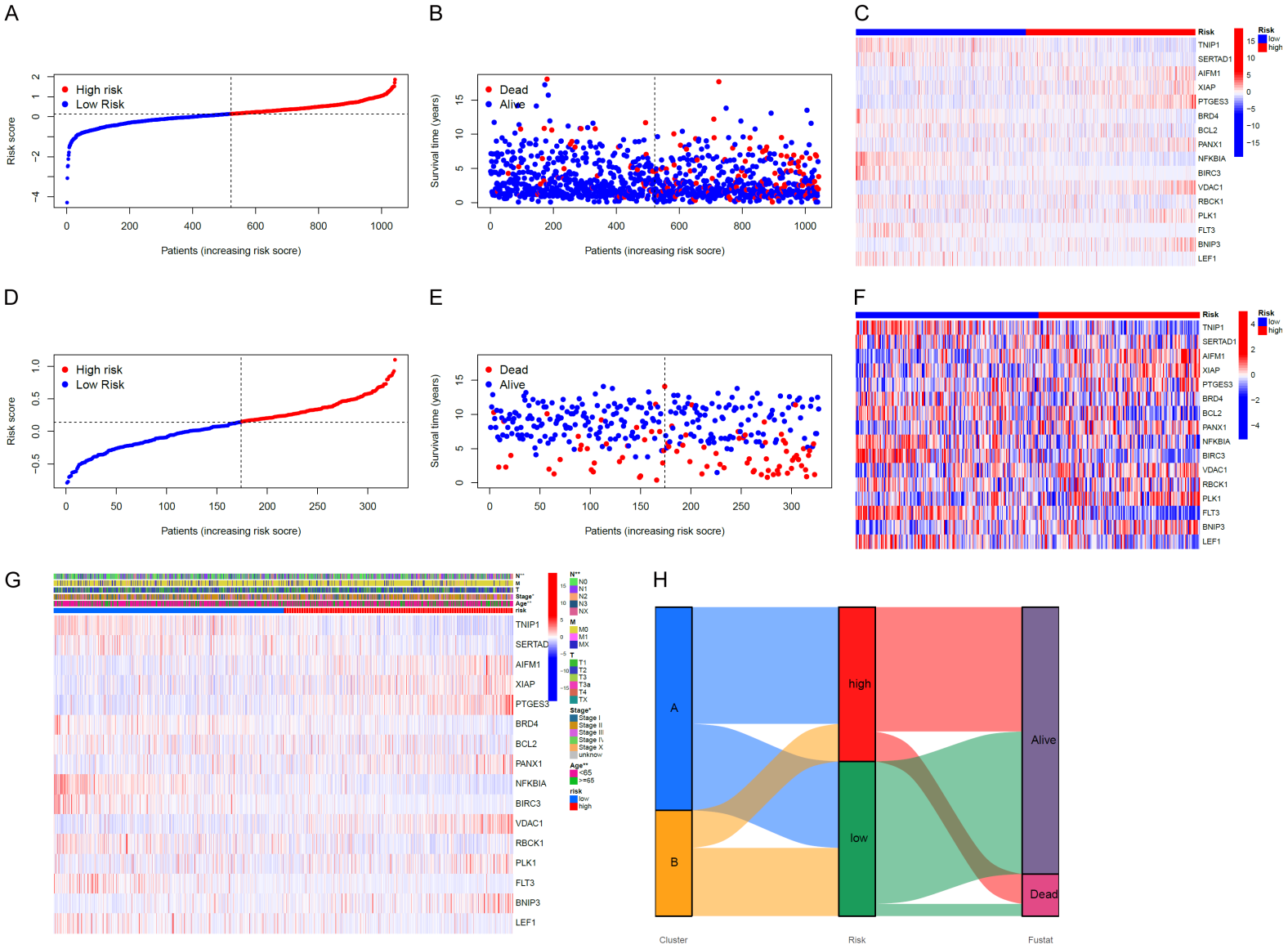


Figure 7. The clinical significance of prognosis related NRGs. (A and D) were the risk score distributions of two risk groups in TCGA and GEO queues, respectively. (B and E) were scatter plots of survival status and risk scores in TCGA and GEO queues, respectively. (C and F) were the expression heat maps of prognosis-related NRGs in TCGA and GEO cohorts, respectively. (G) Expression heat map of prognosis-related NRGs under different clinical factors (T, M, N, stage, age, risk score). (H) Sangji diagram which showed the correlation between risk score, two subtypes, and survival status.

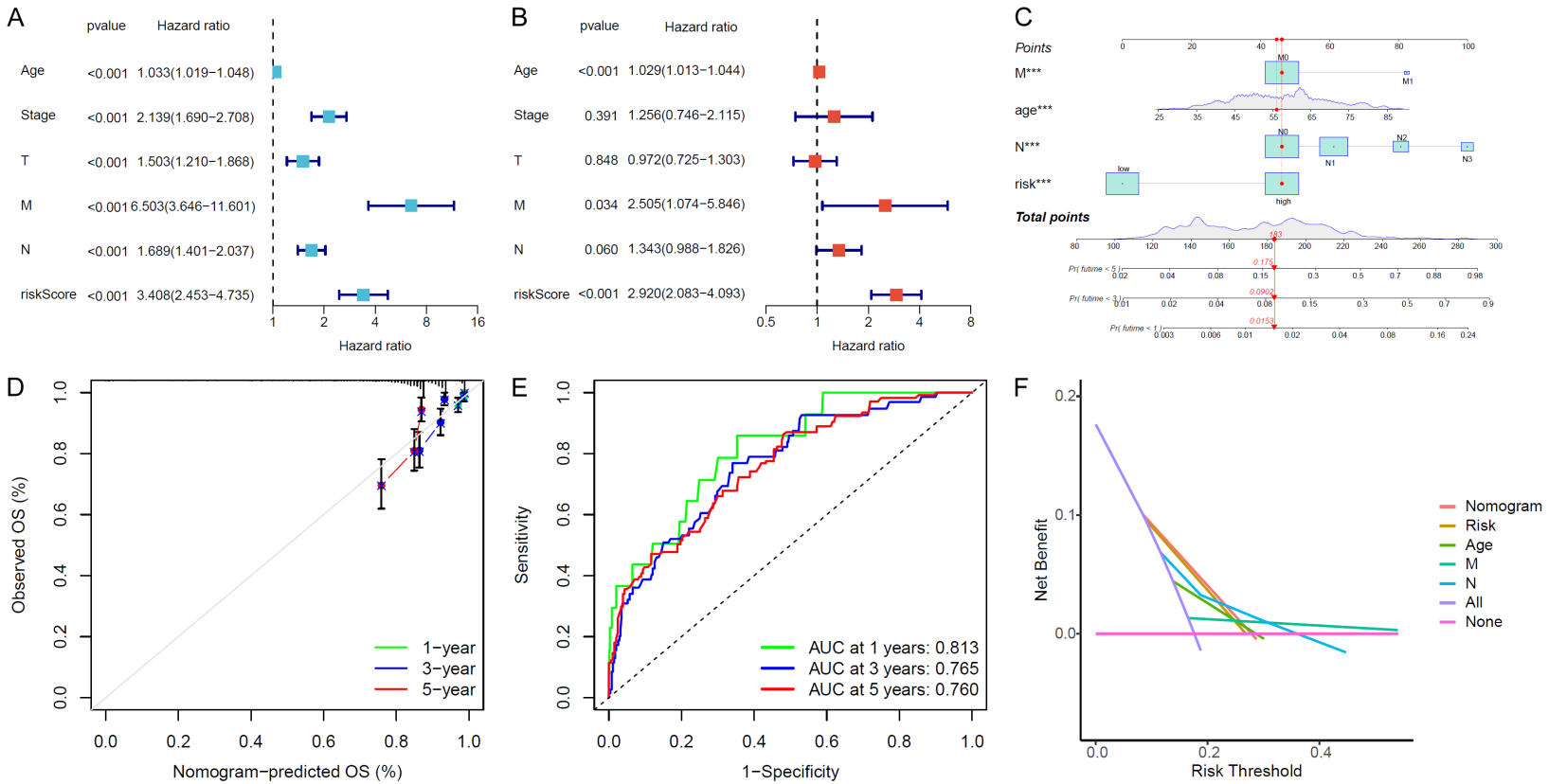


Figure 8. Independent prognostic analysis and construction of nomogram. A. Forest map obtained by single factor Cox regression analysis. B. Forest map obtained by multivariate Cox regression analysis. C. Nomogram model constructed by multivariate Cox regression. D. Calibration curve of multivariate Cox regression. E. ROC curves drawn by 1-, 3-, and 5-year risk models. F. DCA curve of multivariate Cox regression.

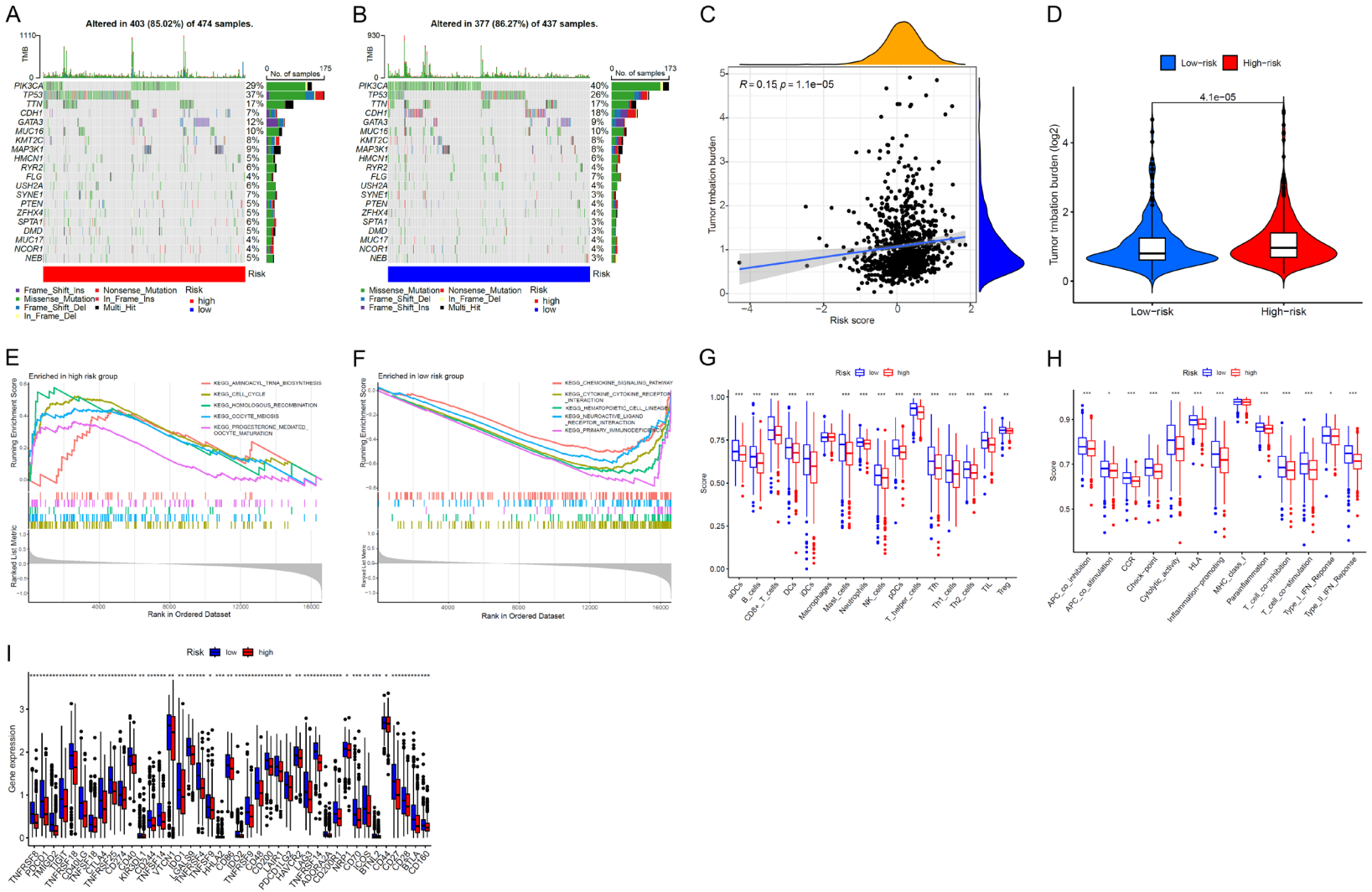


Figure 9. TMB analysis, GSEA analysis, and immune landscape of prognosis model. (A and B) were waterfall maps obtained from TMB analysis of the high-risk group and low-risk group, respectively. (C) Scatter plot between tumor mutation coincidence and risk score. (D) Tide scores of high-risk and low-risk groups. (E and F) Gave the first five paths of high-risk and low-risk groups, respectively. (G-I) Offered the difference box charts of immune cell infiltration, immune function, and immune examination sites in high-risk and low-risk groups, respectively.

