Mathematical Biosciences and Engineering

Research article

Four-lncRNA immune prognostic signature for triple-negative breast cancer

Running title: Immune lncRNAs predict prognosis of TNBC

Yun-xiang Li*, Shi-ming Wang and Chen-quan Li

Department of Breast Surgery, the First Hospital of Shanxi Medical University, Taiyuan 03001, China

* Correspondence: Email: 13453152386@163.com; Tel: +8603514867336; Fax: +8603514867336.

Abstract: *Objective:* We aimed to explore key immune-related long non-coding RNAs (lncRNAs) and their effect in predicting of prognosis of triple-negative breast cancer (TNBC). *Methods:* Four datasets of TNBC were downloaded from TCGA and GEO databases. ImmPort database was utilized to acquire immune-related mRNAs. Single sample gene set enrichment analysis (ssGSEA) and correlation analysis were utilized to screen immune-related lncRNAs. Univariate and multivariate Cox regression analyses were utilized to screen independent prognostic lncRNAs to establish prognostic risk model, and the model was evaluated by survival analysis and nomogram. Differential functions and immune cells infiltration in high and low risk group were analyzed by Gene set variation analysis and ssGSEA. Finally, competitive endogenous RNAs was constructed. *Results:* We revealed 62 immune-related lncRNAs, of which four lncRNAs (RP11-890B15.3, RP11-1024P17.1, MFI2-AS1 and RP11-180N14.1) had independent prognostic value. These four lncRNAs-based prognostic risk model could stratify the TNBC patients into high and low risk groups, and patients with high risk displayed unfavorable outcomes. Nomogram indicated that the prognostic model could indicate TNBC patients survival very well. We further found that high risk group showed significantly enriched immune response to tumor cell, humoral immune response and high infiltrating abundance of regulatory T cell, Type 2 T helper cell, eosinophil, etc. LncRNAs RP11-180N14.1, RP11-1024P17.1 and RP11-890B15.3 regulated more mRNAs by targeting various miRNAs. While MFI2-AS1 regulated three mRNAs by sponging miR-3150a-3p. *Conclusion*: These four lncRNAs were prognostic biomarkers and could be possible therapeutic targets in TNBC.

Keywords: triple-negative breast cancer; long non-coding RNAs; prognosis, immune cells;

competitive endogenous RNAs

1. Introduction

Breast cancer (BC) is one of the most common malignant tumors for women in the world [1]. The global incidence of BC had 1,960,681 incident cases and caused 181,004 deaths in 2017 and the incidence tends to be younger, causing serious troubles to the majority of women [2]. Triple-negative BC (TNBC) refers to the BC with estrogen receptor (ER)-negative, progesterone receptor (PR)-negative and human epidermal growth factor receptor2 (HER-2)-negative, which accounts for about 15% of BC [3]. TNBC has histological grade III or higher histologic grade and is more aggressive than other BC subtypes [4]. Additionally, TNBC is resistant to current HER2-targeted therapy, and chemotherapy is still the most effective method, but its long-term clinical effect is not ideal, with poor prognosis and high recurrence rate [5]. Immunotherapy enhances the tumor microenvironment of antitumor immunity to eliminate cancer cells through the stimulation or mobilization of the immune system of body [6]. Currently, immunotherapy has gained wide attention in terms of tumor therapy.

Long non-coding RNAs (lncRNAs) act as major regulators of gene expression, and their mutations and misregulation have been found to be associated with a variety of biological processes in many diseases and cancers [7]. Especially, lncRNAs may act as competitive endogenous RNAs (ceRNAs) by competitively binding to microRNA (miRNAs) to mediate the expression of their downstream target genes in cancers [8]. For example, lncRNA HCP5 was implicated in pancreatic cancer progression and prognosis by HCP5-miR-29b-3p-MMP9/ITGB1 ceRNA regulatory axis [9]. Due to the important role of lncRNAs in cancers, lncRNAs exert promising application prospect as novel biomarkers and therapeutic targets for cancers, including BC. Dong et al. found that lncRNA TINCR could facilitate trastuzumab resistance and epithelial-mesenchymal transition process by sponging miR-125b to release HER-2 in BC cells, and its high expression showed correlations with worse response and prognosis for HER-2+ BC patients with trastuzumab therapy [10]. Based on systematic bioinformatics analysis, Fan et al. identified that four-lncRNA signature could be prognostic biomarker to independently predict the overall survival of BC patients [11]. Ma et al. constructed a 8 immune-related lncRNAs for the prediction of BC patient survival [12]. De Palma et al. suggested that long intergenic non-coding RNA 01087 could represent a noval biomarker for diagnosis and prognosis of BC and TNBC [13]. However, these findings were investigated based on small sample size or single dataset.

Besides, current studies suggested that lncRNAs are implicated in the tumor immune microenvironment and mediate tumor immune evasion [14,15]. Considering the complexity and heterogeneity of tumor immune microenvironment and the importance of lncRNAs, it is needed to identify potential lncRNAs biomarkers on the basis of the landscape of tumors infiltrate immune cells. Therefore, a integrated analysis combining tumor immune microenvironment and lncRNAs was designed. In addition, only TNBC samples were included, which eliminated the interindividual differences. In this study, we identified immune-related lncRNAs that had independent prognostic value in TNBC by integrating analysis of four datasets from The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO) databases. On the basis of these lncRNAs, we then constructed a prognostic risk model, which could stratify TNBC patients with different prognosis, and the

underlying regulatory mechanism was further investigated. This study will contribute to improve the prognosis prediction and treatment of TNBC.

2. Methods

2.1. Study design and datasets acquisition

A total of four datasets were analyzed in this study, and the analysis procedure was shown in Figure 1. The log2 (FPKM+1) expression data and matched clinical data of breast invasive carcinoma (BRCA) in TCGA were acquired from UCSC Xene database (http://xena.ucsc.edu/). The platform of data was Illumina HiSeq 2000 RNA Sequencing. Based on the inclusion criteria of samples with ER-negative, PR-negative and HER-2-negative, a total of 8 TNBC samples had both tumor tissue and paracancerous tissue were screened. GSE115275 dataset (containing 6 TNBC samples and 6 adjacent nontunor samples) was downloaded from GEO database [16], and the platform of was Agilent-079487 Arraystar Human LncRNA microarray V4 (Probe Name version). GSE115275 was screened with search terms of "Triple negative breast cancer" and selected by inclusion criteria as follows: 1) expression profiling by array as the study type; 2) organism was homo sapiens; 3) publication dates were limited within one year; 4) samples had both tumor tissue and paracancerous tissue; 5) microarray was lncRNA chip. These two datasets (TCGA-TNBC dataset and GSE115275) were used for differential expression analysis.

Datasets used in prognostic model construction were screened by criteria as follows: 1) expression profiling by array as the study type; 2) organism was homo sapiens; 3) publication dates were limited within one year; 4) samples had corresponding data about survival; 5) annotated more than 3000 lncRNAs. GSE135565 was screened with search terms of "Triple negative breast cancer" and selected from GEO database, which contained 84 TNBC samples, and the data platform was GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Additionally, GSE58812 (containing 107 TNBC samples) was screened with search terms of "Triple negative breast cancer" and selected as validation dataset for prognostic model. This dataset was screened with criteria as follows: 1) expression profiling by array as the study type; 2) organism was homo sapiens; 3) publication dates from January 1, 2015 to December 30, 2020; 4) samples had corresponding data about survival; 5) platform was same with GSE135565. The clinical information of the samples in the four datasets was listed in Table 1.

2.2. Data preprocessing and differential expression analysis

For TCGA-TNBC dataset, the Ensemble Gene was converted into gene symbol to obtain mRNAs and lncRNAs utilizing the annotation file (hg38, gencode.v22.annotation.gene.probemap) of GENCODE database [17]. Only the mRNAs and lncRNAs with expression in more than 20% samples were included. For GEO datasets (GSE115275, GSE135565 and GSE58812), the probes were aligned to Reference genome (hg38, gencode.v22) by seqmap software [18] and were annotated to obtain mRNAs and lncRNAs based on the Release22 annotation file in GENCODE database. Probes matching no gene symbol was removed, and mean value was considered as the final expression value of the mRNAs/lncRNAs when different probes matching to one gene symbol.

