

# Development of a Prognostic Model Based on Identification of EMT-related LncRNAs in Triple Negative Breast Cancer

Jiani Guo (✉ [guojiani0218@163.com](mailto:guojiani0218@163.com))

Affiliated Huai'an No1 People's Hospital of Nanjing Medical University: Huaian First People's Hospital  
<https://orcid.org/0000-0002-9899-0924>

Xuesong Yi

The Huaian Clinical College of Xuzhou Medical University

Zhuqing Ji

Affiliated Huai'an No1 People's Hospital of Nanjing Medical University: Huaian First People's Hospital

Mengchu Yao

Affiliated Huai'an No1 People's Hospital of Nanjing Medical University: Huaian First People's Hospital

Yu Yang

Affiliated Huai'an No1 People's Hospital of Nanjing Medical University: Huaian First People's Hospital

Wei Song

Affiliated Huai'an No1 People's Hospital of Nanjing Medical University: Huaian First People's Hospital

Mingde Huang

The Huaian Clinical College of Xuzhou Medical University

---

## Research article

**Keywords:** lncRNA, epithelial-mesenchymal transition, prognostic signature, triple negative breast cancer

**Posted Date:** March 31st, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-356069/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 Journal of Oncology on November 27th, 2021. See the published version at <https://doi.org/10.1155/2021/9219961>.

# Abstract

**Background:** Triple negative breast cancer (TNBC) remains the most incurable subtype of breast cancer owing to high heterogeneity, aggressive nature, and lack of treatment options. It is generally acknowledged that epithelial-mesenchymal transition (EMT) is the key step in tumor metastasis.

**Methods:** With the application of TCGA and GEO database, we identified EMT-related lncRNAs by Cox univariate regression analysis. Optimum risk scores were calculated and used to divide TNBC patients into high/low-risk subgroups by the median value using lasso regression analysis. Kaplan-Meier and ROC curve analyses were applied for model validation. Then we assessed the risk model from multi-omic aspects including immune infiltration, drug sensitivity, mutability spectrum, signaling pathways, and clinical indicators.

**Results:** The risk model was composed of 22 EMT-related long noncoding RNAs (lncRNAs), which seemed to be valuable in prognostic prediction of TNBC patients. The model could act as an independent prognostic factor of TNBC, and showed a robust prognostic ability in the stratification analysis.

**Conclusions:** Together, our study successfully established a risk model with great accuracy and efficacy in prognosis prediction of TNBC patients.

## Background

Triple negative breast cancer (TNBC) is defined as a highly aggressive subtype of breast cancer which lacks estrogen receptor (ER), progesterone receptor (PR) expression, and no amplification of human epidermal growth factor receptor 2 (HER2)(1). TNBC represents almost 20% of all subtypes, and is more likely to be diagnosed in young females under 40 years(2, 3). The pathological characteristics of TNBC, such as high histological grade and central necrosis, make it more likely to develop relapse and visceral metastasis than other subtypes(4, 5). Due to the absence of molecular therapeutic targets, the standard management of TNBC remains chemotherapy and radiotherapy(6). Unfortunately, tumor resistance arises rapidly, followed by patient relapse or metastasize quickly, and results in poor prognosis(7). Even immunotherapy, i.e., immune checkpoint inhibition (ICI), which has been effectively used in several types of solid tumor, have shown little efficacy for TNBC patients(8–10). Hence, there is an urgent need to explore novel biomarkers and potential therapeutic approaches to improve the outcome of TNBC.

It is widely acknowledged that epithelial–mesenchymal transition (EMT) is the most important step that leads to the metastasis of malignant tumors, including TNBC(11). EMT is the process of polar epithelia transforming into the cells capable of free movement, which enhances the invasiveness of tumor cells into peripheral circulation(12). It is recently found that EMT is closely related with multiple signaling pathways including Notch, Hedgehog, PI3K/AKT and Wnt/ $\beta$ -catenin pathways, revealing its key position in TNBC development and great potential in improving clinical outcome of TNBC patients(13).

On the other side, long non-coding RNAs (LncRNAs) are a class of non-coding RNAs (ncRNAs) with a length of more than 200 nucleotides(14). Dysregulation of LncRNAs had been confirmed to be crucial in TNBC progression, including cell proliferation, apoptosis, invasion, metastasis, and regulation of drug resistance (15). Although numerous researches have focused on developing novel lncRNA-based therapeutics, there is still a long way to apply it in clinical practice.