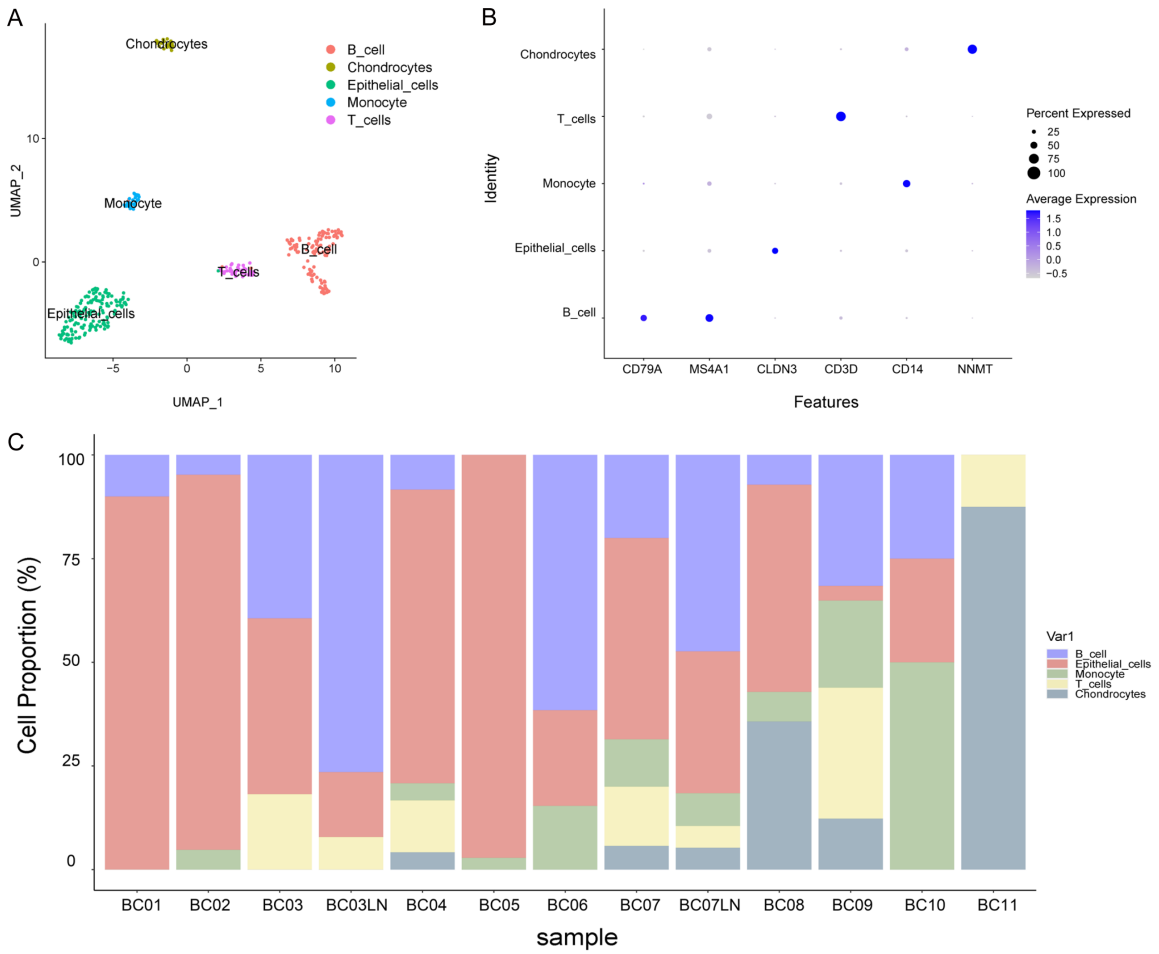


Figure 10. Clustering results of single cell data. A. Umap map of cell clusters. The horizontal and vertical coordinates represent the two classifications obtained by umap algorithm dimensionality reduction, respectively. B. Bubble plot of marker gene expression in cell populations. C. Bar graph of the proportion of cell types in each sample. The horizontal and vertical coordinates represent the sample and the proportion of the number of cells of different cell types, respectively.

KM survival curves of these NRGs in the high and low expression groups.

Enrichment analysis and immune landscape of the prognostic model

The TMB analysis results of the high-risk and low-risk groups are shown in **Figure 9A**, **9B**. The TMB of genes in the low-risk group was more frequent than that in the high-risk group. **Figure 9C** shows the scatter plot of the correlation between the risk score and TMB. As shown in **Figure 9D**, there were significant differences in Tide scores between the high-risk and low-risk groups. **Figure 9E**, **9F** show the GSEA results of the high-risk and low-risk groups, respectively. In addition, we also explored the immune landscape of the prognosis model. There were also significant differences in immune cell infiltration, immune function, and immune examination sites between the two risk groups (**Figure 9G-I**). As shown in **Figure 9G** and **9H**, the immune function and immune cell scores of patients in the low-risk group were generally higher than those in the high-risk group, which indicates that the downregulation of immune function or immune cells in breast cancer patients might lead to a poor prognosis.

Correlation analysis of scRNA-seq data

We performed a detailed analysis of the BC scRNA-seq dataset GSE75688. Then, the KNN clustering algorithm was applied to cluster the dataset, and the singleR algorithm was used to annotate it. **Figure 10A** shows the results of

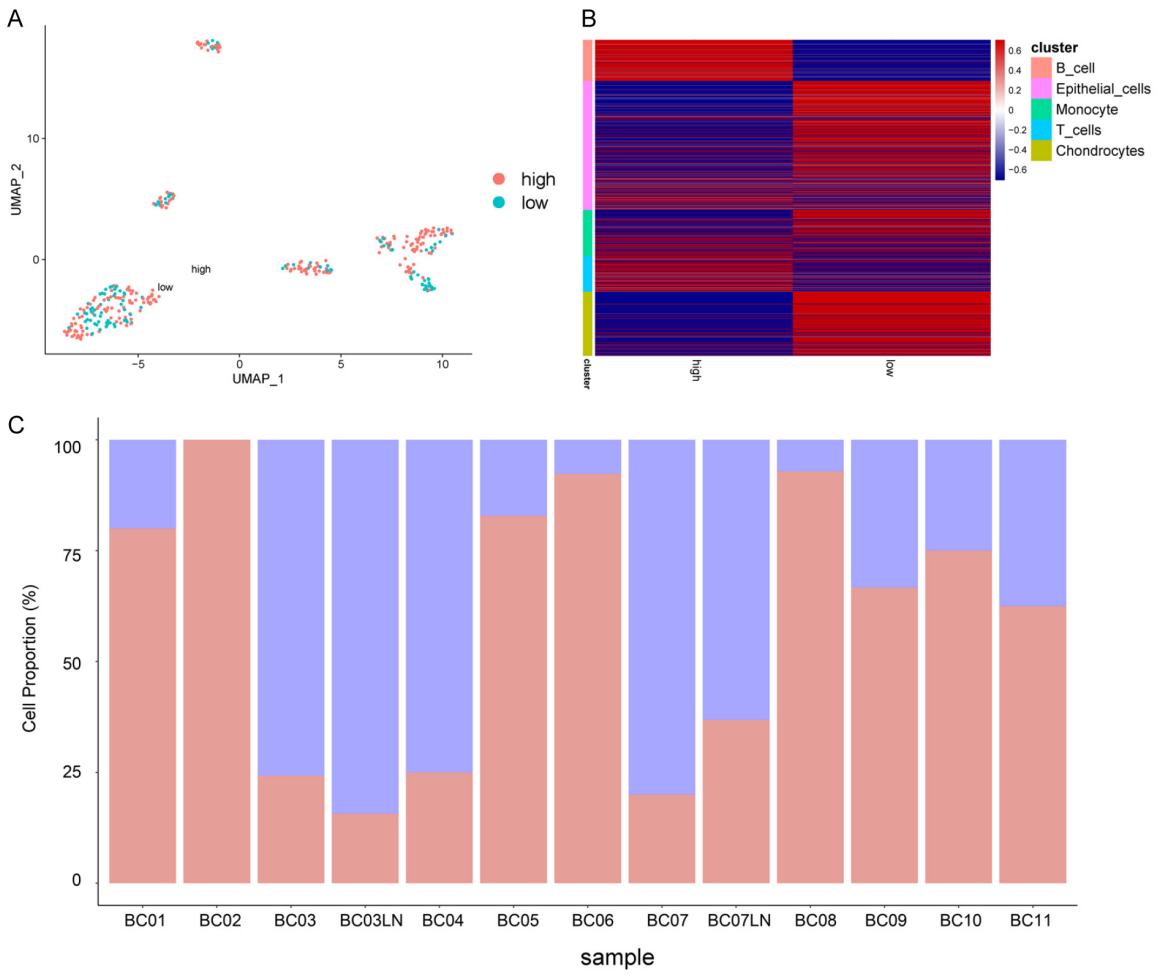


Figure 11. Cell populations were divided into high and low expression groups based on the expression levels of NRGs. A. Umap visualization of cells in the high and low risk groups. B. Expression heatmap of differentially expressed genes between single-cell samples with different NRGs expression levels. C. Bar graph of the ratio between samples with different NRGs expression levels. The horizontal and vertical coordinates represent the proportion of the number of cells in the sample and high and low expression groups, respectively.

the dimensionality reduction visualization of Umap in different cell populations. Different cell types were annotated based on different colours. **Figure 10B** shows the bubble chart of marker gene expression in different cell groups. Six genes were specifically expressed in different types of cell populations. The violin diagram of the expression levels of these genes in different cell groups is shown in Figure S2A-F. **Figure 10C** shows the histogram of the proportion of different cell groups in different samples. The proportion of different cell types in different samples varies greatly. Furthermore, based on the expression levels of NRGs, all cell groups were divided into two groups: the high expression group and the low expression group.

