Identification of an Independent Immune-Genes Prognostic Index for Breast Cancer

Deng Rong (✉ dengrong1977@163.com)

Nanjing Medical University affiliated Nanjing Hospital: Nanjing First Hospital <https://orcid.org/0000-0002-4591-6165>

Yu yun

Nanjing Medical University

Lin chen

Nanjing Medical University affiliated Nanjing Hospital: Nanjing First Hospital

Dong Yuxiang

Nanjing Medical University

Pan Yitong

Nanjing Medical University

Zhang Yuhan

Nanjing Medical University

Wang Junyi

Nanjing Medical University

Chen chen

Nanjing Medical University

Lu Jianing

Nanjing Medical University

Research article

Keywords: Breast cancer, immune genes, prognosis, risk scores model, nomogram

Posted Date: October 8th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-81455/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Cancer on March 30th, 2021. See the published version at <https://doi.org/10.1186/s12885-021-08041-x>.

Abstract

Objective: There are increasing evidences that the immune micro-environment of breast cancer (BC) seriously affects the clinical outcome. This study aims to explore the role of tumor immune genes in the prognosis of BC patients.

Methods: We obtained the list of 2498 immune genes from ImmPort database. In addition, we obtained gene expression data and clinical characteristics data of BC patients from the TCGA database. The prognostic correlation of the differential genes was analyzed through Survival package. Cox regression analysis was performed to analyze the prognostic effect of immune genes. According to the regression coefficients of prognostic immune genes in regression analysis, an immune risk scores model was established. Gene set enrichment analysis (GSEA) was performed to probe the biological correlation of immune gene scores. $P < 0.05$ was considered to be statistically significant.

Results: In total, 556 immune genes were differentially expressed in normal tissues and BC tissues ($p < 0.05$). In univariate cox regression analysis, a total of 66 immune genes were statistically significant for survival risk, of which 30 were associated with overall survival ($P < 0.05$). Through lasso regression analysis, a 15 immune genes risk scores model was established. KM survival analysis revealed that high immune risk scores represented worse survival ($p < 0.001$). ROC curve indicated that the immune genes risk scores model had good a reliability in predicting prognosis (5-year OS, AUC=0.752). The established risk model showed splendid AUC value in the validation dataset (3-year over survival (OS) AUC=0.685, 5-year OS AUC=0.717, $P=0.00048$). Moreover, immune risk scores were proved to be an independent influencing factor for BC patients. Finally, we found that 15 immune genes and risk scores had significant clinical correlations, and were involved in a variety of carcinogenic pathways.

Conclusion: In a word, our study provides a new perspective for the expression of immune genes in BC. In addition, the immune risk scores have considerable advantages in predicting the survival of BC, which are associated with a variety of oncogenic pathways.

Introduction

Breast cancer is one of the main malignant tumors that threaten the lives of women, which has received more and more clinical attention worldwide. It is regarded as the second common malignant tumors in the world. It occupied 25% of all malignant tumors. (1) TNBC was supposed to be the foremost malignant sub-type, accounting for approximately 20%. (2, 3) It is manifested as a large tumor, a high level of differentiation, a high risk of metastasis, and lymph node invasion. (4–6) TNBC is characterized by negative human epidermal growth factor receptor 2 (HER-2), progesterone receptor (PR) and estrogen receptor (ER), thus resistant to endocrine therapy and trastuzumab. (7) Regard to the lack of targeted treatment strategies, chemotherapy remained the unique treatment option. (8) Therefore, it is integral to conduct a comprehensive bioinformatics study on gene expression of breast cancer to identify potential genes that can be used as therapeutic targets in BC.

Previous evidence has shown that the immune system has a contradictory influence on the occurrence and development of cancer, contributing to both cancer progression and inhibition. (9–12) The immune system has a practical impact on the progress of BC. The response of BC patients to immunotherapy and traditional treatment is interfered by immune system. (13, 14) Immune evasion is a tough problem in the immunotherapy of BC. Since there are significant diversities in the expression profiles of immune genes between BC and other cancer types (15, 16), further research is needed to determine which immune genes can play a role as therapeutic targets.