In this study, by using multi-omics analysis, we successfully identified several EMT-related lncRNAs and construct a novel risk score prognostic model with strong efficiency on prognostic prediction. Our aim in this investigation is to understand the potential clinical application of EMT-related lncRNAs in prognosis stratification and their potential significance as biomarkers for targeted TNBC therapy. We systematically analyzed the expression and prognosis of EMT-related lncRNAs, conducted bioinformatics analyses to discuss the molecular mechanisms, and established prognostic markers for TNBC patients. These findings could provide great hope for individual treatment and prognosis prediction in TNBC patients.

## Methods

### Datasets

The lncRNA and mRNA expression profiles extracted respectively from the TNBC raw data, which contained data from 159 TNBC and 113 non-tumor tissues, were downloaded from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). After a careful review, two gene expression profiles (GSE135565, n = 83 and GSE103091, n = 107) were downloaded from the Gene Expression Omnibus (GEO) database. Both of the profiles were based on the Agilent GPL570 platform (Affymetrix Human Genome U133 Plus 2.0 Array). The EMT gene set including 22 EMT related inducible factors, transcription factors and signaling pathway genes was obtained from published literatures(16, 17).

### Construction Of Prognostic Model

The prognostic model was established by Lasso regression analysis after selection of EMT related lncRNAs. Since the expression value of each specific gene was included, the risk score formula of each patient was constructed and weighted with its estimated regression coefficient in lasso regression analysis. According to the formula, the patients were divided into low-risk group and high-risk group using the median risk score as the cut-off point. Survival differences between the two groups were assessed by Kaplan-Meier survival curves using log-rank tests. LASSO regression analysis and stratified analysis were applied for examining the role of the risk score in predicting clinical outcomes. ROC curves were used to investigate the accuracy of model prediction.

### Drug Sensitivity Analysis

Based on the GDSC database (<https://www.cancerrxgene.org/>), an R package “pRRophetic” was employed in the chemosensitivity prediction of each tumor samples. The IC50 of each specific chemotherapy drug was estimated by regression method, and the GDSC training dataset was used for 10-fold cross-validation to test the regression and prediction accuracy. Default values were selected for all parameters, including “combat” to remove batch effect and the average value of repeated gene expression.

## Immunocyte Infiltration Analysis

CIBERSORT algorithm was used to analyze RNA-seq data of different TNBC subgroups to infer the relative proportion of 22 immune infiltrating cells, and Spearman correlation analysis was performed on gene expression and immune cell content,  $P < 0.05$  was considered to be statistically significant.

## Gene Set Variation Analysis (gsva)

Gene set variation analysis (GSVA) is a non-parametric and unsupervised method to evaluate the enrichment of transcriptome gene sets. GSVA changes gene level into pathway level by comprehensively scoring the gene set of interest, and then judges the biological function of samples. In this study, we downloaded gene sets from Molecular signatures database (v7.0 version), and scored each gene set comprehensively to evaluate the potential biological function changes of different samples using GSVA algorithm.

## Gene Set Enrichment Analysis (gsea)

Gene Set Enrichment Analysis (GSEA; <http://www.broadinstitute.org/gsea>) on expression profiles of TNBC patients was applied for identifying differentially expressed genes between high-risk and low-risk groups. The maximum and minimum sizes of 500 and 15 genes were used to filter the gene set. After 100 permutations, rich gene sets were obtained ( $P < 0.05$ , false discovery rate [FDR]  $< 0.25$ ).

## Statistical Analysis

Survival curves were generated by Kaplan Meier method and compared by log rank test. Cox proportional hazards model was used for multivariate analysis. All statistical analyses were conducted with the R language (version 3.6.1). All statistical tests were bilateral, and  $p < 0.05$  was statistically significant.