Figure 11A shows the cell Umap visualization of the high-risk and low-risk groups. The differential gene expression heatmap of the high- and low-expression samples of NRGs is shown in **Figure 11B**. **Figure 11C** displays a histogram of the ratio between samples with different NRG expression levels. The population of highly expressing cells was larger in a few samples. Furthermore, to explore the signalling pathway information of samples with different NRG expression levels, GSVA (Figure S6) and GSEA (**Figure 12A**) were performed on two groups of samples with high and low expression of NRGs. We will discuss in detail the differences between high and low expression samples in terms of the involved pathways in the discussion section. In addition, we provide

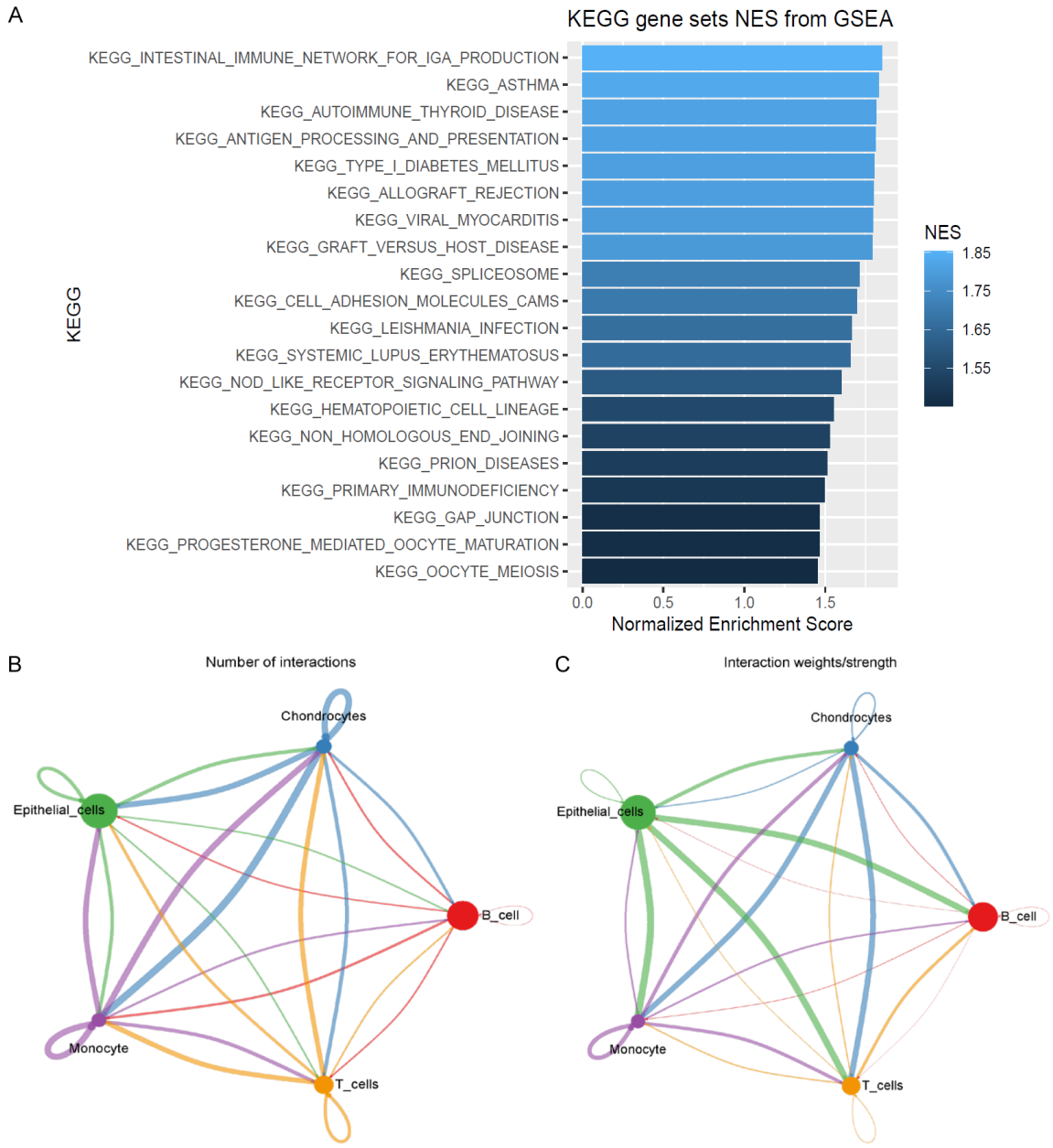


Figure 12. Analysis of intercellular communication. A. GSEA analysis of cells in the high and low expression groups. B. Diagram of the communication network between different cell populations, where the thickness of connecting lines and the size of nodes represent the number of receptor-ligand pairs and the number of cells, respectively. C. Diagram of the communication network between different cell populations, where the thickness of connecting lines and the size of nodes represent the strength of the interaction.

the KEGG and GO analysis results of different genes between single-cell samples with different NRG expression levels in Figure S3. Detailed information on the KEGG pathways obtained by GSEA of differentially expressed genes between single-cell samples with different NRG expression levels is given in Figure S4.

According to the results of the previous annotation of the singleR algorithm, we drew a communication network diagram for the different types of cell groups (**Figure 12B**, **12C**). Specifically, in the cell communication diagram presented in **Figure 12C**, the size of the nodes represents the number of cells, and the thick lines between nodes represent the number of receptor-ligand pairs. In the cell communication diagram shown in **Figure 12B**, the thickness of the connecting lines and the size of the nodes represent the strength of interac-

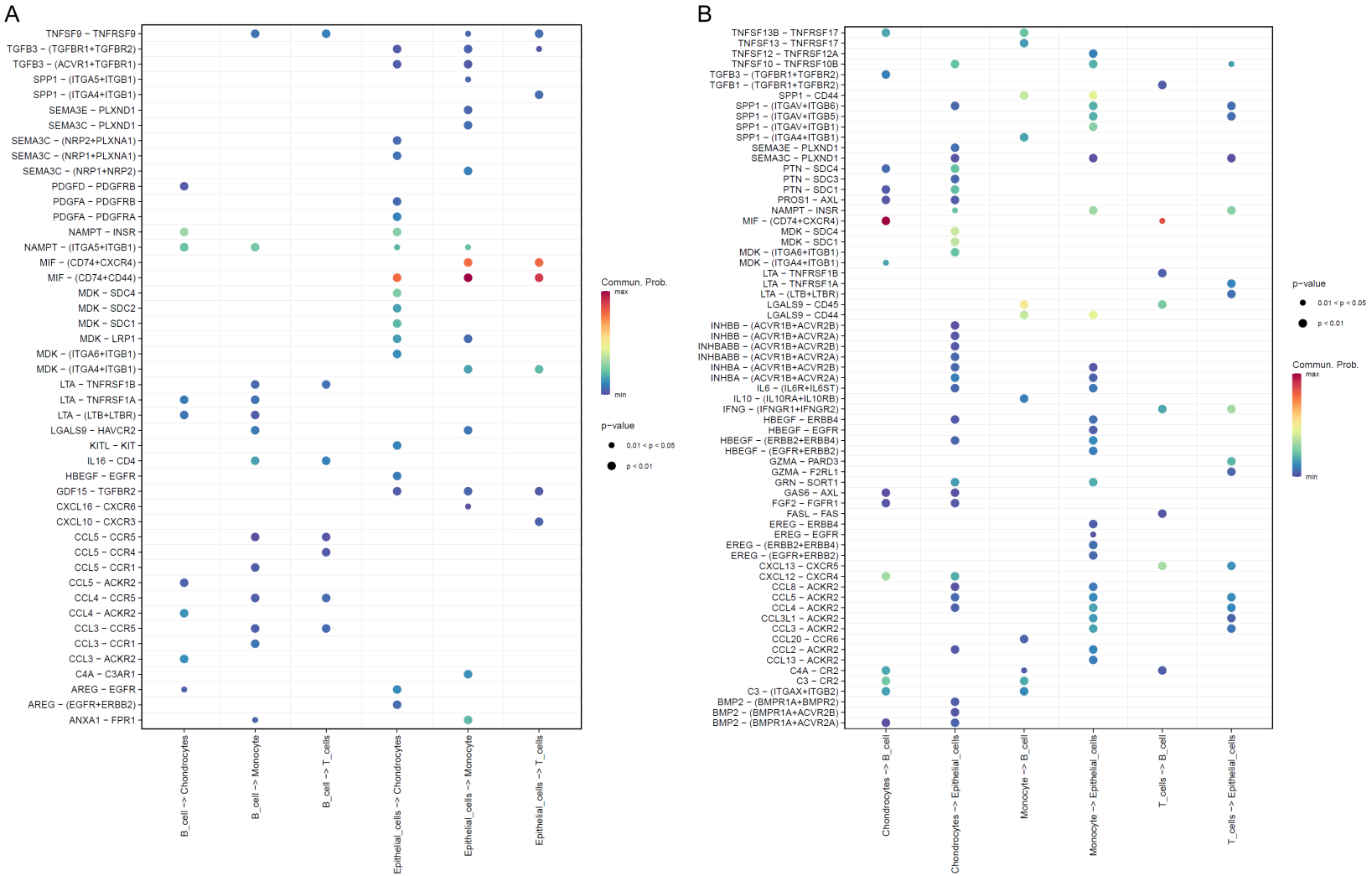


Figure 13. A. Bubble diagram of signal pathway obtained using Epithelial cells and B cells as receptors and other cell groups as ligands. B. Bubble diagram of signal pathway obtained by taking other cell groups as receptors, Epithelial cells, and B cells as ligands.

Table 3. Comparison of contour coefficient with other algorithms

Algorithms	Contour coefficients
NMF	0.1153
Deep semi-NMF	0.1363
DO-DSNMF	0.2704

tion. **Figure 12C** shows that there were more quantitative and stronger relationships between epithelial cells and B cells and other cell groups. Therefore, in **Figure 13A**, we displayed the bubble diagram of the signalling pathway obtained by using epithelial cells and B cells as receptors and other cell groups as ligands. In **Figure 13B**, the bubble diagram of the signal pathway obtained using other cell groups as receptors and epithelial cells and B cells as ligands is presented.

Discussion

In this study, the DO-DSNMF algorithm was proposed to distinguish BC subtypes from the perspective of biological information. First, based on the NRGs of BC samples in the TCGA queue, differential expression analysis was performed, and DENRGs were obtained. Then, BC was divided into different subtypes by using the DO-DSNMF algorithm, and we verified the performance of the algorithm. In the following discussion, the results of the performance comparison between this algorithm and other algorithms are analysed in detail. In particular, to specify the effectiveness of the algorithm, this study compared the proposed algorithm with NMF and deep semi-NMF under the same experimental parameters (**Table 3**). The PCA and the t-sne algorithms were used to reduce the data dimension to two dimensions for visualization (Figure S5).

We analysed the differences in immune-related genes, immune cells, immune function, and immune inspection sites among different subtypes. First, we examined the critical role of different subtypes of sample immune cells (B cells, CD8+ T cells, and NK cells) in breast cancer (**Figure 5A**). B cells play an important role in the immune response, and experiments have confirmed that tumour infiltration of B cells in BC can effectively promote tumour immunity [22]. Nalio Ramos R et al. found that FOLR2 macrophages were related to the infiltration of human BC CD8+ T cells [23]. A recent study showed that hepatic stellate cells can inhibit BC dormancy maintained by NK cells [24]. There are significant differences in immune function among different subtypes (**Figure 5B**), and the increase in apoptosis is related to the cytolytic activity of BC [25]. Reimann H et al. used the BC cell line to verify that HLA-binding new epitopes of breast cancer can be used for personalized treatment of BC [26]. The role of inflammation and its related microRNAs in breast cancer were systematically reviewed [27]. Wuerfel FM et al. evaluated the prognosis and predictive value of HLA-G and HLA-F protein subtype expression patterns in BC patients at immune examination sites with significant differences among different subtypes (**Figure 5E**) [28]. The correlation between HLA-DRB1 gene polymorphisms and BC has been confirmed [29]. In addition, necrotizing apoptosis plays a major role in the progression and metastasis of BC [30]. We discussed the relationship between immune cells and necrotizing apoptosis. For example, tumour necrosis apoptosis can promote breast cancer metastasis, and the level of metastasis is related to the consumption of CD8+ T cells [31]. Necrotizing apoptotic B cells will exhibit mitochondrial dysfunction and hypoxia [32]. The necrotizing apoptosis signal transduction molecules affect the expression level of genes in NK and T lymphocytes of BC patients and are also related to macrophages [33]. An important feature of solid tumours is the death of necrotic cells and the subsequent release of DAMPs. Bone marrow cells (monocytes, dendritic cells (DCs), and granulocytes) die at an early stage, this initiates and regulates the subsequent inflammatory response [34].