In this study, we explored the lineage and expression profiles of immune genes in BC, and its impacts on the prognosis of BC patients. Moreover, an immune-genes based risk scores model and a nomogram is constructed to predict the prognosis of BC.

Materials And Methods

Acquisition of data

First of all, a list of 2498 immune genes were downloaded from ImmPort database. Additionally, we obtained the gene expression profiles of BC patients from the TCGA database, including 112 normal cases and 857 tumor cases. Meanwhile, corresponding clinical data were also obtained (Table 1). $|\text{LogFC}| < 1$ and $P < 0.05$ were used as the criteria for screening differential genes. Because TCGA is an open and publicly available database, ethical approval is not required.

Table 1
Clinical characteristics of included patients in the study

Variables	Total (n = 857)	Training cohort (n = 577)	Validation cohort (n = 280)
Age (year)			
< 60	484	325	159
≥ 60	373	252	121
Gender			
Female	846	570	276
Male	11	7	4
Stage			
I	151	99	52
II	501	336	165
III	188	134	54
IV	17	8	9
T stage			
T1	229	161	68
T2	511	336	175
T3	93	62	31
T4	29	20	9
N stage			
N0	411	272	139
N1	292	196	96
N2	102	73	29
N3	53	36	17
M stage			
M0	840	569	271
M1	17	8	9
Survival			
Dead	118	83	35
Alive	739	494	245

Gene function enrichment analysis

GO enrichment analysis is conducted to comprehend the biological process and molecular function of the differential genes, while KEGG enrichment analysis is applied to identify potential related biological pathways. Gene enrichment analysis (GSEA) is performed between normal tissues and BC tissues in order to probe the biological pathways associated with immune genes risk scores.

Construction and validation of the immune genes risk scores

We use Cox regression tool for survival analysis. On the basis of differential expression, we screened out single factor cox significant and survival-related prognostic immune genes. Further, the least absolute shrinkage and selection operator (LASSO) regression analysis is executed to reduce the dimensionality, so as to screen out the optimal variables. Based on the variables obtained by LASSO and the corresponding regression coefficients, the risk scores were calculated. We use the median value to divide patients into a high- and a low-risk scores group. The prognostic correlation of immune gene risk score was obtained by Kaplan-Meier curve. The credibility and predictive value of the risk scoring model was evaluated through time-related ROC curve.

Analysis of copy number variation data and gene mutation analysis

Based on TCGA breast cancer data, the copy number variation (CNV) was analyzed using R-Circos package and R-ggplot2 package. Furthermore, we used the online tool website-cbioportal to analyze the genetic variation of hub genes. The threshold used was $P < 0.05$.

Statistical analysis

R3.6.1 was used for statistical analysis. The independent t test was used for continuous variables with normal distribution, and the Mann-Whitney U test was used for continuous variables with skewed distribution. A two-sided test was used, and a P value of < 0.05 was considered statistically significant.

Results

Differentially expressed immune genes (DEIGs) screening of BC

A list of 2498 immune genes were obtained from ImmPort database. We obtained mRNA expression data of 857 cases of breast cancer and 122 cases of normal tissues from the TCGA database for further bioinformatics analysis. A total of 556 immune genes were determined as DEIGs between BC tissues and normal tissues, including 402 up-regulated and 154 down regulated ($p < 0.05$, Fig. 1A, Table 2). The heatmap spreaded out the top 10 up-regulated and top 10 down-regulated DEIGs (Fig. 1B).