Differential expression analysis between tumor vs. normal was carried out on the TCGA-TNBC

and GSE115275 datasets utilizing the linear regression and empirical bayesian methods of limma package (Version 3.10.3) [19]. Differentially expressed mRNAs (DEmRNAs) and lncRNAs (DElncRNAs) were screened with $P < 0.05$ and $|logFC| > 0.585$. The overlapped up-regulated and down-regulated DEmRNAs and DElncRNAs were utilized in the following analysis.
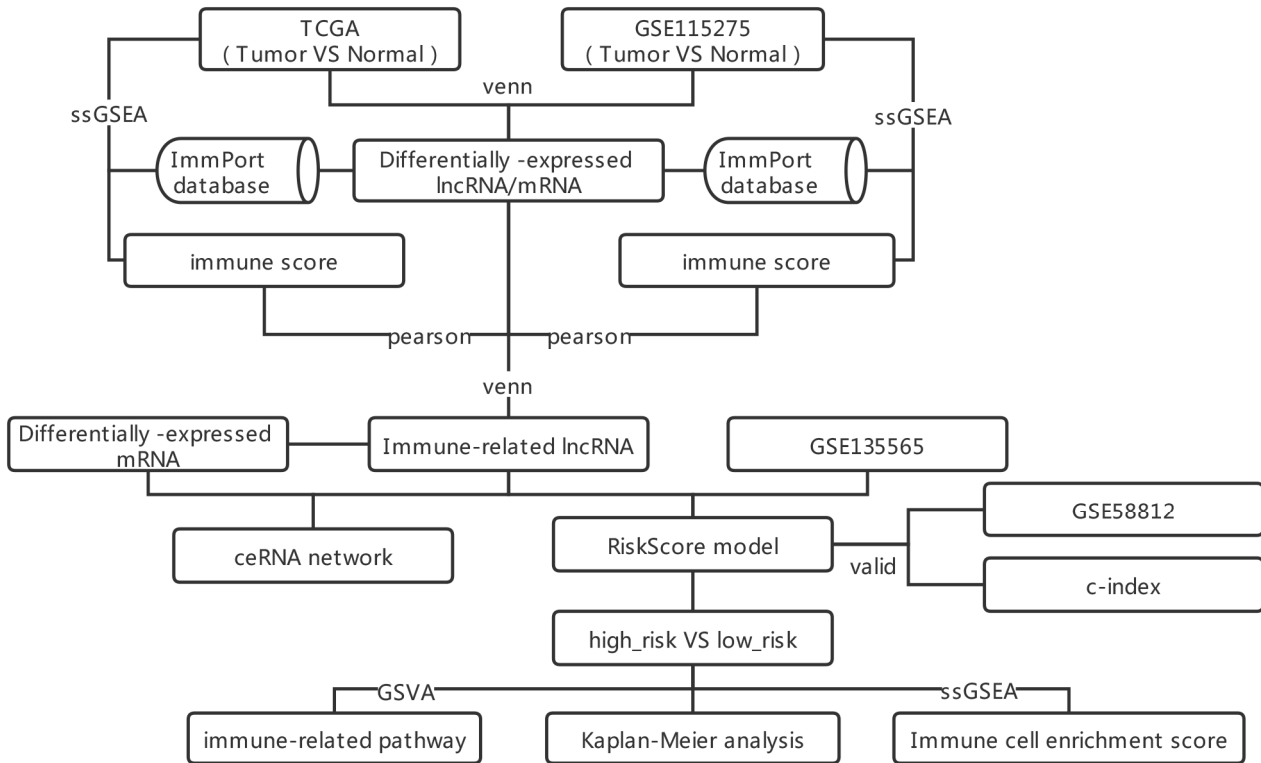

Figure 1. The workflow of this study.

Table 1. The clinical information of the samples in the four datasets.

	TCGA-TNBC	GSE115275	GSE135565	GSE58812
Gender (female)	8	6	84	107
Age (mean ± SD)	51.3 ± 13.3	–	50.2 ± 10.5	57.0 ± 12.8
AJCC stage (1/2/3/4)	2/4/2/0	–	29/50/5/0	–
SUV (mean ± SD)	–	–	7.1 ± 4.8	–
Vital status (Alive/Dead)	6/2	–	77/7	78/29
metastasize status (meta/non-meta)	0/8	–	–	31/76

2.3. Screening of immune-associated lncRNAs

Immune-associated mRNAs were retrieved from ImmPort database [20], and then the calculation of immune score for samples in TCGA-TNBC and GSE115275 datasets was conducted utilizing single sample gene set enrichment analysis (ssGSEA) [21]. Then, the pearson correlation coefficient between DElncRNAs and immune score was calculated. The immune-associated lncRNAs were screened by Benjamini-Hochberg adjusted $P < 0.01$ and $|r| > 0.3$. The overlapped immune-related lncRNAs in TCGA-TNBC and GSE115275 datasets were selected.

2.4. Immune-associated prognostic lncRNAs

GSE135565 dataset was utilized to screen immune-related prognostic lncRNAs. In brief, univariate Cox regression analysis was utilized to explore the correlations of immune-associated lncRNAs with overall survival (OS) and OS time, then the hazard ratio (HR) and P value were obtained. In which, immune-associated lncRNAs with P < 0.05 were regarded as immune- associated prognostic lncRNAs. On the basis of the optimal cutpoint determined by survminer (version 0.4.3) in R package, we categorized samples into high or low expression groups, and Kaplan-Meier (KM) survival curves as well as log-rank test were utilized to perform survival analysis by survival (version 2.42-6) in R package.

2.5. Prognostic risk model construction and validation

Multivariate Cox regression analysis was carried out for the immune-associated prognostic lncRNAs to investigate the correlation coefficient of these lncRNAs with OS and OS time. Then the prognostic risk model was constructed according to the formula: Risk score = βgene1 × exprgene1 + βgene2 × exprgene2 + ... + βgenen × exprgenen. In the formula, β represents the correlation coefficient in multivariate Cox regression analysis, and expr represents the expression value of the lncRNAs. We categorized samples into high or low risk groups on the basis of the optimal cutpoint determined by survminer (version 0.4.3) in R package, and KM survival curves as well as log-rank test were utilized to perform survival analysis by survival (version 2.42-6) in R package. GSE58812 dataset was utilized to validate the prognostic risk model. Additionally, GSE58812 contained the information of metastasis free survival (MFS) and metastasize status. We further constructed metastatic risk model based on the lncRNAs of prognostic risk model. We categorized samples into high or low risk groups on the basis of the optimal cutpoint, followed by MFS survival analysis. The selection of optimal cutpoint and the methods for survival analysis were same to the details described above. Then the predivtive and distinguishing ability of the risk model within 3, 5 and 10 years was evaluated using receiver operating characteristic (ROC) curve analysis.

2.6. Nomogram model construction

In order to investigate whether the prognostic risk model had the independent value for predicting prognosis, univariate and multivariate Cox regression analysis were carried out on GSE135565 dataset. First, univariate Cox regression analysis was performed for clinical variables and risk group, in which, the variables with P < 0.05 were further enrolled in multivariate Cox regression analysis, and factors with P < 0.05 were further selected to construct nomogram model.

2.7. Differential immune-related GO-BP in high vs. low risk groups in GSE135565 dataset

The immune-related Gene Ontology-biological processes (GO-BP) and gene sets were searched from c5.bp.v7.1.symbols.gmt gene set of MSIGdb v7.1 [22] by term of "immune". Then Gene set variation analysis (GSVA) [23] was utilized to calculate the enrichment score of each biological process to obtain an enrichment score matrix. Differential analysis for immune-related GO-BP in high vs. low risk groups were performed utilizing limma package. Benjamini-Hochberg was utilized

for multiple-testing correction. Adjusted P < 0.05 was utilized to screen differential immune-related GO-BP. The heatmap was plotted utilizing pheatmap (version: 1.0.12) in R package.

2.8. Immune cells infiltration in high vs. low risk groups in GSE135565 dataset

The gene sets for labeling tumor infiltrating immune cells in tumor microenvironment were obtained from the study of Charoentong et al. [24]. Subsequently, on the basis of ssGSEA algorithm, the enrichment score of each immune cell type in samples were calculated to show the abundance of tumor infiltrating immune cells. Boxplot was visualized utilizing ggplot2 (version: 3.2.1) in R package. Wilcox test was utilized to compare the difference in high vs. low risk groups.