In addition, we also analysed the signalling pathways involved in subtypes by GSVA (**Figure 5F**). Zhao P et al. verified the possible role of glycosyl phosphatidylinositol (GPI) ankyrin in BC progression [35]. MAL2 in breast tumour cells can reduce antigen presentation on tumour cells and promote immune escape in BC [36]. Drugs for allograft rejection have tremendous therapeutic potential in the treatment of BC [37]. The relationship between these pathways and necrotizing apoptosis has also been confirmed in the literature. The inhibition of the synthesis of the GPI anchor can lead to autophagy and possibly necrotizing

apoptosis of Aspergillus fumigatus [38]. Extracellular vesicles (EVs) released by necrotizing apoptotic cells contain new tumour antigens and are rich in components of antigen processing and presentation [39]. Zhuo DX et al. indicated that the mitochondrial pathway in allograft necrosis, apoptosis, and inflammatory signal transduction is a new direction for heart transplantation in the future [40].

In this study, prognosis-related DENRGs were screened by Lasso-Cox regression analysis, and KM survival analysis was carried out to verify the effect of these genes on the BC subtype. Examples include BCL2, BIRC3, BNIP3, BRD4, PLK1, and FLT3. BCL2 plays an important role in the prognosis of BC [41]. BIRC3 has been shown to be involved in the chemical resistance to adriamycin in BC cells [42]. By investigating the role of BNIP3 in normal and neoplastic breasts, Tan EY et al. confirmed that BNIP3 could be used as a progressive marker of primary human breast cancer [43]. Some targets or pathways, such as BRD4 and PLK1, have also been confirmed to be related to BC [44]. FLT3 has also been shown to be associated with the immune infiltration of breast cancer [45].

Furthermore, we divided all BC samples into high-risk and low-risk groups according to the sample risk score obtained from the prognosis model. We also conducted TMB analysis and GSEA on the two groups and discussed their immune landscape. In the enrichment pathway of the high-risk group (**Figure 9E**), aminoacyl-tRNA biosynthesis was highly correlated with the BC subtype [46]. Cell cycle-related long noncoding RNA is a key regulator of breast cancer progression and metastasis [47]. Mutations in many genes encoding homologous recombinant (HR) proteins may increase BC risk [48]. In the enrichment pathway of the low-risk group (**Figure 9F**), regulating the haematopoietic lineage potential of bone marrow can effectively prevent bone metastasis of BC [49]. When analysing the antitumour mechanism of cinobufotalin in BC, researchers found that the “neuroactive ligand-receptor interaction” pathway is crucial for cinobufotalin in BC [50]. We confirmed the significant role of immune cells, such as DCs, macrophages, and neutrophils. There were substantial differences in these cells between the high-risk and low-risk groups, as shown in **Figure 9G** and in the literature. Met O et al. [51] used human dendritic cells (DCs) modified by survivin mRNA to analyse survivin-specific T cells in breast cancer patients. Macrophages play an important role in the BC immune microenvironment [52]. Neutrophils are associated with lung metastasis of breast cancer [53]. As shown in **Figure 9H**, immunotherapy targeting T-cell inhibition may become one of the therapeutic directions of human BC [54]. Furthermore, the results showed that the immune level of high-risk group patients was lower than that of low-risk group patients. Therefore, improving the immune level of breast cancer patients may improve the prognosis of breast cancer patients. In a study of iron death-related gene markers in breast cancer patients, Wang D et al. found that the type I IFN response and type II IFN response were low in the high-risk population [55]. As shown in **Figure 9I**, we found that the immune examination sites TMIGD2 and CD27 in the high-risk and low-risk groups are consistent with a previous report [56], but the expression patterns of these two genes in BC are different.

In this research, the scRNA-seq data of BC were analysed according to NRGs. Specifically, we clustered and annotated the scRNA-seq samples of different BCs. Then, based on the expression level of NRGs, the cell groups were divided into high-expression and low-expression groups, and the proportion of high-expression and low-expression cells in different cell groups was analysed. GSEA and GSVA were carried out on single-cell samples with different NRG expression levels. In addition, KEGG and GO pathway analyses involving the differentially expressed genes of high-expression and low-expression cells are presented in Figure S3A-D. The GO enrichment analysis of upregulated genes is shown in Figure S2A, and Eoghan P McGrath et al. reviewed the unfolded protein reaction in BC [57]. Stacey Aggarwal et al. found that the depletion of the dAKAP1 protein kinase A signal island from the mitochondrial outer membrane changed the metabolism and motility of BC cells [58]. Maitham A Khajah and others found that Na+/K+ ATPase activity promoted the invasion of endocrine-resistant BC cells [59]. As shown in the GO enrichment analysis of downregulated genes in Figure S2B, the relationship between some pathways and BC was also confirmed. Caveolin-1 is a 22-kD

transmembrane protein that is highly expressed in plasma membrane invaginations called caveolae, and it is regulated by single molecules of several key signalling pathways [60]. Nicolini et al. showed that progressive defects in cell-mediated immunity appear in the first few months after BC operation, and it takes a long time to recover after apparent metastatic disease [61]. In accordance with the KEGG enrichment analysis of upregulated genes shown in Figure S2C, Lorenza Sisinni and others systematically reviewed BC endoplasmic reticulum stress and the unfolded protein response [62]. Valproic acid has been shown to upregulate oestrogen receptors (ERs) in breast cancer and prostate cancer. Samir Rabadiya and others evaluated the therapeutic effect of magnesium valproate (MgV) on cardiac complications related to type 1 diabetes in rats and found that magnesium valproate improved type 1 diabetes and cardiomyopathy in diabetic rats through oestrogen receptors [63]. We also analysed the relationship between several pathways in Figure S3 and necrotizing apoptosis. FKBP12 is a protein that regulates protein folding, and it may mediate necrotic apoptosis [64]. P53 can regulate the death of necrotic/necrotic apoptotic cells through the mitochondrial membrane [65]. Ubiquitin plays an important role in diseases related to necrotizing apoptosis [66].

Finally, we discussed two kinds of cells with more interactions with other cell groups as receptor- and ligand-mediated signalling pathways. In **Figure 12E**, we analysed the important roles of the GDF15-TGFBR2, CD74-(MIF+CD74), and CCl4-CCR5 genes in BC. GDF15 is a potential therapeutic target of BC radiotherapy [67]. TGFBR2 is a potential receptor of GDF15, and it is inactivated in the carcinogenesis of many types of cancer [68]. The interactions between CD74 and MIF and between CD74 and CD44 may be potential tumour markers of BC cells [69]. The CCL4-CCR5 axis plays an important role in bone fibroblasts in BC bone metastasis [70]. As shown in **Figure 12F**, we confirmed the close correlation between the FGF2-FGFR1 and GAS6-AXL gene pairs and BC. There is an interaction between cancer-related fibroblasts and BC cells in functional FGF2/FGFR1 [71]. The GAS6/AXL axis has also been shown to be related to chemical resistance and metastasis of BC [72].

Conclusion

This study proposed a novel clustering algorithm for analysing BC subtypes and constructed an NRG prognostic model for BC. The prognosis and immune landscape of BC patients were evaluated by this model. The cell experiment supported its role in BC, which provides a potential therapeutic target for the treatment of BC.

Acknowledgements

The work was supported by the Seedling Plan of Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region (GXWCH-YMJH-2018005) and Youth Science Foundation of Guangxi Medical University (GXMUYSF-202224).

Disclosure of conflict of interest

None.

Address correspondence to: Shipeng Ning, Department of Breast Surgery, Guangxi Medical University Cancer Hospital, Nanning 530000, Guangxi, China. E-mail: nspdoctor@163.com; Jianlan Mo, Department of Anesthesiology, The Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, Nanning 530000, Guangxi, China. E-mail: mojianlan0771@163.com

References

[1] Waks AG and Winer EP. Breast cancer treatment: a review. JAMA 2019; 321: 288-300.
[2] Siegel RL, Miller KD and Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020; 70: 7-30.
[3] Droin N, Guéry L, Benikhlef N and Solary E. Targeting apoptosis proteins in hematological malignancies. Cancer Lett 2013; 332: 325-334.
[4] Liedtke C and Trautwein C. The role of TNF and Fas dependent signaling in animal models of inflammatory liver injury and liver cancer. Eur J Cell Biol 2012; 91: 582-589.
[5] Perier C, Bové J and Vila M. Mitochondria and programmed cell death in Parkinson's disease: apoptosis and beyond. Antioxid Redox Signal 2012; 16: 883-895.
[6] Wu W, Liu P and Li J. Necroptosis: an emerging form of programmed cell death. Crit Rev Oncol Hematol 2012; 82: 249-258.
[7] Chan VSH, Lam TPW and Lam WWM. Nephroptosis: the wandering kidney. Kidney Res Clin Pract 2018; 37: 306-307.
[8] Degterev A, Hitomi J, Germscheid M, Ch'en IL, Korkina O, Teng X, Abbott D, Cuny GD, Yuan C,

Wagner G, Hedrick SM, Gerber SA, Lugovskoy A and Yuan J. Identification of RIP1 kinase as a specific cellular target of necrostatins. Nat Chem Biol 2008; 4: 313-321.

[9] Seifert L, Werba G, Tiwari S, Giao Ly NN, Alothman S, Alqunaibit D, Avanzi A, Barilla R, Daley D, Greco SH, Torres-Hernandez A, Pergamo M, Ochi A, Zambirinis CP, Pansari M, Rendon M, Tippens D, Hundeyin M, Mani VR, Hajdu C, Engle D and Miller G. The necrosome promotes pancreatic oncogenesis via CXCL1 and Mincle-induced immune suppression. Nature 2016; 532: 245-249.

[10] Wang N and Liu D. Identification and validation a necroptosis-related prognostic signature and associated regulatory axis in stomach adenocarcinoma. Onco Targets Ther 2021; 14: 5373-5383.

[11] Wang Q, Li M, Yang M, Yang Y, Song F, Zhang W, Li X and Chen K. Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: implications for immune checkpoint blockade therapy. Aging (Albany NY) 2020; 12: 3312-3339.

[12] Jiang C, Liu Y, Wen S, Xu C and Gu L. In silico development and clinical validation of novel 8 gene signature based on lipid metabolism related genes in colon adenocarcinoma. Pharmacol Res 2021; 169: 105644.

[13] Liu L, Zhang SW, Zhang YC, Liu H, Zhang L, Chen R, Huang Y and Meng J. Decomposition of RNA methylome reveals co-methylation patterns induced by latent enzymatic regulators of the epitranscriptome. Mol Biosyst 2015; 11: 262-274.

[14] Sherborne AL, Davidson PR, Yu K, Nakamura AO, Rashid M and Nakamura JL. Mutational analysis of ionizing radiation induced neoplasms. Cell Rep 2015; 12: 1915-1926.

[15] de Campos CP, Rancoita PM, Kwee I, Zucca E, Zaffalon M and Bertoni F. Discovering subgroups of patients from DNA copy number data using NMF on compacted matrices. PLoS One 2013; 8: e79720.

[16] Gaujoux R and Seoighe C. Semi-supervised nonnegative matrix factorization for gene expression deconvolution: a case study. Infect Genet Evol 2012; 12: 913-921.

[17] Hartsperger ML, Blöchl F, Stümpflen V and Theis FJ. Structuring heterogeneous biological information using fuzzy clustering of k-partite graphs. BMC Bioinformatics 2010; 11: 522.

[18] Wang X, Fu X, Zhang J, Xiong C, Zhang S and Lv Y. Identification and validation of m(6)A RNA methylation regulators with clinical prognostic value in papillary thyroid cancer. Cancer Cell Int 2020; 20: 203.

[19] Wang T, Dang N, Tang G, Li Z, Li X, Shi B, Xu Z, Li L, Yang X, Xu C and Ye K. Integrating bulk and single-cell RNA sequencing reveals cellular heterogeneity and immune infiltration in hepatocellular carcinoma. Mol Oncol 2022; 16: 2195-2213.