Table 2
Analysis of differentially expressed immune genes in TCGA breast cancer

Up regulated	<p> RAC2ITKADARRFXANKGDF11IRF1CMTM1UCN2MX2TGFB1KRASIFNA17PLXNC1IFNA2DHX58ADRM1PIK3R3PRDX1SOCS1S100A11HLA-DQA1VEGFAF2ROGFRPSMC4SEMA4A1L1RNHLA-DQB1IL32CALRCD48INHBCDEFB104BTNFRSF12ACD3ECCR6SERPINA3HSPA2IL31RANR1I3NFKBIBINHBEISG20L2IFIH1LEAP2CACYPBZAP70CXCR4GIPRTNFRSF13CTGFB3DEFB103ADDX58GNRH2CCL25RFX5OPRL1SRCNCR2LATPSMD14S100A16IGF1RNCR3BCL3HLA-DQA2CARD11RELBICD79AMCHR1CD86RBP5IFITM1UNC93B1IL2RAPTPN6RLN1FASLGSTAT1PGCMAPTPSME2AQP9IRF5IL2RGHDGFCCL19BMP10NFATC4LYZRBP1DEFB105APTGER1LCKTFRCSH2D1ACD3DIL12RB1RARRES3IFNA5MSR1KIR3DL2AMBNPDFHNF4GKLRC2SEMA3FIFNA14CCR5CD1EHAMPIL23AFCGR3ABST2CD22SPAG11BTMSB15BGREM1VAV2PPP4CITGALRARANR2F6PAK1CXCL13HLA-GPRLRTNFSF13BPLAUICD72BLNKMDKPAK4PSMD3SLC29A3FGF22SEMA5BIL2RBNFKBIEAPOBEC3HCSHL1MC1RSLC10A2PAEPMC4RRLN3CSF3RIRF9BMP15GNRHRISG20IFNA21DEFB136DEFA5PLXNA3INSL5VAV3CCR3GUCA2APNOCITOR2ATAP1BMP8BRAC3CLEC11ACSPG5IL18IKBKEABEP2IFNA13CTLA4TNFRSF4NOX3FAM19A5IFI30MIFCBLINOX5GPR33RNASE2EPGNGPHB5IFNW1OAS1CD1BFLT3INS-IGF2NOX1TGNR2E3MICBSECTM1GZMBSEMA7ACD19IL24C8GMBL2HSPA6OSMAGTMMX1IL17FHNF4ASDC1RSAD2APOBEC3ATYMPHTR3ESR1RETNSLC11A1PPYCCCL15NOD2DEFB121UCNPIK3R2ANGPTL6IFNA6CGB2AGRPCCR7CXCR3PLAURCRABP2IFNA7DEFB129RASGRP1GDF2IL3LECT2IFNGFGFR4LEFTY1S100A7ULBP2WFIKKN1RXFP1MCHR2IRF7CCR4COLEC10AVPR1BAZU1PDCD1TMSB15ATHPOGALPIDO1CCL17LTAALALR3MLN1IL11TNFSF4HTR3BFSHBRLN2OBP2APRKCGKIR2DL4ICOSSPP1CGB8IL1F10INSL4LTBCELA1FABP12DEFB134IL27GALR2SSTR2PGLYRP2RBP2CXCR5IFNA16S100A14ADM2UTS2IL12BLCN12OLR1MMP9CCL1SCG2IFNA4MMP12OASLDEFB108BIL9AMELXGDF15IL9RKCND2CTSEDEFB110FGFR3CSH1CCL20MC2RGPHA2EDN2TMPRSS6GALSEMGI1BMP8APTHROBO2RETNLBHTR1ADEFB128PMCHRXFP3HRGGH2DEFB113PTH2IL21RTNFRSF9PROCOTR3AAMHTNFRSF18ESM1MTNR1BCXCL9PYYGCGRINHAACGB5LCN9DEFB112ISG15OPRD1SLURP1IFNA10GDF9CD1AUMODL1FGF23ULBP1IL17CKIR3DL3IL21CXCL10ARTNINHBAACCR8BIRC5SCTVGFOTFR2HTN3SSTR5IL20PRLHFGF21GIPR3HDMLCXCL11KNG1TUBB3CCL7S100A7ALCN1ORM2APOHEPOPGLYRP4FGF3FGF5IFNB1PGLYRP3BMPR1BCCL11FABP6SEMGI2CAMP1S100PMUC5ACDEFB126GHSRDEFB123DEFB115ORM1GCGDEFB116TRHCSH2FGF4TCHHL1IL19HTN1REG1APCSK1IAPPINSICST4CGAUCN3 </p>
Down regulated	<p> LEPADIPOQACVR1CFABP4RBP4DEFB132OXTRANGPTL7SAA1ANGPTL5CSF3LALBAACXCL2BMP3MASP1NPR1GLP2RPENKNOS1PPARGGDF10ANGPT4CCL14TSLPPLXNA4SAA2GHRDESANGPTL1CMA1S100BLHCGRIL6IL33LEPRFOSSEMA3GSCTRFABP9CX3CL1PTNCCL28FGF2ADCYAP1R1STAB2ADRB2ANGPT1EDN3RXRGCD209LIFRTGFB3RNASE7CNTFRAVPR2OSTNCCL21TACR1GNAI1PF4OGNIGF1PAK3NTF4GFAPTGFB2IFNA8NRG2RBP7APODCCL24LCN6KLPTH1RFGF1BMP2NGFRNEDNRBGPR17PTGFRNR4A3ELANE1S1PR1CCL13CCL16CATCXCL12IL17BANGPTL4SOCS3ACO1NRG1NR4A1CYR61LTBP4NR3C2PDGFDCCCL23PPBPSEMA3DNPR3NMBSCGB3A1ANGPTL2TINAGL1ESR2CRIM1CXCL3NR3C1METTEKIL17DBMP6EGFRVIPCTSGVIMLRP1GREM2FGF7PTGS2JUNPIK3R1ROBO3LCN10IL17RDTSHBCSRP1AHNAKSEMA5APLA2G2AMARCOADMMPMP2FAM3DTNFRSF10DSEMA3ASEMA6DEDN1NOVPLTPLGR6PDGFRA1TLR4SSTR1AVPR1APDGFATPM2PTGER4THRBEGFIL11RACRHR2CER1ICAM2A2MPTGDS1TAC1SLIT2LGR4BACH2PDGFRLC3FGF16 </p>