2.9. LncRNA-miRNA-mRNA network

Correlation analysis between lncRNAs in prognostic risk model and DEmRNAs were performed. After Benjamini-Hochberg correction, lncRNA-mRNA pairs with adjusted P < 0.01 and correlation coefficient r > 0 were selected. The clusterProfiler [25] in R package was utilized to explore the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway of mRNAs to predict the function of lncRNAs. KEGG pathways with P < 0.05 and enrichment gene count ≥ 2 were considered as significantly enriched terms. In addition, DIANA-LncBase v2 [26] was utilized to predict the lncRNA-miRNA pairs based on the lncRNAs in prognostic risk model, and the lncRNA-miRNA pairs with score > 0.9 was selected. The miRNA-mRNA pairs were predicted utilizing miRWalk2.0 database [27], which contained the data in miRWalk, Microt4, miRanda, miRDB, PITA, RNA22, RNAhybrid and Targetscan databases. The miRNA-mRNA interactions existed in at least five databases were selected. Finally, the lncRNA-miRNA-mRNA interactions axis were integrated by the obtained interactions pairs of lncRNA-mRNA, lncRNA-miRNA as well as miRNA-mRNA, and the lncRNA-miRNA-mRNA network was built utilizing Cytoscape software (version:3.4.0) [28].

3. Results

3.1. Screening of DEmRNAs and DElncRNAs

From the TCGA-TNBC dataset, a total of 3884 DEmRNAs and 412 DElncRNAs were screened between tumor vs. normal tissue samples. While for GSE115275 dataset, there were 1559 DEmRNAs and 1978 DElncRNAs between tumor vs. normal tissue samples. Table 2 listed the number of up-regulated and down-regulated DEmRNAs and DElncRNAs in the two datasets. After merging, a total of 128 overlapped DElncRNAs (Supplementary Table 1) (43 up-regulated and 85 down-regulated) and 436 overlapped DEmRNAs (Supplementary Table 2) (182 up-regulated and 254 down-regulated) were obtained (Figure 2A,B).

3.2. Immune-associated lncRNAs

We firstly retrieved 1181 immune-associated genes from ImmPort database, and followed by the calculation of immune score for samples in TCGA-TNBC and GSE115275 datasets utilizing

ssGSEA. Then, the correlation coefficient between the 128 overlapped DElncRNAs and immune score was calculated, and a total of 71 and 103 immune- associated lncRNAs were obtained with adjusted $P < 0.01$ and $|r| > 0.3$ from TCGA-TNBC and GSE115275 datasets, respectively. Of which, there were 62 common immune-related lncRNAs (Figure 2C; Supplementary Table 3).
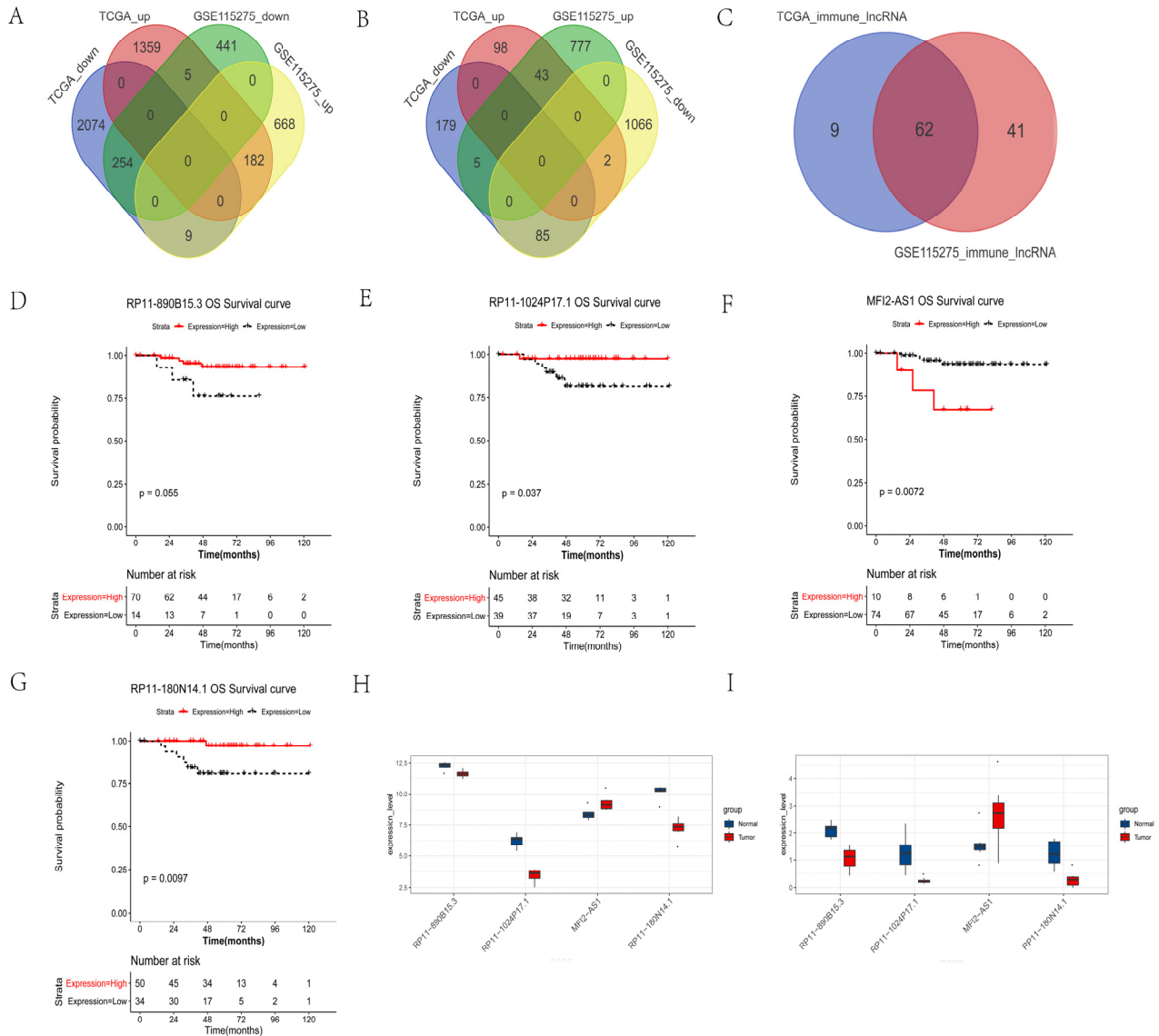

Figure 2. Screening of immune-related prognostic lncRNAs. Venn plots show the overlapped DEmRNAs (A), DElncRNAs (B) and immune-related lncRNAs (C) in TCGA-TNBC and GSE115275 datasets; KM survival curves show the prognostic value of the four immune-related prognostic lncRNAs RP11-890B15.3 (D), RP11-1024P17.1 (E), MFI2-AS1 (F) and RP11-180N14.1 (G); Boxplots show the expression levels of the four lncRNAs in GSE115275 dataset (H) and TCGA cohort (I). DEmRNAs, differentially expressed mRNAs; DElncRNAs, differentially expressed long non-coding RNAs; TCGA: The Cancer Genome Atlas; TNBC: triple-negative breast cancer; K-M: Kaplan-Meier; lncRNA: long non-coding RNA.