## Results

### Identification of EMT associated lncRNAs in TNBC Cohort

Our study has downloaded the original mRNA expression data of TNBC(FRKM) from TCGA database, and extracted 21 EMT related regulators. Firstly, we screened a total of 14 differential EMT genes in the expression profile between tumor group and normal group by differential analysis ( $\log_{2}FC > 1$  &  $\log_{2}FC < -1$ , and  $p < 0.05$ ), including 7 up-regulated genes and 7 down-regulated genes (Fig. 1A). After that, the expression data of 3234 LncRNAs from TNBC, as well as data from EMT genes, were screened by correlation analysis to find the LncRNAs highly correlated with EMT ( $|r| > 0.3$  &  $P \text{ value} < 0.001$ ). It revealed that a total of 1033 LncRNAs were highly associated with EMT (Table S1). Among them, 20 LncRNAs and 14 EMT genes were randomly selected to show the correlation in the form of heat map (Fig. 1B). Finally, the significantly down-regulated LncRNAs were screened out (the expression level was 0 in more than half of the samples, or the average expression level was less than 0.3 in the samples), and ultimately 536 LncRNAs were used as candidate gene sets for further modeling and analysis.

## Gain Of Prognostic Genes And Construction Of Prognostic Model

In order to further identify the key genes in the screened LncRNAs set, we collected clinical information of TNBC patients and screened out the feature genes in TNBC by Cox univariate regression and lasso regression feature selection algorithm (Fig. 2A-2C). It was demonstrated that 285 LncRNAs (shared genes of candidate gene set and external validated data set) were screened by Cox univariate regression analysis to find the prognostic genes (Table S2), in which 22 prognostic genes with significance ( $P \text{ value} < 0.05$ ) were obtained as follows: YTHDF3-AS1, UBE2E2-AS1, SOCS2-AS1, TINCR, A2M-AS1, CYB561D2, TUG1, NIFK-AS1, LINC00667, NDUFB2-AS1, CASC15, PINK1-AS, ZSCAN16-AS1, EPB41L4A-AS1, TRIM52-AS1, LINC00839, ASB16-AS1, RGS5, LINC01023, SLC16A1-AS1, MBNL1-AS1, LINC01315. The patients from TCGA were randomly divided into training dataset and testing dataset at a ratio of 4:1, and we used lasso regression analysis to get the best risk score value for further analysis (Risk Score =  $NIFK-AS1 \times (-0.33684413) + LINC01315 \times (-0.322333697) + LINC00667 \times (-0.288743288) + ASB16-AS1 \times (-0.161362422) + PINK1-AS \times (-0.07986676) + RGS5 \times 0.169608076 + UBE2E2-AS1 \times 0.26526301 + YTHDF3-AS1 \times 0.268478279 + ZSCAN16-AS1 \times 0.271742665 + SOCS2-AS1 \times 0.371413484 + TINCR \times 0.398092567 + NDUFB2-AS1 \times 0.484513929$ ). According to the median of risk score, patients were divided into high-risk group and low-risk group (median value of TCGA training dataset: -0.209563547920045; median value of TCGA testing dataset: -0.294563095125158) and analyzed by Kaplan-Meier curve. The overall survival (OS) of high-risk group in both sets was significantly lower than low-risk group (Fig. 2D, 2E). Additionally, ROC curve showed that the C-index of both sets are 0.91 and 0.79 respectively, (Fig. 2F, 2G), indicating the model's better verification efficiency.

### Clinical Predictive Value of the Model based on Multi-Omics Analysis

Tumor microenvironment is mainly composed of tumor-associated fibroblasts, immune cells, extracellular matrix, multiple growth factors, inflammatory factors, specific physical and chemical characteristics, and cancer cells. Tumor microenvironment significantly affects the diagnosis, survival

outcome, and sensitivity of clinical treatment in cancers. Through analyzing the relationship between risk score and tumor immune infiltration, we further investigated the potential molecular mechanism of risk score in TNBC development, which demonstrated that risk score was positive related with Macrophages M2, Mast cells resting, NK cells activated, Mast cells activated, etc., and negative correlated with T cells CD4 memory activated, Dendritic cells resting, T cells CD4 memory resting, B cells naive, etc. (Fig. 3A). Since surgery combined with chemotherapy is effective in early breast cancer, our research was based on the drug sensitivity data of GDSC database, and the sensitivity of each tumor sample was predicted by R-package "pRRophetic" to further explore the relationship between risk score and sensitivity of common chemotherapy drugs. The results showed that risk score significantly affected the sensitivity of patients to Bicalutamide, Bryostatins, Dasatinib, Gefitinib, Lapatinib, and Metformin (Fig. 3B). By investigating the mutation spectrum of high/low-risk groups, we found that there was significant difference between the two groups in the mutation proportion of multiple genes (Fig. 3C).