Functional annotation of these 556DEIGs

To study the potential mechanisms and molecular functions of the identified 556 DEIGs, we conducted the GO and KEGG analysis. The top three enriched GO terms for up-regulated DEIGs and down-regulated DEIGs were: T cell activation, lymphocyte differentiation and response to virus; cell chemotaxis, positive regulation of response to external stimulus and leukocyte migration, respectively (Fig. 2A). KEGG analysis revealed the top enriched pathway for up-regulated DEIGs and down-regulated DEIGs were: Cytokine – cytokine receptor interaction, JAK – STAT signaling pathway and Chemokine signaling pathway; Cytokine – cytokine receptor interaction, JAK – STAT signaling pathway and EGFR tyrosine kinase inhibitor resistance pathway, respectively (Fig. 2B).

Establishment of immune prognosis model

Among the identified 556 DEIGs, 66 prognostic DEIGs were identified by utilizing univariate cox regression analyses (Fig. 3A). KM survival analysis showed that 30 of them were significantly correlated with OS. TCGA BC samples were randomly separated into two sets (training set : validation set, 7 : 3). Then, lasso regression analysis was applied to increase the robustness and select the optimal variables based on training set. 15 DEIGs were got for the construction of immune prognostic index (Fig. 3B, Fig. 3C, Table 3). According to the median risk score, BC patients were separated into high risk and low risk (Fig. 3D, Fig. 3F). Heat map was utilized to visualize the difference of gene expression profile in low- and high- risk patients in BC training set (Fig. 2E). The results from KM analysis revealed that high risk patients possessed lower overall survival in both training group and validation group ($P < 0.001$) (Fig. 4A, Fig. 4B, Fig. 4C). The ROC curve prompted that the risk scores model had dominant credibility and predictive value ($AUC = 0.813$, $AUC = 0.704$ for 5 years overall survival in training and validation group, respectively) (Fig. 4D, Fig. 4E, Fig. 4F). In further univariate cox analysis, age, pathological stage, pathological T, N, M stage and high risk scores were associated with poor survival (Fig. 5A). In multivariate Cox model, only age and risk score worked as independent predicted factors (Fig. 5B). To establish a quantitative visualization model of breast cancer prognosis, we combined multiple clinical factors to establish a nomogram (Fig. 5C, 5D, 5E).