Table 2. The number of up-regulated and down-regulated DEmRNAs and DElncRNAs in TCGA-TNBC and GSE115275 datasets.

		up	down	total
TCGA-TNBC	mRNA	1546	2338	3884
	lncRNA	143	269	412
GSE115275	mRNA	859	700	1559
	lncRNA	825	1153	1978

3.3. Immune-related prognostic lncRNAs

GSE135565 dataset was utilized to screen immune-related prognostic lncRNAs. Among the 62 immune-related lncRNAs, only 28 lncRNAs were annotated in GSE135565 dataset (Supplementary Table 3). Univariate Cox regression analysis showed that RP11-890B15.3 (hazard ratio [HR] = 0.124, P = 0.027), RP11-1024P17.1 (HR = 0.110, P = 0.027) and RP11-180N14.1 (HR = 0.003, P = 0.035) overexpression levels were significantly associated with a good prognosis, and MFI2-AS1 (HR = 7.642, P = 0.030) higher expression was correlated with a poor prognosis (Table 3). In addition, survival analysis showed higher expression of RP11-890B15.3 (84 samples, P = 0.055), RP11-1024P17.1 (84 samples, P = 0.037) and RP11-180N14.1 (84 samples, P = 0.0097) were associated with favorable prognosis. While high expression of MFI2-AS1 (84 samples, P = 0.0072) was associated with worse prognosis (Figure 2D–G). Additionally, the expression levels of these four lncRNAs were performed in GSE115275 dataset and TCGA cohort, and the results shown that RP11-890B15.3, RP11-1024P17.1 and RP11-180N14.1 were downregulated, and MFI2-AS1 was upregulated (Figure 2H,I). Therefore, these four lncRNAs (RP11-890B15.3, RP11-1024P17.1, MFI2-AS1 and RP11-180N14.1) were considered as immune-related prognostic lncRNAs.

3.4. Prognostic risk model construction and validation

The four lncRNAs associated with both immune and prognosis were utilized to establish prognostic risk model (OS risk model). As described in method, GSE135565 dataset was utilized as training set and GSE115275 was utilized as validation set. The samples were categorized into high or low risk groups on the basis of the optimal cutpoint. It could be seen from Figure 3A, the risk score could divide patients into two groups with different risk, and patients with high risk tended to display unfavorable outcomes. The heatmap of the four lncRNAs showed that the expression of RP11-890B15.3, RP11-1024P17.1 and RP11-180N14.1 were increased in high risk group, while the expression of MFI2-AS1 was decreased in high risk group. The K-M curves confirmed that patients with high risk had poor OS than patients with low risk both in training set (Figure 3B) and validation set (Figure 3C). In addition, we also constructed MFS risk score utilizing the information of MFS and metastasize status in GSE115275 dataset. We categorized samples into high or low risk groups on the basis of the optimal cutpoint, and K-M curve showed patients with high risk had poor MFS than that of low risk group (Figure 3D). The AUC values of survival prediction within 3, 5 and 10 years were 0.814, 0.817 and 0.817, respectively, suggesting the favorable predictive ability of the lncRNA model (Figure 3E). The results suggested that these four lncRNAs could be predictive biomarkers for both OS and MFS in TNBC.
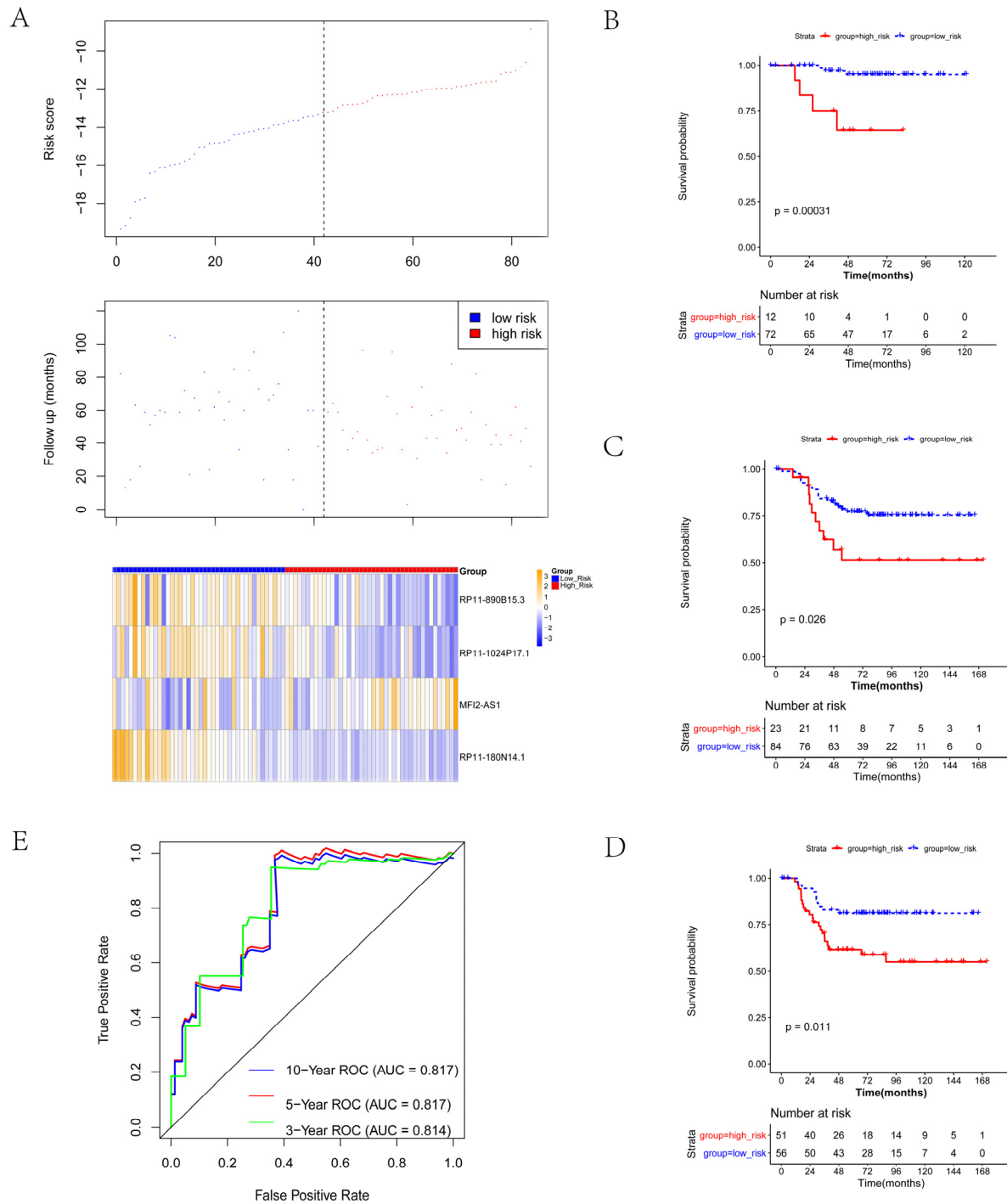
Figure 3. Prognostic performance of the prognostic risk model. Prognostic risk score distribution (upper panel), overall survival time distribution scatter plot of samples (middle panel) and heatmaps of methylation level pattern of the four lncRNAs with prognostic risk score changes (lower panel) in GSE135565 dataset (A); K-M survival curves show the OS of patients with high and low risk score in GSE135565 training set (B) and GSE115275 validation set (C); K-M survival curve shows the MFS of patients with high and low risk score in GSE115275 validation set (D); The ROC curves based on risk score level (E). K-M: Kaplan-Meier; OS: overall survival.

Table 3. Survival analysis and univariate Cox regression analysis of 4 immune-related lncRNAs.

	K-M analysis	Univariate Cox regression analysis				Differential expression analysis
lncRNA	p.value	HR	lower.95	upper.95	p.value	up_down
RP11-890B15.3	0.055	0.123567	0.019408	0.786726	0.026832	down_lnc
RP11-1024P17.1	0.037	0.110413	0.015648	0.779107	0.027081	down_lnc
MFI2-AS1	0.0072	7.642445	1.222606	47.77251	0.029637	up_lnc
RP11-180N14.1	0.0097	0.002625	1.05E-05	0.659276	0.035054	down_lnc