[20] Jiang A, Wang J, Liu N, Zheng X, Li Y, Ma Y, Zheng H, Chen X, Fan C, Zhang R, Fu X and Yao Y. Integration of single-cell RNA sequencing and bulk RNA sequencing data to establish and validate a prognostic model for patients with lung adenocarcinoma. Front Genet 2022; 13: 833797.

[21] Trigeorgis G, Bousmalis K, Zafeiriou S and Schuller BW. A deep matrix factorization method for learning attribute representations. IEEE Trans Pattern Anal Mach Intell 2017; 39: 417-429.

[22] Garaud S, Buisseret L, Solinas C, Gu-Trantien C, de Wind A, Van den Eynden G, Naveaux C, Lodewyckx JN, Boisson A, Duvillier H, Craciun L, Ameye L, Veys I, Paesmans M, Larsimont D, Piccart-Gebhart M and Willard-Gallo K. Tumor infiltrating B-cells signal functional humoral immune responses in breast cancer. JCI Insight 2019; 5: e129641.

[23] Nalio Ramos R, Missolo-Koussou Y, Gerber-Ferder Y, Bromley CP, Bugatti M, Núñez NG, Tosello Boari J, Richer W, Menger L, Denizeau J, Sedlik C, Caudana P, Kotsias F, Niborski LL, Viel S, Bohec M, Lameiras S, Baulande S, Lesage L, Nicolas A, Meseure D, Vincent-Salomon A, Reyal F, Dutertre CA, Ginhoux F, Vimeux L, Donnadieu E, Buttard B, Galon J, Zelenay S, Vermi W, Guermonprez P, Piaggio E and Helft J. Tissue-resident FOLR2(+) macrophages associate with CD8(+) T cell infiltration in human breast cancer. Cell 2022; 185: 1189-1207. e25.

[24] Correia AL, Guimaraes JC, Auf der Maur P, De Silva D, Trefny MP, Okamoto R, Bruno S, Schmidt A, Mertz K, Volkmann K, Terracciano L, Zippelius A, Vetter M, Kurzeder C, Weber WP and Bentires-Alj M. Hepatic stellate cells suppress NK cell-sustained breast cancer dormancy. Nature 2021; 594: 566-571.

[25] Murthy V, Oshi M, Tokumaru Y, Endo I and Takabe K. Increased apoptosis is associated with robust immune cell infiltration and cytolytic activity in breast cancer. Am J Cancer Res 2021; 11: 3674-3687.

[26] Reimann H, Nguyen A, Sanborn JZ, Vaske CJ, Benz SC, Niazi K, Rabizadeh S, Spilman P, Mackensen A, Ruebner M, Hein A, Beckmann MW, van der Meijden ED, Bausenwein J, Kretschmann S, Griffioen M, Schlom J, Gulley JL, Lee KL, Hamilton DH, Soon-Shiong P, Fasching PA and Kremer AN. Identification and validation of expressed HLA-binding breast cancer neoepitopes for potential use in individualized cancer therapy. J Immunother Cancer 2021; 9: e002605.

[27] Bahiraee A, Ebrahimi R, Halabian R, Aghabozorgi AS and Amani J. The role of inflammation and its related microRNAs in breast cancer: a narrative review. J Cell Physiol 2019; 234: 19480-19493.
[28] Wuerfel FM, Huebner H, Häberle L, Gass P, Hein A, Jud SM, Hack CC, Wunderle M, Schulz-Wendtland R, Erber R, Hartmann A, Ekici AB, Beckmann MW, Fasching PA and Ruebner M. HLA-G and HLA-F protein isoform expression in breast cancer patients receiving neoadjuvant treatment. Sci Rep 2020; 10: 15750.
[29] Liu L, Sun X, Yuan C and Liu H. Relationship between HLA-DRB1 gene polymorphism and breast cancer: a protocol for systematic review and meta-analysis. Medicine (Baltimore) 2021; 100: e25078.
[30] Yu H, Lv W, Tan Y, He X, Wu Y, Wu M and Zhang Q. Immunotherapy landscape analyses of necroptosis characteristics for breast cancer patients. J Transl Med 2022; 20: 328.
[31] Liu Z, Choksi S, Kwon HJ, Jiao D, Liu C and Liu ZG. Tumor necroptosis-mediated shedding of cell surface proteins promotes metastasis of breast cancer by suppressing anti-tumor immunity. Breast Cancer Res 2023; 25: 10.
[32] Fan H, Liu F, Dong G, Ren D, Xu Y, Dou J, Wang T, Sun L and Hou Y. Activation-induced necroptosis contributes to B-cell lymphopenia in active systemic lupus erythematosus. Cell Death Dis 2014; 5: e1416.
[33] Stoll G, Ma Y, Yang H, Kepp O, Zitvogel L and Kroemer G. Pro-necrotic molecules impact local immunosurveillance in human breast cancer. Oncoimmunology 2017; 6: e1299302.
[34] Lotfi R, Kaltenmeier C, Lotze MT and Bergmann C. Until death do us part: necrosis and oxidation promote the tumor microenvironment. Transfus Med Hemother 2016; 43: 120-132.
[35] Zhao P, Nairn AV, Hester S, Moremen KW, O'Regan RM, Oprea G, Wells L, Pierce M and Abbott KL. Proteomic identification of glycosylphosphatidylinositol anchor-dependent membrane proteins elevated in breast carcinoma. J Biol Chem 2012; 287: 25230-25240.
[36] Fang Y, Wang L, Wan C, Sun Y, Van der Jeught K, Zhou Z, Dong T, So KM, Yu T, Li Y, Eyvani H, Colter AB, Dong E, Cao S, Wang J, Schneider BP, Sandusky GE, Liu Y, Zhang C, Lu X and Zhang X. MAL2 drives immune evasion in breast cancer by suppressing tumor antigen presentation. J Clin Invest 2021; 131: e140837.
[37] Steelman LS, Martelli AM, Cocco L, Libra M, Nicoletti F, Abrams SL and McCubrey JA. The therapeutic potential of mTOR inhibitors in breast cancer. Br J Clin Pharmacol 2016; 82: 1189-1212.
[38] Yan J, Du T, Zhao W, Hartmann T, Lu H, Lü Y, Ouyang H, Jiang X, Sun L and Jin C. Transcriptome and biochemical analysis reveals that suppression of GPI-anchor synthesis leads to autophagy and possible necroptosis in Aspergillus fumigatus. PLoS One 2013; 8: e59013.
[39] Shlomovitz I, Erlich Z, Arad G, Edry-Botzer L, Zargarian S, Cohen H, Manko T, Ofir-Birin Y, Cooks T, Regev-Rudzki N and Gerlic M. Proteomic analysis of necroptotic extracellular vesicles. Cell Death Dis 2021; 12: 1059.
[40] Zhuo DX, Ginder K and Hardin EA. Markers of immune function in heart transplantation: implications for immunosuppression and screening for rejection. Curr Heart Fail Rep 2021; 18: 33-40.
[41] Callagy GM, Webber MJ, Pharoah PD and Caldas C. Meta-analysis confirms BCL2 is an independent prognostic marker in breast cancer. BMC Cancer 2008; 8: 153.
[42] Mendoza-Rodríguez M, Arévalo Romero H, Fuentes-Pananá EM, Ayala-Sumuano JT and Meza I. IL-1β induces up-regulation of BIRC3, a gene involved in chemoresistance to doxorubicin in breast cancer cells. Cancer Lett 2017; 390: 39-44.
[43] Tan EY, Campo L, Han C, Turley H, Pezzella F, Gatter KC, Harris AL and Fox SB. BNIP3 as a progression marker in primary human breast cancer; opposing functions in in situ versus invasive cancer. Clin Cancer Res 2007; 13: 467-474.
[44] Liu T, Song S, Wang X and Hao J. Small-molecule inhibitors of breast cancer-related targets: potential therapeutic agents for breast cancer. Eur J Med Chem 2021; 210: 112954.
[45] Chen R, Wang X, Fu J, Liang M and Xia T. High FLT3 expression indicates favorable prognosis and correlates with clinicopathological parameters and immune infiltration in breast cancer. Front Genet 2022; 13: 956869.
[46] Alakwaa FM and Savelieff MG. Bioinformatics analysis of metabolomics data unveils association of metabolic signatures with methylation in breast cancer. J Proteome Res 2020; 19: 2879-2889.
[47] Zangouei AS, Zangoue M, Taghehchian N, Zangooie A, Rahimi HR, Saburi E, Alavi MS and Moghbeli M. Cell cycle related long non-coding RNAs as the critical regulators of breast cancer progression and metastasis. Biol Res 2023; 56: 1.
[48] Ali RMM, McIntosh SA and Savage KI. Homologous recombination deficiency in breast cancer: implications for risk, cancer development, and therapy. Genes Chromosomes Cancer 2021; 60: 358-372.
[49] Ubellacker JM, Baryawno N, Severe N, DeCristo MJ, Sceneay J, Hutchinson JN, Haider MT, Rhee CS, Qin Y, Gregory WM, Garrido-Castro