## Prognostic Model Related Signal Mechanism

Subsequently, we analyzed the signaling pathways involved in high/low-risk models to explore the potential molecular mechanism of risk score affecting tumor progression. Results of GSVA revealed that the differential pathways of the two groups were mainly enriched in UV\_RESPONSE\_UP, ADIPOGENESIS, UNFOLDED\_PROTEIN\_RESPONSE, HYPOXIA, APICAL\_SURFACE, and other pathways (Fig. 4A). Finally, we found that there were significant enrichments in various related pathways through GSEA analysis. Some of the highly significant signaling pathways were shown (Fig. 4B, 4C), which suggested that the disturbance of these signaling pathways in high/low-risk groups affected the prognosis of TNBC.

## Robustness Analysis By External Datasets

We downloaded the data of TNBC patients with survival data processed in GEO database (GSE135565 and GSE103091), predicted the clinical classification of TNBC based on the model, evaluated the survival differences between two groups through Kaplan-Meier analysis, and investigated the stability of the prediction model. The results demonstrated that the OS of high-risk group was obviously lower than low-risk group in both GEO external verification sets (Fig. 5A, 5B). In order to verify the accuracy of the model, we did ROC curve analysis using external data sets, which showed that the model had a strong efficiency on prognostic prediction (GSE135565-C-index = 0.72, GSE103091-C-index = 0.65) (Fig. 5C, 5D).

## Risk And Independent Prognostic Analysis

Since the samples were divided into high/low-risk groups by the median value of risk score, the results of regression analysis were displayed by nomogram. Among which, the logistic regression analysis revealed that the different stages of TNBC were obviously associated with the distribution of risk score value obtained by our model analysis in all the samples (Fig. 6A). Through the analysis of general linear model

(GLM) and Cox proportional hazards (CoxPH) model, we found that the distribution of risk score and several clinical parameters (such as age, stage, T, etc.) had different contributions to the scoring in distinct stages of cancer (Fig. 6B, 6D). At the same time, we also did some prediction analyses on the five-year and seven-year periods (Fig. 6C), and identified that risk score was an independent prognostic factor for TNBC patients through univariate and bivariate analysis (Fig. 6E, 6F).

## Correlation Analysis Of Risk And Multiple Clinical Parameters

We grouped all the risk score values by different clinical parameters, which was shown in the form of boxplot graph (Fig. 7A-7D), and found that these risk scores were significant among the groups with multiple clinical indicators through Kruskal-Wallis test ( $P < 0.05$ ) (Fig. 7A, 7C). As the risk score rose, the stage grade and lymph node involvement increased.

## Discussion

TNBC has remained an unmet medical challenge for decades, since prone to recurrence and metastasis after operation, and no therapeutic targets have been identified(18, 19). It is widely acknowledged that metastasis of TNBC is correlated with aberrant activation of EMT(20). EMT is a multistep, plastic and reversible process that allows tumor cells acquire a mesenchymal phenotype(21). The important characteristics of EMT includes the downregulation of cell-adhesion molecules (such as E-cadherin), activation of transcription factor (such as Snail2), and upregulation of mesenchymal cell markers (such as vimentin)(22). However, the entirety accomplishment of the EMT progression demands an intricate genetic procedure, and the precise role of transcriptional and epigenetic regulators in modulating diverse EMT processes in tumorigenesis (including TNBC) are still not fully understood(19, 21, 23). Recent studies have focused on the biological role of lncRNAs in malignant evolution and EMT. With their multifunction, lncRNAs are proved to be related with EMT in a wide spectrum of physiological and pathological processes(24). The promoting and suppressing effects of lncRNAs on EMT underly the complexity and plasticity of tumor cells(25). For example, lncRNA CAR10 was reported to be an EMT promoter. CAR10 could induce EMT by directly binding with miR-30 and miR-203, and then regulating the expression of Snail1 and Slug in lung adenocarcinoma metastasis(26). On the contrary, Han et.al(27) investigated the inhibitory effect of lncRNA CRCMSL in colorectal cancer. They pointed out that CRCMSL could bind to protein HMGB2 and stabilize the localization in the cytoplasm, hence attenuating the interaction between HMGB2 and OCT4 and inhibiting EMT. In TNBC, multiple lncRNAs had been identified to regulate EMT pathways and tumor invasion via interacting with various molecules, such as LINC01638(28), GAS5(29), UCA1(30), ARNILA(31), and NNT-AS1(32). A better understanding of how lncRNAs regulate EMT process at diverse molecular levels can accelerate the development of therapeutic strategies and prognostic targets.