3.5. Nomogram model construction

In order to investigate whether the prognostic risk model had the independent value for predicting prognosis, univariate and multivariate Cox regression analysis were carried out. The results showed that risk group (P = 0.003) and AJCC stage (P = 0.008) had significant impact on prognosis in univariate analysis. Multivariate analysis indicated that AJCC stage and risk group were all independent prognostic factors in TNBC with a hazard ratio of 18.327 (P = 0.008) and 17.619 (P = 0.013), respectively (Table 4). We further constructed a nomogram model to predict the 3-year, 5-year, and 10-year survival of the TNBC patients utilizing independent prognostic factors risk group and AJCC stage. The results indicated that prognostic model could predict the survival for TNBC patients very well (Figure 4).
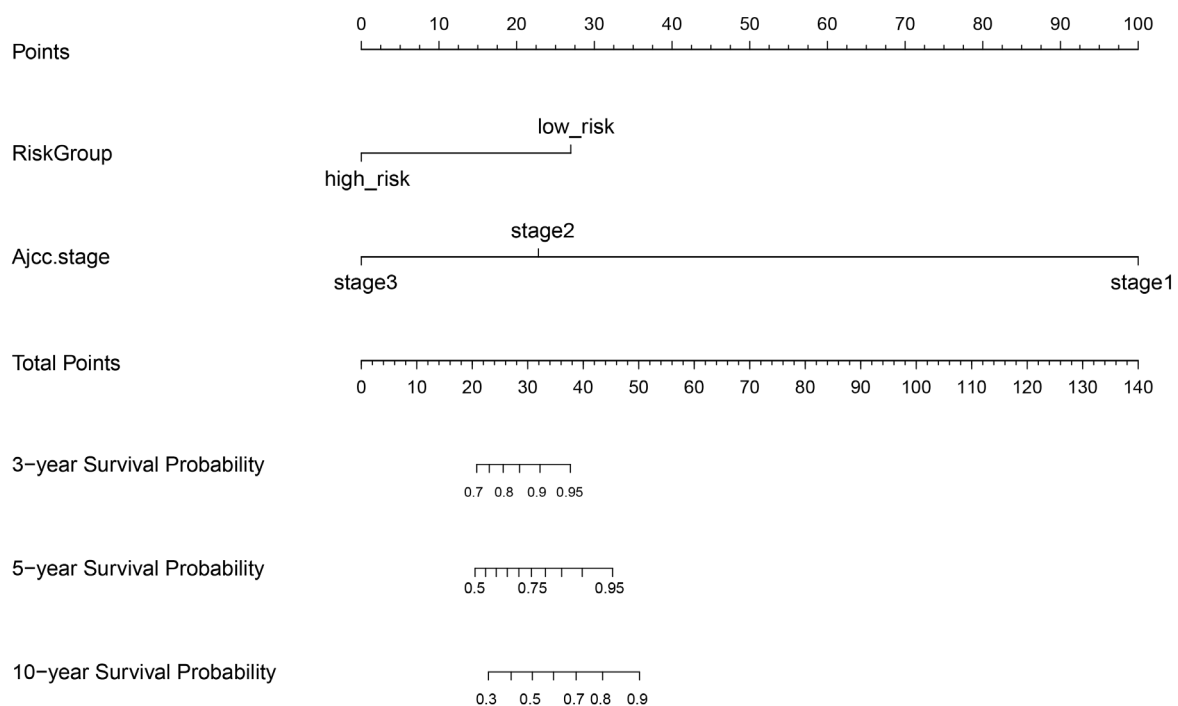

Figure 4. Nomogram model. Nomogram model was constructed to predict the 3-year, 5-year, and 10-year survival of the TNBC patients utilizing independent prognostic factors risk group and AJCC stage. TNBC: triple-negative breast cancer; AJCC: American Joint Committee on Cancer.

Table 4. Univariate and multivariate Cox regression analysis of the clinical factors.

	Univariate		Multivariate	
Variables	HR	p.value	HR	p.value
Age	1.000	0.990		
AJCC stage	5.705	0.008	18.327	0.008
Suv	0.934	0.483		
RiskGroup	9.633	0.003	17.619	0.013

3.6. Differential immune-related GO-BP in high vs. low risk groups

A total of 84 immune-related GO-BP and gene sets were obtained, and then GSVA was utilized to calculate the enrichment score of each biological process to obtain an enrichment score matrix. After differential analysis in high vs. low risk groups, a total of eight immune-related GO-BPs were significant differential (Figure 5; Supplementary Tables 4 and 5). Of which, low risk group was involved in 5 GO-BPs, for example, activation of innate immune response. Differently, high risk group was obviously implicated in three GO-BPs, for instance, immune response to tumor cell.
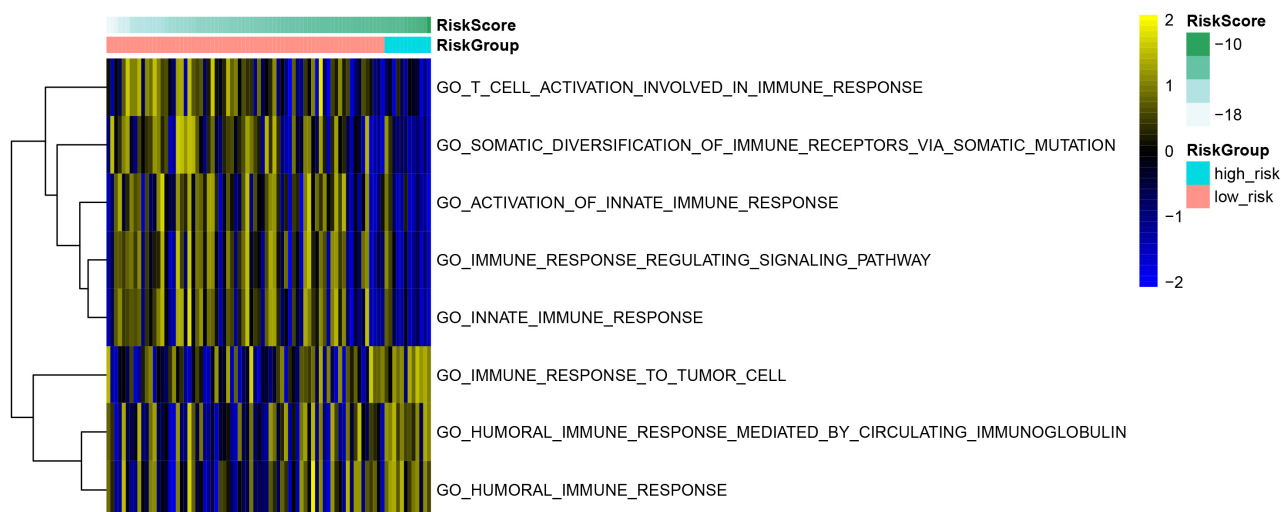

Figure 5. Gene set variation analysis. Heatmap show the differential immune-related GO-BP in high risk and low risk groups in gene set variation analysis. The raw color from blue to yellow represents significant from low to high. GO-BP: Gene ontology-biological process.