AC, Holen I, Brown JE, Coleman RE, Scadden DT and McAllister SS. Modulating bone marrow hematopoietic lineage potential to prevent bone metastasis in breast cancer. Cancer Res 2018; 78: 5300-5314.
[50] Li J, Rong MH, Dang YW, He RQ, Lin P, Yang H, Li XJ, Xiong DD, Zhang LJ, Qin H, Feng CX, Chen XY, Zhong JC, Ma J and Chen G. Differentially expressed gene profile and relevant pathways of the traditional Chinese medicine cinobufotalin on MCF-7 breast cancer cells. Mol Med Rep 2019; 19: 4256-4270.
[51] Met O and Svane IM. Analysis of survivin-specific T cells in breast cancer patients using human DCs engineered with survivin mRNA. Methods Mol Biol 2013; 969: 275-292.
[52] Li Y, Zhao X, Liu Q and Liu Y. Bioinformatics reveal macrophages marker genes signature in breast cancer to predict prognosis. Ann Med 2021; 53: 1019-1031.
[53] Xiao Y, Cong M, Li J, He D, Wu Q, Tian P, Wang Y, Yang S, Liang C, Liang Y, Wen J, Liu Y, Luo W, Lv X, He Y, Cheng DD, Zhou T, Zhao W, Zhang P, Zhang X, Xiao Y, Qian Y, Wang H, Gao Q, Yang QC, Yang Q and Hu G. Cathepsin C promotes breast cancer lung metastasis by modulating neutrophil infiltration and neutrophil extracellular trap formation. Cancer Cell 2021; 39: 423-437.e427.
[54] Janakiram M, Abadi YM, Sparano JA and Zang X. T cell coinhibition and immunotherapy in human breast cancer. Discov Med 2012; 14: 229-236.
[55] Wang D, Wei G, Ma J, Cheng S, Jia L, Song X, Zhang M, Ju M, Wang L, Zhao L and Xin S. Identification of the prognostic value of ferroptosis-related gene signature in breast cancer patients. BMC Cancer 2021; 21: 645.
[56] Fang J, Chen F, Liu D, Gu F, Chen Z and Wang Y. Prognostic value of immune checkpoint molecules in breast cancer. Biosci Rep 2020; 40: BSR20201054.
[57] McGrath EP, Logue SE, Mnich K, Deegan S, Jäger R, Gorman AM and Samali A. The unfolded protein response in breast cancer. Cancers (Basel) 2018; 10: 344.
[58] Aggarwal S, Gabrovsek L, Langeberg LK, Golkowski M, Ong SE, Smith FD and Scott JD. Depletion of dAKAP1-protein kinase A signaling islands from the outer mitochondrial membrane alters breast cancer cell metabolism and motility. J Biol Chem 2019; 294: 3152-3168.
[59] Khajah MA, Mathew PM and Luqmani YA. Na+/K+ ATPase activity promotes invasion of endocrine resistant breast cancer cells. PLoS One 2018; 13: e0193779.
[60] Anwar SL, Wahyono A, Aryandono T and Haryono SJ. Caveolin-1 in breast cancer: single molecule regulation of multiple key signaling pathways. Asian Pac J Cancer Prev 2015; 16: 6803-6812.
[61] Nicolini A, Ferrari P, Spinelli R, Carpi A, Sagripanti A and Ambrogi F. Cell-mediated immunity in breast cancer patients. Biomed Pharmacother 1996; 50: 337-343.
[62] Sisinni L, Pietrafesa M, Lepore S, Maddalena F, Condelli V, Esposito F and Landriscina M. Endoplasmic reticulum stress and unfolded protein response in breast cancer: the balance between apoptosis and autophagy and its role in drug resistance. Int J Mol Sci 2019; 20: 857.
[63] Rabadiya S, Bhadada S, Dudhrejiya A, Vaishnav D and Patel B. Magnesium valproate ameliorates type 1 diabetes and cardiomyopathy in diabetic rats through estrogen receptors. Biomed Pharmacother 2018; 97: 919-927.
[64] Wang Z, Feng J, Yu J and Chen G. FKBP12 mediates necroptosis by initiating RIPK1-RIPK3-MLKL signal transduction in response to TNF receptor 1 ligation. J Cell Sci 2019; 132: jcs227777.
[65] Dashzeveg N and Yoshida K. Cell death decision by p53 via control of the mitochondrial membrane. Cancer Lett 2015; 367: 108-112.
[66] Chen Y, Ren W, Wang Q, He Y, Ma D and Cai Z. The regulation of necroptosis by ubiquitylation. Apoptosis 2022; 27: 668-684.
[67] Zhao X, Liu X, Hu S, Pan Y, Zhang J, Tai G and Shao C. GDF15 contributes to radioresistance by mediating the EMT and stemness of breast cancer cells. Int J Mol Sci 2022; 23: 10911.
[68] Tarfiei GA, Shadboorestan A, Montazeri H, Rahmanian N, Tavosi G and Ghahremani MH. GDF15 induced apoptosis and cytotoxicity in A549 cells depends on TGFBR2 expression. Cell Biochem Funct 2019; 37: 320-330.
[69] Ssadh HA, Abdulmonem WA, Rasheed Z, Madar IH, Alhoderi J, Eldeen SKN, Alradhwan A, Alasmael N, Alkhamiss A and Fernández N. Knockdown of CD-74 in the proliferative and apoptotic activity of breast cancer cells. Open Access Maced J Med Sci 2019; 7: 3169-3176.
[70] Pervaiz A, Zepp M, Mahmood S, Ali DM, Berger MR and Adwan H. CCR5 blockage by maraviroc: a potential therapeutic option for metastatic breast cancer. Cell Oncol (Dordr) 2019; 42: 93-106.
[71] Santolla MF, Vivacqua A, Lappano R, Rigiracciolo DC, Cirillo F, Galli GR, Talia M, Brunetti G, Miglietta AM, Belfiore A and Maggiolini M. GPER mediates a feedforward FGF2/FGFR1 paracrine activation coupling CAFs to cancer cells toward breast tumor progression. Cells 2019; 8: 223.
[72] Wang C, Jin H, Wang N, Fan S, Wang Y, Zhang Y, Wei L, Tao X, Gu D, Zhao F, Fang J, Yao M and Qin W. Gas6/Axl axis contributes to chemoresistance and metastasis in breast cancer through Akt/GSK-3β/β-catenin signaling. Theranostics 2016; 6: 1205-1219.

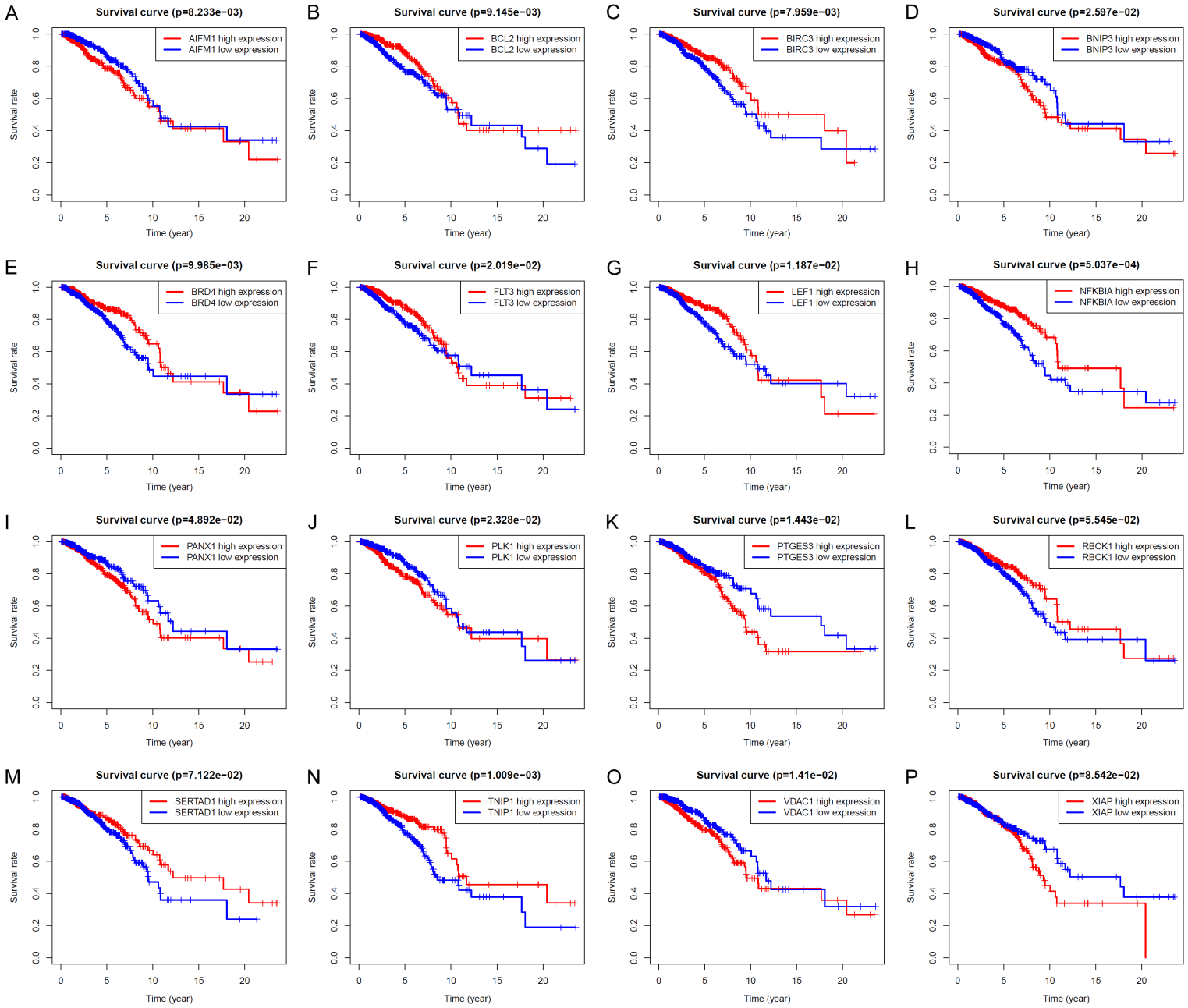

A
Survival curve (p=8.233e−03)
AIFM1 high expression
AIFM1 low expression
B
Survival curve (p=9.145e−03)
BCL2 high expression
BCL2 low expression
C
Survival curve (p=7.959e−03)
BIRC3 high expression
BIRC3 low expression
D
Survival curve (p=2.597e−02)
BNIP3 high expression
BNIP3 low expression
E
Survival curve (p=9.985e−03)
BRD4 high expression
BRD4 low expression
F
Survival curve (p=2.019e−02)
FLT3 high expression
FLT3 low expression
G
Survival curve (p=1.187e−02)
LEF1 high expression
LEF1 low expression
H
Survival curve (p=5.037e−04)
NFKBIA high expression
NFKBIA low expression
I
Survival curve (p=4.892e−02)
PANX1 high expression
PANX1 low expression
J
Survival curve (p=2.328e−02)
PLK1 high expression
PLK1 low expression
K
Survival curve (p=1.443e−02)
PTGES3 high expression
PTGES3 low expression
L
Survival curve (p=5.545e−02)
RBCK1 high expression
RBCK1 low expression
M
Survival curve (p=7.122e−02)
SERTAD1 high expression
SERTAD1 low expression
N
Survival curve (p=1.009e−03)
TNIP1 high expression
TNIP1 low expression
O
Survival curve (p=1.41e−02)
VDAC1 high expression
VDAC1 low expression
P
Survival curve (p=8.542e−02)
XIAP high expression
XIAP low expression
Survival rate
Time (year)

Figure S1. KM survival curve of prognosis-related NRGs.

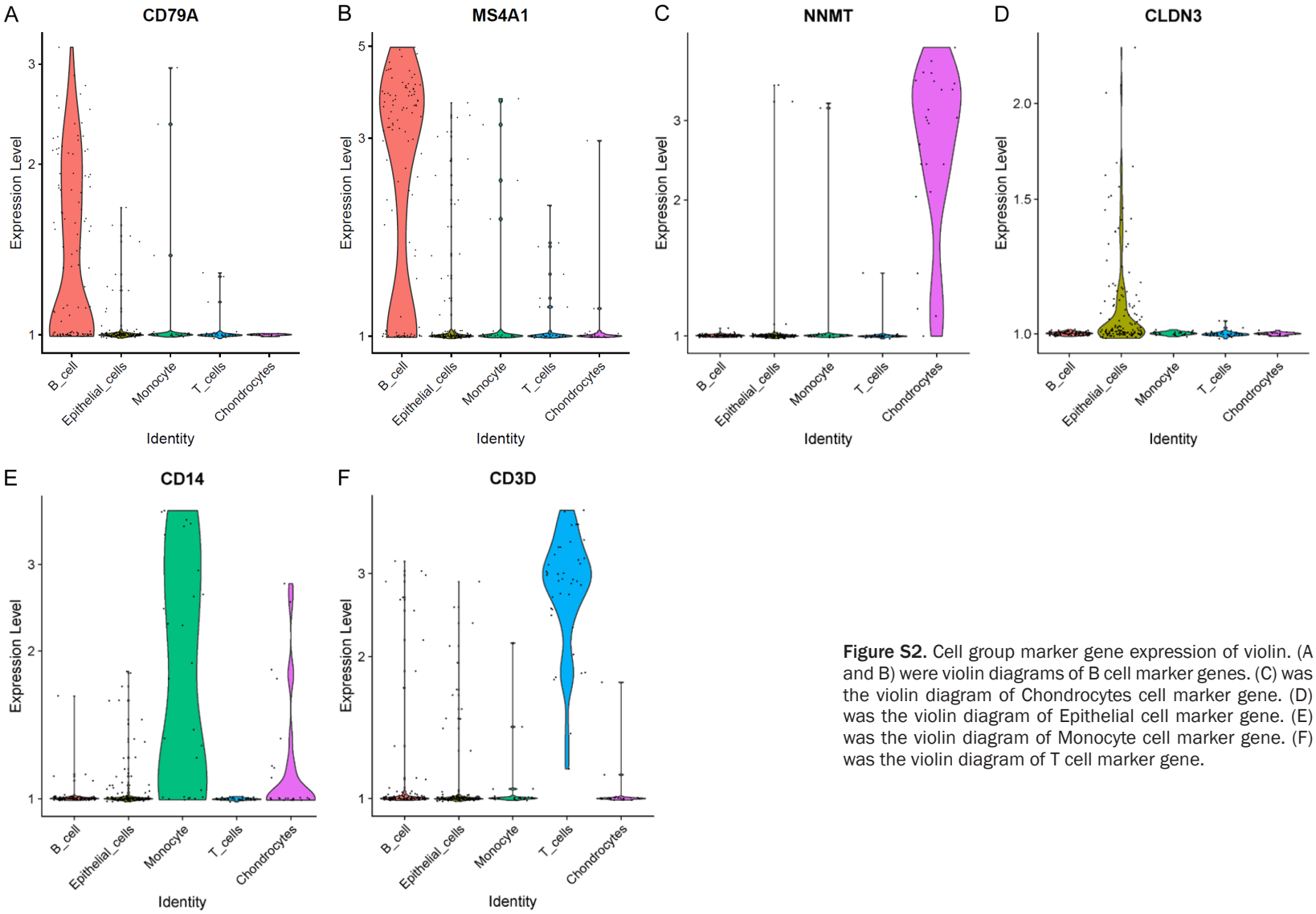


Figure S2. Cell group marker gene expression of violin. (A and B) were violin diagrams of B cell marker genes. (C) was the violin diagram of Chondrocytes cell marker gene. (D) was the violin diagram of Epithelial cell marker gene. (E) was the violin diagram of Monocyte cell marker gene. (F) was the violin diagram of T cell marker gene.