Table 3
Multivariate cox regression analysis to
establish RNA binding proteins risk
prediction model

Gene	Coef
TSLP	-0.703829357640691
IL17B	-0.0870608394604504
NR3C2	-0.0255482484720901
RAC2	-0.130057137304801
SERPINA3	-0.0898937544948299
HSPA2	-0.120788735486787
CD79A	-0.0431127011058176
UNC93B1	0.513946621757904
NFKBIE	-0.329152003213528
SDC1	0.0854293362952585
IFNG	-0.220305753667004
IRF7	-0.171479153154717
GALP	2.91458293196349
TNFRSF18	-0.129391946165935
ULBP1	0.174787641983627

Recognition of gene sets for genome variation

Based on TCGA breast cancer data, we analyzed the copy number variation (CNV) of 15 model genes and showed the frequency of copy number variation using R-Circos package and R-ggplot2 package (Figure S1A, Figure S2B). The results showed that PAC2, ULBP and SERPINA3 had the most variation frequency (Figure S2B). Furthermore, we analyzed the single nucleotide polymorphism composition (SNPs) of 15 model genes (Figure S2C). The results showed that NR3C2 had the most SNPs, including missense mutation and silent. Finally, we used the online tool website-cbioportal to analyze the genetic variation of 15 immune genes (Figure S2D).

Clinical and prognostic correlation of 15 model genes and immune genes risk score

The proportion of 15 model genes in different clinical and pathological stages was investigated. We demonstrated that IL17B, NFKBIE and SERPINA3 mainly drove development of breast cancer (Fig. 6). In

addition, survival analysis showed that all model genes were significantly associated with survival (Figure S2). Meanwhile, we found that the expression of RAC2, CD79A and IFNG were associated with the infiltration of Macrophage M0 and Macrophage M2 (Fig. 7). Regard to the immune genes risk score, a strong correlation with age, gender, pathological stage and clinical T stage was identified (Fig. 8).

Gene set enrichment analysis of risk scores

To explore the biological correlation of risk scores involved in progression of breast cancer, we performed a GSEA analysis of risk scores based on the TCGA breast cancer cohort. GSEA analysis indicated high risk scores was associated with E2F_TARGETS, G2M_CHECKPOINT, GLYCOLYSIS, MTORC1_SIGNALING and PROTEIN_SECRETION pathway (Fig. 9A). In addition, low risk scores was associated with APOPTOSIS, COMPLEMENT, IL2_STAT5_SIGNALING, INFLAMMATORY_RESPONSE and P53 pathway (Fig. 9B).

Discussion

BC is regarded as the most common malignant tumor in women. Although great efforts have been made to improve diagnosis and treatment strategies, it still poses a fatal threat to patients. Accumulation of evidence have shown that Cancer immunotherapy, especially the treatment of immune checkpoint inhibitors, has become an important part of the treatment of certain types of cancer, and has provided a continuous therapeutic effect for specific groups of patients. (17) Immune genes, such as cytokines, not only act locally, but rapidly spread within the tumor and affect the activation and dissemination of tumor immune cells. (18, 19) Obviously, different types of cancer have different immune gene subgroups. Therefore, the examination of immune gene subgroups is essential for judging the risk of tumors and exploring immunotherapy.

In our research, we performed a detailed and comprehensive evaluation of immune genes in BC. All gene expression data and patients clinical characteristics information were downloaded from TCGA dataset. 2498 immune genes from ImmPort database were analyzed between breast cancer and normal tissues. Eventually, 556 DEIGs were verified. Moreover, we identified and constructed a 15 immune genes risk scores model for breast cancer through univariate and lasso regression analysis, including TSLP, IL17B, NR3C2, RAC2, SERPINA3, HSPA2, CD79A, UNC93B1, NFKBIE, SDC1, IFNG, IRF7, GALP, TNFRSF18 and ULBP1. Furthermore, to study the clinical and biological relevance of risk scores, we performed the KM, ROC and GSEA analysis. Indeed, the high risk group received a lower survival, and possessed a higher histological grade.