3.7. Immune cells infiltration in high vs. low risk groups

The enrichment score of 23 immune cell types were calculated to show the abundance of tumor infiltrating immune cells. As shown in Figure 6, infiltrating abundance of nine immune cell types had significant differences in high vs. low risk groups. Of which, activated dendritic cell, eosinophil, immature B cell, plasmacytoid dendritic cell, regulatory T cell (Treg cell) and Type 2 T helper cell (Th2) had high infiltrating abundance in high risk group, while CD56dim natural killer cell, monocyte and neutrophil had high infiltrating abundance in low risk group.
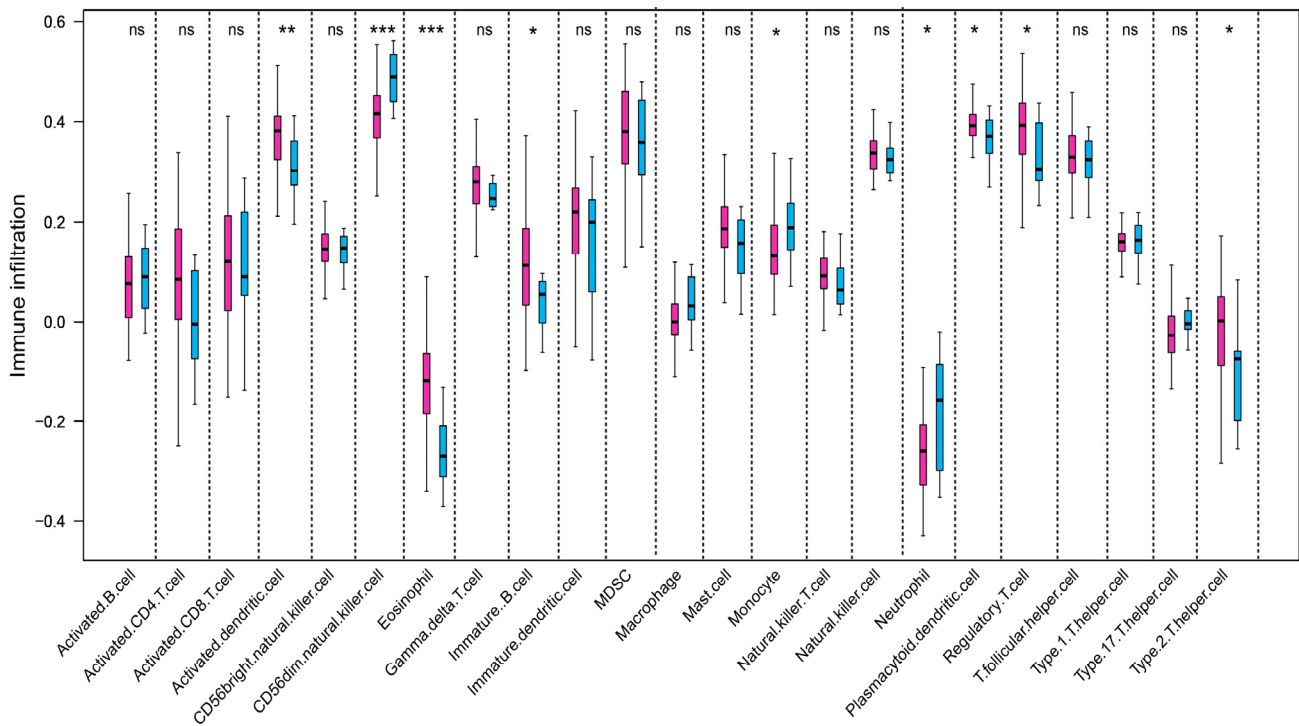
Figure 6. Boxplot of tumor infiltrating immune cells. Boxplot shows the infiltrating abundance of 23 tumor infiltrating immune cells in high risk and low risk groups. Red box represents high risk group, and blue bos represents low risk group.*P < 0.05; **P < 0.01; ***P < 0.001; ns: not significant.

3.8. ceRNA network

Correlation analysis was performed for the four lncRNAs and DEmRNAs, and 110 positive lncRNA-mRNA pairs with P < 0.01 and correlation coefficient r > 0 were selected (Supplementary Table 6). Enrichment analysis was further performed for the co-expressed mRNAs to predict the function of lncRNAs. Significant pathways were enriched for three lncRNAs, including RP11-180N14.1, MFI2-AS1 and RP11-890B15.3 (Figure 7A; Supplementary Table 7). It could be seen that these three lncRNAs were associated with different KEGG pathways. MFI2-AS1 was mainly associated with apoptosis and protein processing in endoplasmic reticulum. RP11-180N14.1 was mainly involved in ascorbate and aldarate metabolism, pentose and glucuronate interconversions, drug metabolism-other enzymes, etc. While RP11−890B15.3 was mainly implicated in ABC transporters and breast cancer pathways. Additionally, 62 lncRNA-miRNA pairs were obtained for the four lncRNAs, and 13852 miRNA-mRNA interactions were obtained for the mRNAs in lncRNA-mRNA pairs. Finally, 324 lncRNA-miRNA-mRNA interactions axis was integrated, containing 4 lncRNAs, 42 miRNAs and 48 mRNAs (Figure 7B; Supplementary Tables 8 and 9). LncRNAs RP11-180N14.1, RP11-1024P17.1 and RP11-890B15.3 regulated more mRNAs by targeting various miRNAs. While MFI2-AS1 regulated three mRNAs [TNF receptor associated factor 2 (TRAF2), suppressor APC domain containing 2 (SAPCD2) and G protein regulated inducer of neurite outgrowth 1 (GPRIN1)] by sponging miR-3150a-3p.
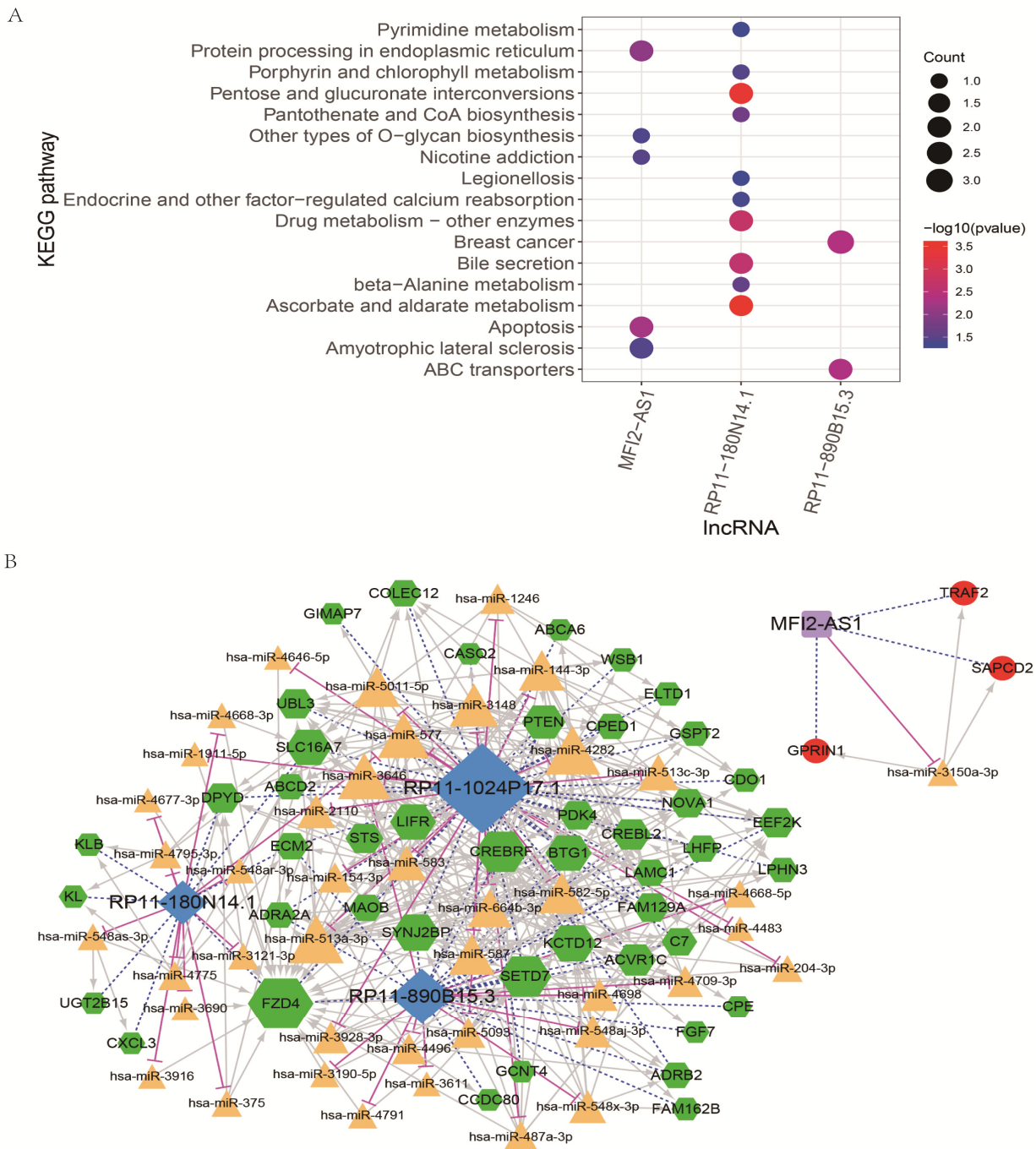
Figure 7. LncRNA-miRNA-mRNA network. (A), bubble diagram shows the significant enriched KEGG pathways of co-expressed mRNAs for lncRNAs; bubble size represents gene count; bubble color from blue to red represent significant from low to high; (B), lncRNA-miRNA-mRNA network. Yellow triangle represents miRNA; blue rhombus represents down-regulated lncRNA, and purple square represents up-regulated lncRNA; green hexagon represents down-regulated mRNA, and red node represents up-regulated mRNA; imaginary line represents the co-expression between lncRNA and mRNA; grey arrow represents miRNA-mRNA pairs and purple line represent lncRNA-miRNA pairs. KEGG: Kyoto Encyclopedia of Genes and Genomes; lncRNA: long non-coding RNA; miRNA: microRNA.