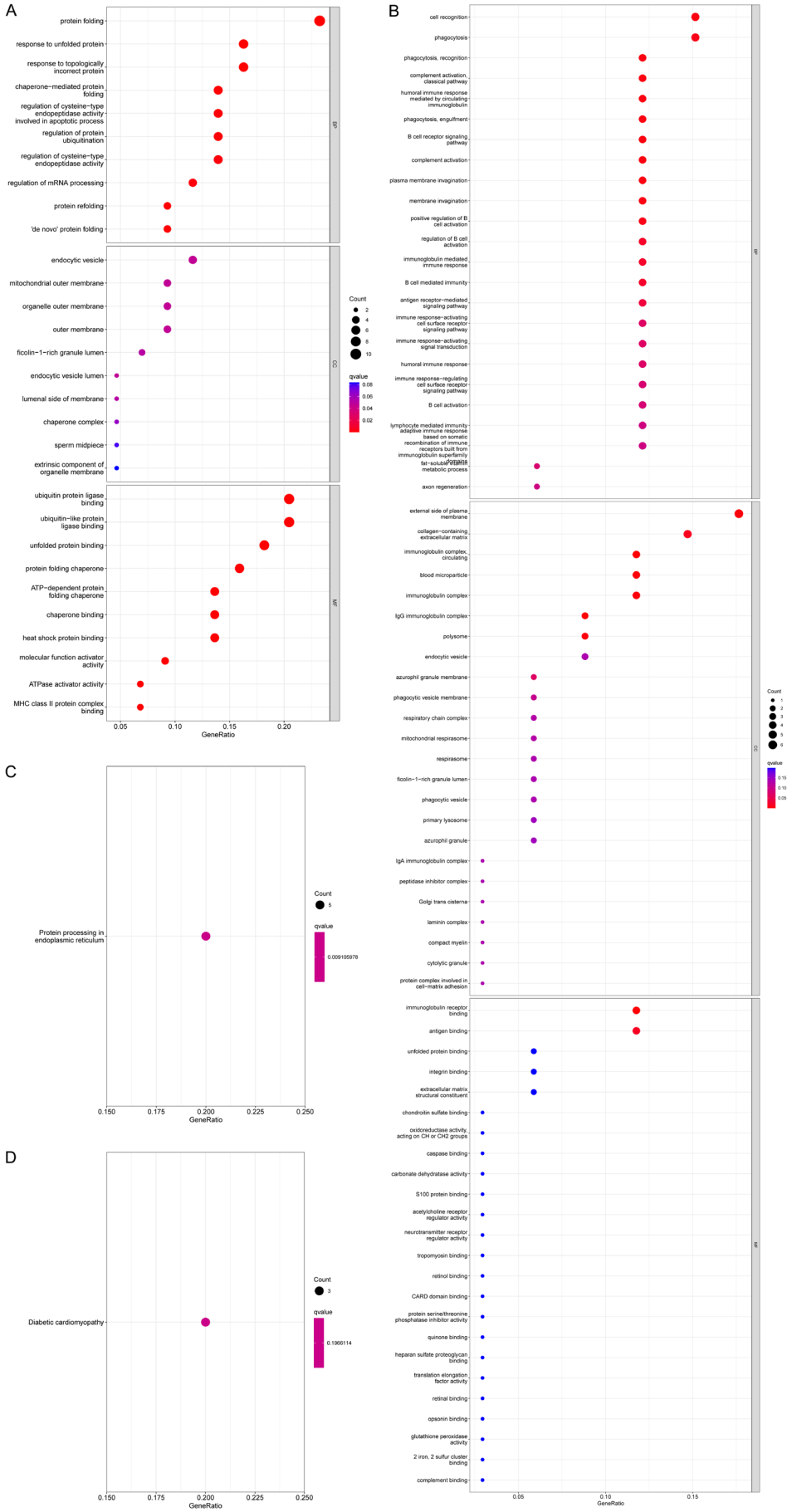
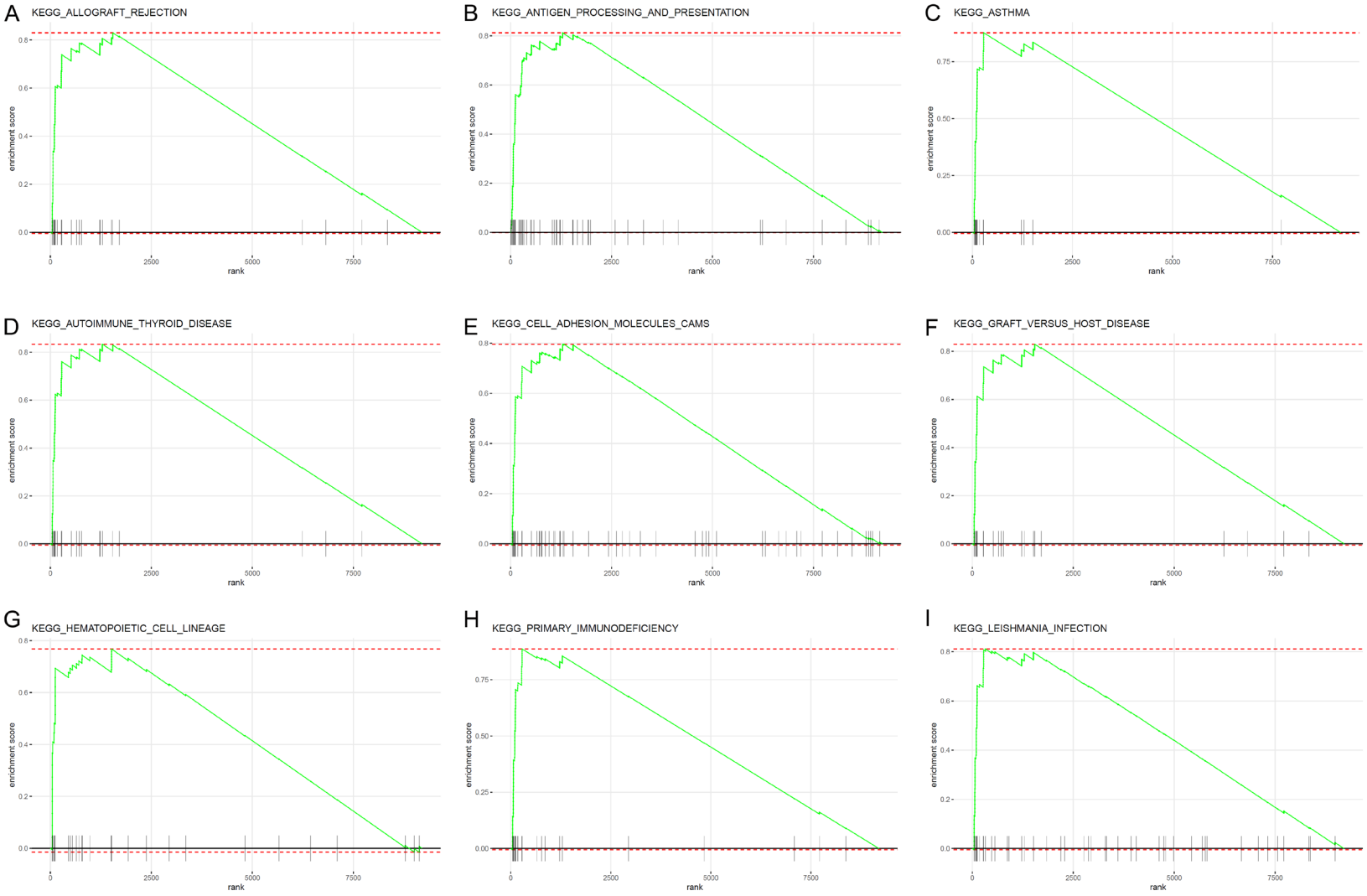

A
protein folding
response to unfolded protein
response to topologically incorrect protein
chaperone−mediated protein folding
regulation of cysteine−type endopeptidase activity involved in apoptotic process
regulation of protein ubiquitination
regulation of cysteine−type endopeptidase activity
regulation of mRNA processing
protein refolding
'de novo' protein folding
BP
endocytic vesicle
mitochondrial outer membrane
organelle outer membrane
outer membrane
ficolin−1−rich granule lumen
endocytic vesicle lumen
lumenal side of membrane
chaperone complex
sperm midpiece
extrinsic component of organelle membrane
CC
ubiquitin protein ligase binding
ubiquitin−like protein ligase binding
unfolded protein binding
protein folding chaperone
ATP−dependent protein folding chaperone
chaperone binding
heat shock protein binding
molecular function activator activity
ATPase activator activity
MHC class II protein complex binding
MF
0.05
0.10
0.15
0.20
GeneRatio
Count
2
4
6
8
10
qvalue
0.08
0.06
0.04
0.02
B
cell recognition
phagocytosis
phagocytosis, recognition
complement activation, classical pathway
humoral immune response mediated by circulating immunoglobulin
phagocytosis, engulfment
B cell receptor signaling pathway
complement activation
plasma membrane invagination
membrane invagination
positive regulation of B cell activation
regulation of B cell activation
immunoglobulin mediated immune response
B cell mediated immunity
antigen receptor−mediated signaling pathway
immune response−activating cell surface receptor signaling pathway
immune response−activating signal transduction
humoral immune response
immune response−regulating cell surface receptor signaling pathway
B cell activation
lymphocyte mediated immunity
adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily
fat−soluble vitamin metabolic process
axon regeneration
BP
external side of plasma membrane
collagen−containing extracellular matrix
immunoglobulin complex, circulating
blood microparticle
immunoglobulin complex
IgG immunoglobulin complex
polysome
endocytic vesicle
azurophil granule membrane
phagocytic vesicle membrane
respiratory chain complex
mitochondrial respirasome
respirasome
ficolin−1−rich granule lumen
phagocytic vesicle
primary lysosome
azurophil granule
IgA immunoglobulin complex
peptidase inhibitor complex
Golgi trans cisterna
laminin complex
compact myelin
cytolytic granule
protein complex involved in cell−matrix adhesion
CC
immunoglobulin receptor binding
antigen binding
unfolded protein binding
integrin binding
extracellular matrix structural constituent
chondroitin sulfate binding
oxidoreductase activity, acting on CH or CH2 groups
caspase binding
carbonate dehydratase activity
S100 protein binding
acetylcholine receptor regulator activity
neurotransmitter receptor regulator activity
tropomyosin binding
retinol binding
CARD domain binding
protein serine/threonine phosphatase inhibitor activity
quinone binding
heparan sulfate proteoglycan binding
translation elongation factor activity
retinal binding
opsonin binding
glutathione peroxidase activity
2 iron, 2 sulfur cluster binding
complement binding
MF
0.05
0.10
0.15
GeneRatio
Count
1
2
3
4
5
6
qvalue
0.15
0.10
0.05
C
Protein processing in endoplasmic reticulum
0.150
0.175
0.200
0.225
0.250
GeneRatio
Count
5
qvalue
0.009105978
D
Diabetic cardiomyopathy
0.150
0.175
0.200
0.225
0.250
GeneRatio
Count
3
qvalue
0.1966114

Figure S3. KEGG and GO analysis of differential genes among single cell samples with different NRGs expression levels. (A and B) were the GO enrichment analysis results of up-regulated genes and down-regulated genes, respectively. (C and D) were the results of KEGG enrichment analysis of up-regulated genes and down-regulated genes, respectively.