- Several DEIGs in the immune genes risk scores model have been investigated in human cancers. Thymic interstitial lymphopoietin (TSLP), a key inflammatory cytokine that induces type 2 inflammation, predicts a poor prognosis in oropharyngeal squamous cell carcinoma (OPSCC). (20) Interleukin-17 (IL-17), a member of the interleukin family, is a cytokine that plays a role in inflammation and cancer, and can enhance lung cancer invasion/migration ability. (21) In addition,

studies have found that knocking down RAC2 can inhibit the progression of osteosarcoma by inhibiting the wnt signaling pathway. (22) The up regulation of hnRNP-K transcriptional activity mediated by SERPINA3 promotes the survival and proliferation of HCC cells, which may be an indicator of poor prognosis in HCC patients. (23) AS a putative oncoprotein, Heat shock protein family a member 2 (HSPA2) is often up-regulated in human malignancies and promotes aggressive phenotype of tumors. (24) NFKBIE aberrations are common genetic events in trans-b-cell malignancies, and NFKBIE deletion is a new marker of poor prognosis in primary mediastinal B-cell lymphoma (PMBL). (25) The remaining genes have also been confirmed to be interrelated to malignant origin, aggressive behaviours of tumors.

Similarly, Lai et al. (26) established a panel of 4 autophagy-related genes (ARG) signatures consisting of SERPINA1, ATG4A, NRG1 and IFNG to predict the prognosis of breast cancer, which can help clinicians make judgments and decisions on determining effective treatment strategies. Wang et al. (27) identified a six differentially expressed genes (DEGs) model consisting of IGHA2, SERPINA1, GFALS, SPDYC, PAX7, and ADRB1 by using cox regression survival modeling for breast cancer. In another study (28), the authors constructed a prognostic risk scoring system containing 6 genes (SCUBE3, RDH16, SPC24, SPC25, CCDC69 and DGAT2), suggesting that these mRNAs may serve a driving role in the progression of Her2-positive BC. The construction of this risk scoring system is conducive to identifying high-risk HER2-positive BC patients, and it subserve to help achieve personalized targeted therapy. Different from previous studies, we first focused on DEIGs, and established and verified a novel DEIGs risk scores prediction model.

Nevertheless, there still remain some weak points in our research. Firstly, our results are based on bulk RNA sequencing of single omics. The heterogeneity and diversity between cells in the tumor microenvironment is ignored. Secondly, only gene expression and gene mutation levels are concerned, while tumor burden, methylation levels and other equally important events in tumor progression are ignored.

Conclusion

In conclusion, our study reveals the biological effects of immune genes in the origin and development of BC. The immune gene risk score model has advantages in predicting the prognosis of BC, which is an independent factor affecting the prognosis of BC. In addition, our findings may be of great guiding value in make a thorough inquiry of novel strategies for cancer immunological diagnosis and treatment.

Abbreviations

OS: over survival; TNBC: Triple negative breast cancer; ER: estrogen receptor; PR: progesterone receptor; HER-2: human epidermal growth factor receptor 2; TIICs: tumor infiltrating immune cells; KEGG: Kyoto Encyclopedia of Genes and Genomes; GO: gene ontology; DAVID: Database for Annotation, Visualization, and Integrated Discovery; GSEA: Gene enrichment analysis; CNV: copy number variation;

LASSO: least absolute shrinkage and selection operator; DEIGs: differentially expressed immune genes; SNPs: single nucleotide polymorphism; TME: Tumor micro-environment; TSLP: Thymic stromal lymphopoietin; OPSCC: oropharyngeal squamous cell carcinoma; HSPA2: Heat shock protein family a member 2; PMBL: primary mediastinal B-cell lymphoma; ARGs: autophagy-related genes; DEGs: differentially expressed genes.