4. Discussion

In recent years, great efforts have been made to explore the pathogenesis and improve the clinical outcomes of TNBC. However, the long-term clinical effect of TNBC is not ideal, and the prognosis remains worse. Accurate prognosis has been proved to be helpful for healthcare decisions and developing individualized therapy strategy [29]. Therefore, increasing prognostic biomarkers are identified, which can help to stratify the patients by predicting the potential outcome of tumor-related progression, relapse, or survival [30,31]. Immunotherapy has becoming a powerful treatment strategy for cancers, and the recent clinical successes emphasize the significance of understanding tumor immunity to the clinical translation in cancers treatment [32]. In this study, based on integrate analysis of four datasets from TCGA and GEO databases, we identified 62 immune-related lncRNAs, of which four lncRNAs (RP11-890B15.3, RP11-1024P17.1, MFI2-AS1 and RP11-180N14.1) with independent prognostic value were identified. These four lncRNAs-based prognostic risk model could stratify the TNBC patients into two groups with different risk, and patients in high risk group displayed unfavorable worse outcomes.

In order to investigate the underlying mechanism, we firstly investigated the differences on immune-related GO-BP and tumor-infiltrating immune cells in high vs. low risk groups. We found that high risk group showed significantly enriched immune response to tumor cell, humoral immune response and high infiltrating abundance of Treg cell, Th2, eosinophil and so on. Treg cells act crucial roles in developing immunosuppressive tumor microenvironment [33]. Hashemi et al. had showed that high level of regulatory T cells was related to worse overall survival in BC patients, and Treg cell frequency showed independent prognostic value for patients with high risk of recurrence [34]. The Th1/Th2 balance was found to be involved in the antitumor immunity in BC. Plant extracts xanthohumol and saikosaponin A all showed anti-tumor activity by shifting Th1/Th2 balance toward Th1 in BC [35,36]. Eosinophils had been demonstrated to infiltrate tumors, and was found to be correlated with prognosis. However, whether high level of eosinophils was associated with favorable or adverse prognosis remains debatable, which depends on various factors, for example tumor types [37]. Tumor-associated tissue eosinophilia was found in 15% of TNBC, and was correlated with increased neoantigen load and nonsilent mutation rate [38]. Tsuda et al. suggested that B cells was involved in tumor immunity by producing antibodies and immunosuppressive, and increased proportion of total B cells was found in BC patients than that of normal [39]. Therefore, we speculated that higher level of these immune cells might explain the poor outcomes of TNBC patients with high risk to some extent.

Additionally, we identified four independent immune-related prognostic lncRNAs, including RP11-890B15.3, RP11-1024P17.1, MFI2-AS1 and RP11-180N14.1. LncRNA RP11 had been found to implicate in the development and progression of various cancers, containing BC [40,41]. Gao et al. showed that lncRNA RP11-480I12.5 displayed anti-tumor effect in BC by sponging miR-490-3p to elevate AURKA and Wnt/β-catenin pathway [42]. LncRNA MFI2-AS1 had also been reported to play important roles in different cancers, such as glioma [43], thyroid cancer [44], colorectal cancer [45], and clear-cell renal cell carcinoma [46]. For example, Wei et al. indicated that MFI2-AS1 could facilitate progression and metastasis of hepatocellular carcinoma by sponging miR-134 to increase the expression of FOXM1 [47]. In addition, Luo et al. suggested that MFI2-AS1 displayed independent value to indicate prognosis and might be served as a possible therapeutic target in colorectal cancer [48]. However, the role of MFI2-AS1 had not been investigated in BC. In our study,

we found that MFI2-AS1 might regulate the expression of TRAF2, SAPCD2, GPRIN1 by sponging miR-3150a-3p. Studies had suggested that TRAF2 was implicated in tumor cells apoptosis [49], invasion and migration of BC cells [50]. Zhang et al. indicated that SAPCD2 could facilitate the invasion and migration of BC cells [51]. Hence, we concluded that these four lncRNAs might play important roles in TNBC.

However, there were still some limitations. First, the exact roles of the four lncRNAs and their involved regulatory axis should be further investigated in future. Second, the clinical value of the prognostic model should be confirmed in clinical. Third, the landscape of tumor-infiltrating immune cells should be further confirmed in clinical.

In conclusion, a prognostic risk model was established on the basis of four independent prognostic lncRNAs, including RP11-890B15.3, RP11-1024P17.1, MFI2-AS1 and RP11-180N14.1. These four lncRNAs-based prognostic risk model could stratify the TNBC patients into high risk and low risk groups, and patients in high risk group displayed worse outcomes. Patients with high risk showed significant differences on immune-related GO-BP and tumor-infiltrating immune cells compared with patients with low risk. This might be one reason to explain the poor outcomes of high risk patients.

Acknowledgements

Conflict of interests

The authors declare that they have no competing interests.

References

1. A. B. Hanker, D. R. Sudhan, C. L. Arteaga, Overcoming endocrine resistance in breast cancer, *Cancer Cell*, **37** (2020), 496–513.
2. N. Li, Y. Deng, L. Zhou, T. Tian, S. Yang, Y. Wu, et al., Global burden of breast cancer and attributable risk factors in 195 countries and territories, from 1990 to 2017: results from the Global Burden of Disease Study 2017, *J. Hematol. Oncol.*, **12** (2019), 140.
3. L. Yin, J. J. Duan, X. W. Bian, S. C. Yu, Triple-negative breast cancer molecular subtyping and treatment progress, *Breast Cancer Res.*, **22** (2020), 61.
4. M. A. Medina, G. Oza, A. Sharma, L. G. Arriaga, J. M. Hernández, V. M. Rotello, et al., Triple-negative breast cancer: a review of conventional and advanced therapeutic strategies, *Int. J. Environ. Res. Public Health*, **17** (2020), 2078.
5. S. Zhao, W. J. Zuo, Z. M. Shao, Y. Z. Jiang, Molecular subtypes and precision treatment of triple-negative breast cancer, *Ann. Transl. Med.*, **8** (2020), 499.

6. T. E. Keenan, S. M. Tolaney, Role of immunotherapy in triple-negative breast cancer, *J. Natl. Compr. Cancer Network*, **18** (2020), 479–489.

7. A. Bhan, M. Soleimani, S. S. Mandal, Long noncoding RNA and cancer: a new paradigm, *Cancer Res.*, **77** (2017), 3965–3981.

8. W. X. Peng, P. Koirala, Y. Y. Mo, LncRNA-mediated regulation of cell signaling in cancer, *Oncogene*, **36** (2017), 5661–5667.

9. W. Wang, W. Lou, B. Ding, B. Yang, H. Lu, Q. Kong, et al., A novel mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network associated with prognosis of pancreatic cancer, *Aging*, **11** (2019), 2610–2627.

10. H. Dong, J. Hu, K. Zou, M. Ye, Y. Chen, C. Wu, et al., Activation of LncRNA TINCR by H3K27 acetylation promotes Trastuzumab resistance and epithelial-mesenchymal transition by targeting MicroRNA-125b in breast Cancer, *Mol. Cancer*, **18** (2019), 3.

11. C. N. Fan, L. Ma, N. Liu, Systematic analysis of lncRNA-miRNA-mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer, *J. Transl. Med.*, **16** (2018), 264.

12. W. Ma, F. Zhao, X. Yu, S. Guan, H. Suo, Z. Tao, et al., Immune-related lncRNAs as predictors of survival in breast cancer: a prognostic signature, *J. Transl. Med.*, **18** (2020), 442.

13. F. D. E. De Palma, V. Del Monaco, J. G. Pol, M. Kremer, V. D'Argenio, G. Stoll, 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** (2020), 105249.

14. M. Xu, X. Xu, B. Pan, X. Chen, K. Lin, K. Zeng, et al., LncRNA SATB2-AS1 inhibits tumor metastasis and affects the tumor immune cell microenvironment in colorectal cancer by regulating SATB2, *Mol. Cancer*, **18** (2019), 135.

15. D. Huang, J. Chen, L. Yang, Q. Ouyang, J. Li, L. Lao, et al., NKILA lncRNA promotes tumor immune evasion by sensitizing T cells to activation-induced cell death, *Nat. Immunol.*, **19** (2018), 1112–1125.