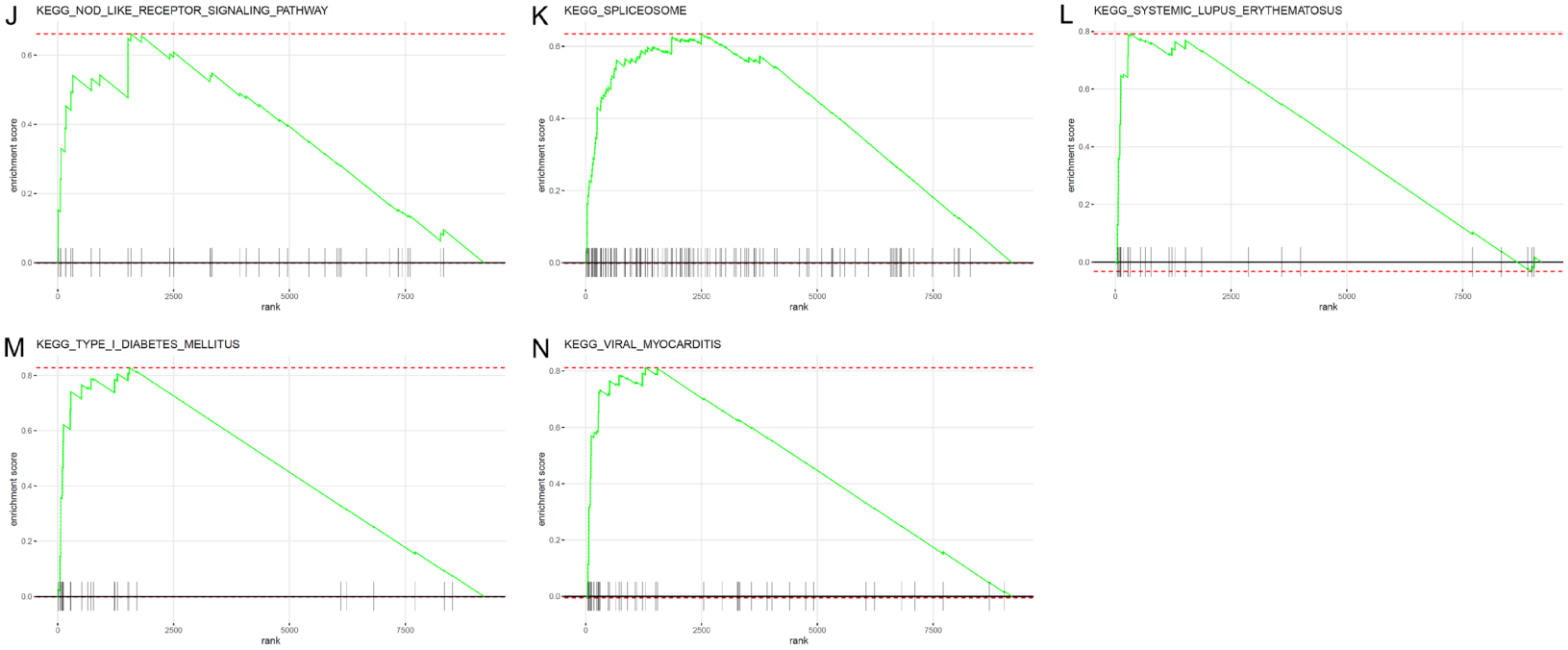


Figure S4. KEGG pathway information obtained from GSEA analysis of differential genes among single cell samples with different NRGs expression levels.

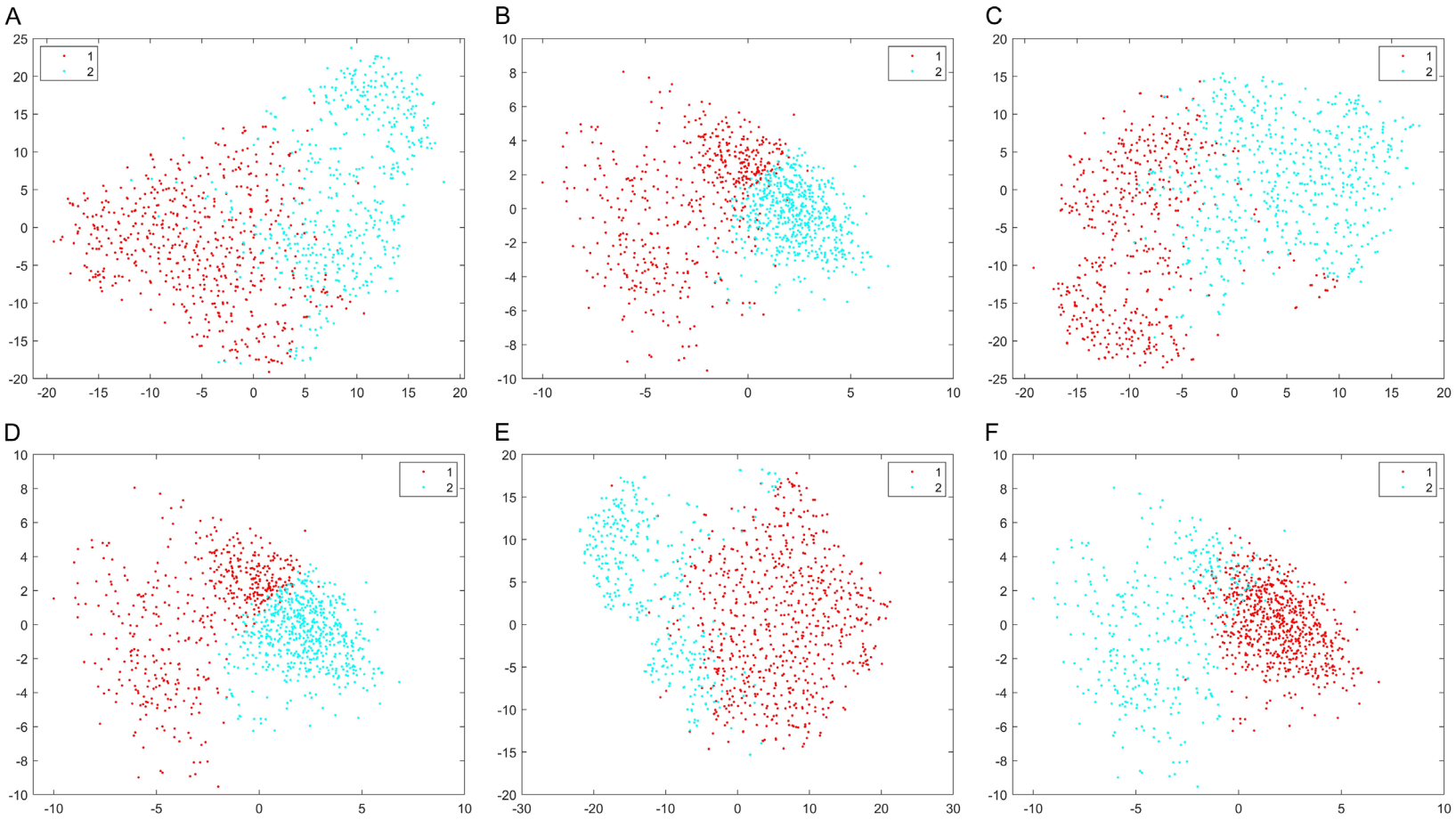


Figure S5. Visualization results of different algorithms clustering under the same experimental conditions (the number of clusters was 2). (A, C and E) were the clustering results of NMF, Deep semi-NMF, and DO-DSNMF obtained by using the t-sne algorithm, respectively. (B, D and F) were the clustering results of NMF, Deep semi-NMF, and DO-DSNMF obtained by the PCA algorithm, respectively.

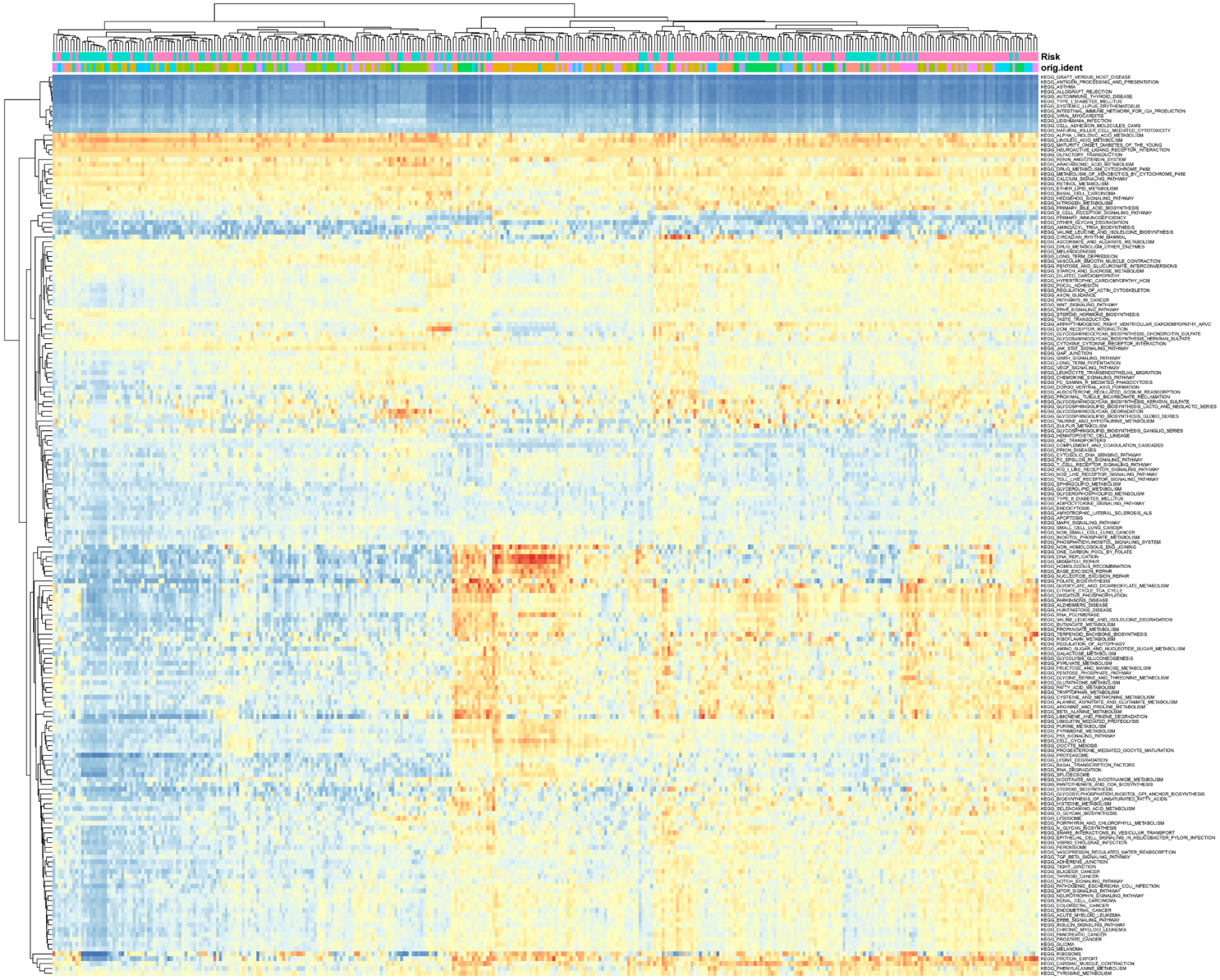


Figure S6. Results of GSEA analysis of cells in the high and low expression groups.