Declarations

Acknowledgments

None.

Authors contributions

Rong Deng, Yun Yu and Lin Chen designed this work. Lin Chen and Yuxiang Dong wrote the manuscript. Yuxiang Dong and Yitong Pan performed the bioinformatics analysis. Chen Chen, Yuhan Zhang, Junyi Wang and Jianing Lu performed the data review. All authors have read and approved the manuscript.

Funding

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work (if applied, including full data access, integrity of the data and the accuracy of the data analysis) in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359-86.
2. Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review. *Histopathology*. 2008;52(1):108–18.
3. Schneider BP, Winer EP, Foulkes WD, et al. Triple-negative breast cancer: risk factors to potential targets. *Clin Cancer Res*. 2008;14(24):8010–8.
4. van Roozendaal LM, Smit LHM, Duijsens G, et al. Risk of regional recurrence in triple-negative breast cancer patients: a Dutch cohort study. *Breast Cancer Res Treat*. 2016;156(3):465–72.
5. Lehmann BD, Bauer JA, Chen X, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*. 2011;121(7):2750–67.

6. Burstein MD, Tsimelzon A, Poage GM, et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clin Cancer Res.* 2015;21(7):1688–98.
7. Liedtke C, Rody A. New treatment strategies for patients with triple-negative breast cancer. *Curr Opin Obstet Gynecol.* 2015;27(1):77–84.
8. Mayer IA, Abramson VG, Lehmann BD, et al. New strategies for triple-negative breast cancer—deciphering the heterogeneity. *Clin Cancer Res.* 2014;20(4):782–90.
9. Woo SR, Corrales L, Gajewski TF. Innate immune recognition of cancer. *Annu Rev Immunol.* 2015;33:445–74.
10. Biswas SK. Metabolic Reprogramming of Immune Cells in Cancer Progression. *Immunity.* 2015;43(3):435–49.
11. Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. *Nat Immunol.* 2013;14(10):1014–22.
12. Li B, Severson E, Pignon JC, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 2016;17(1):174.
13. Varn FS, Mullins DW, Arias-Pulido H, et al. Adaptive immunity programmes in breast cancer. *Immunology.* 2017;150(1):25–34.
14. Vonderheide RH, Domchek SM, Clark AS. Immunotherapy for Breast Cancer: What Are We Missing? *Clin Cancer Res.* 2017;23(11):2640–6.
15. Bates JP, Derakhshandeh R, Jones L, et al. Mechanisms of immune evasion in breast cancer. *BMC Cancer.* 2018;18(1):556.
16. Emens LA. Breast Cancer Immunotherapy: Facts and Hopes. *Clin Cancer Res.* 2018;24(3):511–20.
17. Nagarsheth N, Wicha MS, Zou W. Chemokines in the cancer microenvironment and their relevance in cancer immunotherapy. *Nat Rev Immunol.* 2017;17(9):559–72.
18. Bremnes RM, Al-Shibli K, Donnem T, et al. The role of tumor-infiltrating immune cells and chronic inflammation at the tumor site on cancer development, progression, and prognosis: emphasis on non-small cell lung cancer. *J Thorac Oncol.* 2011;6(4):824–33.
19. Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* 2017;77(21):e108-e10.
20. Lin CM, Lin LW, Chen YW, et al. The expression and prognostic impact of proinflammatory cytokines and their associations with carcinogens in oropharyngeal squamous cell carcinoma. *Cancer Immunol Immunother.* 2020;69(4):549–58.
21. Yang YF, Lee YC, Lo S, et al. A positive feedback loop of IL-17B-IL-17RB activates ERK/beta-catenin to promote lung cancer metastasis. *Cancer Lett.* 2018;422:44–55.
22. Xia P, Gao X, Shao L, et al. Down-regulation of RAC2 by small interfering RNA restrains the progression of osteosarcoma by suppressing the Wnt signaling pathway. *Int J Biol Macromol.* 2019;137:1221–31.