16. T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, D. Rudnev, C. Evangelista, et al., NCBI GEO: mining tens of millions of expression profiles--database and tools update, *Nucleic Acids Res.*, **35** (2007), D760–765.

17. J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, et al., GENCODE: the reference human genome annotation for The ENCODE Project, *Genome Res.*, **22** (2012), 1760–1774.

18. H. Jiang, W. H. Wong, SeqMap: mapping massive amount of oligonucleotides to the genome, *Bioinformatics*, **24** (2008), 2395–2396.

19. G. K. Smyth, limma: Linear Models for Microarray Data, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor,* Springer, (2005), 397–420.

20. S. Bhattacharya, S. Andorf, L. Gomes, P. Dunn, H. Schaefer, J. Pontius, et al., ImmPort: disseminating data to the public for the future of immunology, *Immunol. Res.*, **58** (2014), 234–239.

21. D. A. Barbie, P. Tamayo, J. S. Boehm, S. Y. Kim, S. E. Moody, I. F. Dunn, et al., Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1, *Nature*, **462** (2009), 108–112.

22. A. Liberzon, A. Subramanian, R. Pinchback, H. Thorvaldsdóttir, P. Tamayo, J. P. Mesirov, Molecular signatures database (MSigDB) 3.0, *Bioinformatics*, **27** (2011), 1739–1740.

23. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data, *BMC Bioinf.*, **14** (2013), 7.

24. P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, et al., Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade, *Cell Rep.*, **18** (2017), 248–262.

25. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters, *Omics*, **16** (2012), 284–287.

26. M. D. Paraskevopoulou, I. S. Vlachos, D. Karagkouni, G. Georgakilas, I. Kanellos, T. Vergoulis, et al., DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts, *Nucleic Acids Res.*, **44** (2016), D231–238.

27. H. Dweep, N. Gretz, miRWalk2.0: a comprehensive atlas of microRNA-target interactions, *Nat. Methods*, **12** (2015), 697.

28. M. Kohl, S. Wiese, B. Warscheid, Cytoscape: software for visualization and analysis of biological networks, in *Data mining in proteomics,* Humana Press, (2011), 291–303.

29. E. Almagro, C. S. González, E. Espinosa, Prognostic factors of early breast cancer, *Med. Clin.*, **146** (2016), 167–171.

30. P. Costa-Pinheiro, D. Montezuma, R. Henrique, C. Jerónimo, Diagnostic and prognostic epigenetic biomarkers in cancer, *Epigenomics*, **7** (2015), 1003–1015.

31. M. J. Kwon, Emerging immune gene signatures as prognostic or predictive biomarkers in breast cancer, *Arch. Pharm. Res.*, **42** (2019), 947–961.

32. Y. Yang, Cancer immunotherapy: harnessing the immune system to battle cancer, *J. Clin. Invest.*, **125** (2015), 3335–3337.

33. L. Wang, D. L. Simons, X. Lu, T. Y. Tu, S. Solomon, R. Wang, et al., Connecting blood and intratumoral T(reg) cell activity in predicting future relapse in breast cancer, *Nat. Immunol.*, **20** (2019), 1220–1230.

34. V. Hashemi, L. A. Maleki, M. Esmaily, A. Masjedi, G. Ghalamfarsa, A. Namdar, et al., Regulatory T cells in breast cancer as a potent anti-cancer therapeutic target, *Int. Immunopharmacol.*, **78** (2020), 106087.

35. X. Zhao, J. Liu, S. Ge, C. Chen, S. Li, X. Wu, et al., Saikosaponin a inhibits breast cancer by regulating Th1/Th2 balance, *Front. Pharmacol.*, **10** (2019), 624.

36. W. Zhang, Y. Pan, P. Gou, C. Zhou, L. Ma, Q. Liu, et al., Effect of xanthohumol on Th1/Th2 balance in a breast cancer mouse model, *Oncol. Rep.*, **39** (2018), 280–288.

37. S. Sakkal, S. Miller, V. Apostolopoulos, K. Nurgali, Eosinophils in cancer: favourable or unfavourable?, *Curr. Med. Chem.*, **23** (2016), 650–666.

38. K. Chouliaras, Y. Tokumaru, M. Asaoka, M. Oshi, K. M. Attwood, K. Yoshida, et al., Prevalence and clinical relevance of tumor-associated tissue eosinophilia (TATE) in breast cancer, *Surgery*, (2020), in press.

39. B. Tsuda, A. Miyamoto, K. Yokoyama, R. Ogiya, R. Oshitanai, M. Terao, et al., B-cell populations are expanded in breast cancer patients compared with healthy controls, *Breast Cancer*, **25** (2018), 284–291.

40. A. Giro-Perafita, L. Luo, A. Khodadadi-Jamayran, M. Thompson, B. Akgol Oksuz, A. Tsirigos, et al., LncRNA RP11-19E11 is an E2F1 target required for proliferation and survival of basal breast cancer, *NPJ Breast Cancer*, **6** (2020), 1.

41. D. Sun, H. Liu, T. Wang, Long noncoding RNA RP11-334E6.12 promotes the proliferation, migration and Invasion of breast cancer cells through the EMT pathway by activating the STAT3 cascade, *Cancer Manag. Res*, **12** (2020), 1113–1120.

42. X. Gao, Y. Lai, Z. Zhang, Y. Ma, Z. Luo, Y. Li, et al., Long Non-coding RNA RP11-480I12.5 promotes the proliferation, migration, and invasion of breast cancer cells through the miR-490-3p-AURKA-Wnt/β-Catenin axis, *Front. Oncol.*, **10** (2020), 948.

43. K. Liu, J. Liu, Q. F. Bo, MFI2-AS1 regulates the aggressive phenotypes in glioma by modulating MMP14 via a positive feedback loop, *Eur. Rev. Med. Pharmacol. Sci.*, **23** (2019), 5884–5895.

44. T. Yu, L. Tong, Y. Ao, G. Zhang, Y. Liu, H. Zhang, Upregulation of TRIAP1 by the lncRNA MFI2-AS1/miR-125a-5p axis promotes thyroid cancer tumorigenesis, *Onco. Targets Ther.*, **13** (2020), 6967–6974.

45. C. Li, F. Tan, Q. Pei, Z. Zhou, Y. Zhou, L. Zhang, et al., Non-coding RNA MFI2-AS1 promotes colorectal cancer cell proliferation, migration and invasion through miR-574-5p/MYCBP axis, *Cell Prolif.*, **52** (2019), e12632.

46. R. Flippot, R. Mouawad, J. P. Spano, M. Rouprêt, E. Compérat, M. O. Bitker, et al., Expression of long non-coding RNA MFI2-AS1 is a strong predictor of recurrence in sporadic localized clear-cell renal cell carcinoma, *Sci. Rep.*, **7** (2017), 8540.

47. Y. Wei, Z. Wang, Y. Zong, D. Deng, P. Chen, J. Lu, LncRNA MFI2-AS1 promotes HCC progression and metastasis by acting as a competing endogenous RNA of miR-134 to upregulate FOXM1 expression, *Biomed. Pharmacother.*, **125** (2020), 109890.

48. R. Luo, J. Song, W. Zhang, L. Ran, Identification of MFI2-AS1, a novel pivotal lncRNA for prognosis of stage III/IV colorectal cancer, *Dig. Dis. Sci.*, **65** (2020), 3538–3550.

49. C. Jiao, W. Chen, X. Tan, H. Liang, J. Li, H. Yun, et al., Ganoderma lucidum spore oil induces apoptosis of breast cancer cells in vitro and in vivo by activating caspase-3 and caspase-9, *J. Ethnopharmacol.*, **247** (2020), 112256.

50. Y. Yao, K. Zhao, Z. Yu, H. Ren, L. Zhao, Z. Li, et al., Wogonoside inhibits invasion and migration through suppressing TRAF2/4 expression in breast cancer, *J. Exp. Clin. Cancer Res.*, **36** (2017), 103.

51. Y. Zhang, J. L. Liu, J. Wang, SAPCD2 promotes invasiveness and migration ability of breast cancer cells via YAP/TAZ, *Eur. Rev. Med. Pharmacol. Sci.*, **24** (2020), 3786–3794.