23. Ko E, Kim JS, Bae JW, et al. SERPINA3 is a key modulator of HNRNP-K transcriptional activity against oxidative stress in HCC. *Redox Biol.* 2019;24:101217.
24. Yang YL, Zhang Y, Li DD, et al. RNF144A functions as a tumor suppressor in breast cancer through ubiquitin ligase activity-dependent regulation of stability and oncogenic functions of HSPA2. *Cell Death Differ.* 2020;27(3):1105–18.
25. Mansouri L, Noerenberg D, Young E, et al. Frequent NFKBIE deletions are associated with poor outcome in primary mediastinal B-cell lymphoma. *Blood.* 2016;128(23):2666–70.
26. Lai J, Chen B, Mok H, et al. Comprehensive analysis of autophagy-related prognostic genes in breast cancer. *J Cell Mol Med.* 2020.
27. Wang F, Tang C, Gao X, et al. Identification of a six-gene signature associated with tumor mutation burden for predicting prognosis in patients with invasive breast carcinoma. *Ann Transl Med.* 2020;8(7):453.
28. Gao C, Zhuang J, Li H, et al. Development of a risk scoring system for evaluating the prognosis of patients with Her2-positive breast cancer. *Cancer Cell Int.* 2020;20:121.

Figures

Figure 1

Identification of DEIGs. (A) volcano plots of 556 DEIGs in breast cancer and normal tissues from TCGA database. (B) Heatmap plots of top 10 up-regulated and top 10 down-regulated DEIGs. The colors in the heatmaps from green to red represent expression level from low to high. The red dots in the volcano plots represent up-regulation, the green dots represent down-regulation and black dots represent genes without differential expression

A

B

Figure 2

GO (A) and KEGG(B) enrichment analysis of DEIGs.

Figure 3

(A) Univariate survival analysis by Cox proportional hazards models to select prognostic key immune genes. (B-C) LASSO Cox regression model for 19 prognostic immune genes used to construct immune genes risk score model. (D) Distribution of immune risk scores in breast cancer patients. (E) Distribution of survival status in breast cancer patients. (F) Distribution of specific risk factors in the high- and low-risk groups (divided by median value). (* $P < 0.05$)

Figure 4

(A)Kaplan-Meier curve analysis of high-risk and low-risk patients in the training cohort. (B) Kaplan-Meier curve analysis of high-risk and low-risk patients in the testing cohort.(C) Kaplan-Meier curve analysis of high-risk and low-risk patients in the entire TCGA cohort. (D) Timedependent ROC curve analysis of the training cohort. (E) Timedependent ROC curve analysis of the testing cohort. (F) Time-dependent ROC curve analysis of the entire TCGA cohort.

Figure 5

Cox's proportional hazard model of correlative factors in breast cancer patients. (A) Univariate COX regression analysis for seven clinicopathological parameters affecting the overall survival. (B) Multivariate COX regression analysis for seven clinicopathological parameters affecting the overall survival. (C) An established nomogram to predict breast cancer survival based on cox model. (D-E) Plots displaying the calibration of each model comparing predicted and actual 3- and 5-year overall survival.

Figure 6

Correlation analysis between TNM&Stage and 15 model genes in breast cancer cases. (A) Correlation analysis between tumor stage and 15 model genes expression in breast cancer cases. (B) Correlation analysis between node stage and 15 model genes expression in breast cancer cases. (C) Correlation analysis between metastasis stage and 15 model genes in breast cancer cases. (D) Correlation analysis between pathologic stage and 15 model genes expression in breast cancer cases.

Figure 7

Correlation between 15 model immune genes and immune cell infiltration.

Figure 8

Correlation between immune genes risk scores and various clinical factors. (A) Age. (B) Gender. (C) Stage. (D) T stage. (E) N stage. (F) M stage.

Figure 9

Gene set enrichment analysis of immune genes risk scores. (A) high risk scores. (B) low risk scores.

Supplementary Files

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

- [FigureS2.pdf](#)
- [FigureS1.pdf](#)