Currently, there are some applications of risk models with prognostic function in clinical. The most widely used model is the 21-gene expression assay (Oncotype DX, Genomic Health), which can provide prognostic information in hormone-receptor–positive breast cancer(33). Nevertheless, there is still lack of simple and effective prognostic prediction model in TNBC. Researchers have begun to pay close attention to establish signatures with combination of coding and noncoding RNAs in clinical prognosis. Recently, Lin et.al(34) constructed a hypoxia signature in glioma groups. The hypoxia risk model could reflect overall immune response intensity of tumor microenvironment and predict prognosis. Another research established a m6A-related lncRNA prognostic signature, which could predict the OS of lower-grade glioma patients(35). Furthermore, Hong et.al(36) identified a novel signature. Unlike previous strategies, they paid attention to the immune-related gene pairing and built a reasonable model with two-lncRNA combinations to predict the immune landscape in hepatocellular carcinoma.

In this study, we firstly established a novel risk score prediction model based on EMT-related lncRNAs in TNBC. In combination with TCGA and GEO database, along with 14 screened EMT factors, we performed a differential co-expression analysis to classify 536 candidate lncRNAs. Twenty-two of them were confirmed to have prognostic value in both datasets and used to establish a model for predicting the OS of TNBC patients. According to the median value of risk score, patients were divided into high/low-risk groups with significant difference of OS. Our results demonstrated that the risk score was an independent risk factor in TNBC. Since the prediction model was preliminary built, its accuracy and efficacy were carefully compared and verified from several aspects, including tumor immune infiltration, drug sensitivity, mutability spectrum, signaling pathways, and clinical parameters (age, stage, grade, clinical classification, lymph nodes involvement et.al).

Among the lncRNAs involved in the model, several of them were reported to be associated with tumor progression, such as lncRNA TINCR(37–39) and TUG1(40–42). A recent study revealed that serum lncRNA TINCR level was significantly increased in TNBC and correlated with clinical outcome(43). Tang et.al(44) reported that lncRNA TUG1 could act as a miR-197 sponge to enhance cisplatin sensitivity in TNBC. Additionally, LINC01315 was newly identified as a prognostic biomarker in TNBC(45). However, most of the lncRNAs were not fully investigated in TNBC, and we hope that EMT-related lncRNAs might target novel insights in TNBC development.

On the other side, there existed several shortcomings and limitations. For instance, the raw data used for our analysis was totally obtained from TCGA and GEO database, which meant the original information was incomplete and lack of regional specificity, making the final model unreliable. More independent TNBC cohorts should be collected for further validation. Besides, the roles of the identified lncRNAs in TNBC progression need to be clarified in clinical samples and biological experiments.

## Conclusions

In conclusion, our study demonstrated that an effective prognosis model constructed by EMT-related lncRNAs could serve as an independent risk factor, and provide new strategies for TNBC patients.

# Abbreviations

TNBC: triple negative breast cancer

ER: estrogen receptor

PR: progesterone receptor

HER2: human epidermal growth factor receptor 2

ICI: immune checkpoint inhibition

EMT: epithelial–mesenchymal transition

LncRNAs: long non-coding RNAs

ncRNAs: non-coding RNAs

TCGA: The Cancer Genome Atlas

GEO: Gene Expression Omnibus

GSVA: gene set variation analysis

GSEA: gene Set Enrichment Analysis

OS: overall survival

GLM: general linear model

CoxPH: Cox proportional hazards

# Declarations

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data and materials

The datasets analyzed during the current study are available in the TCGA (<https://portal.gdc.cancer.gov/>, TNBC dataset) and GEO (<https://www.ncbi.nlm.nih.gov/geo/>, GSE135565 and GSE103091 dataset) repository. Data from screening are included in the Additional files.

## Competing interests

The authors declare that they have no competing interests.

## Funding

This work was supported by the Six Talent Peaks Project in Jiangsu Province (No. LGY2017051).

## Authors' contributions

MH and WS designed the research. JG and XY performed the analyses. ZJ, MY and YY collected and analyzed part of the data. JG wrote the manuscript. MH critically commented and edited the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

Not applicable.

## Authors' information

MH is the head of department I of medical oncology in the hospital, and chief physician specializing in breast cancer.

## References

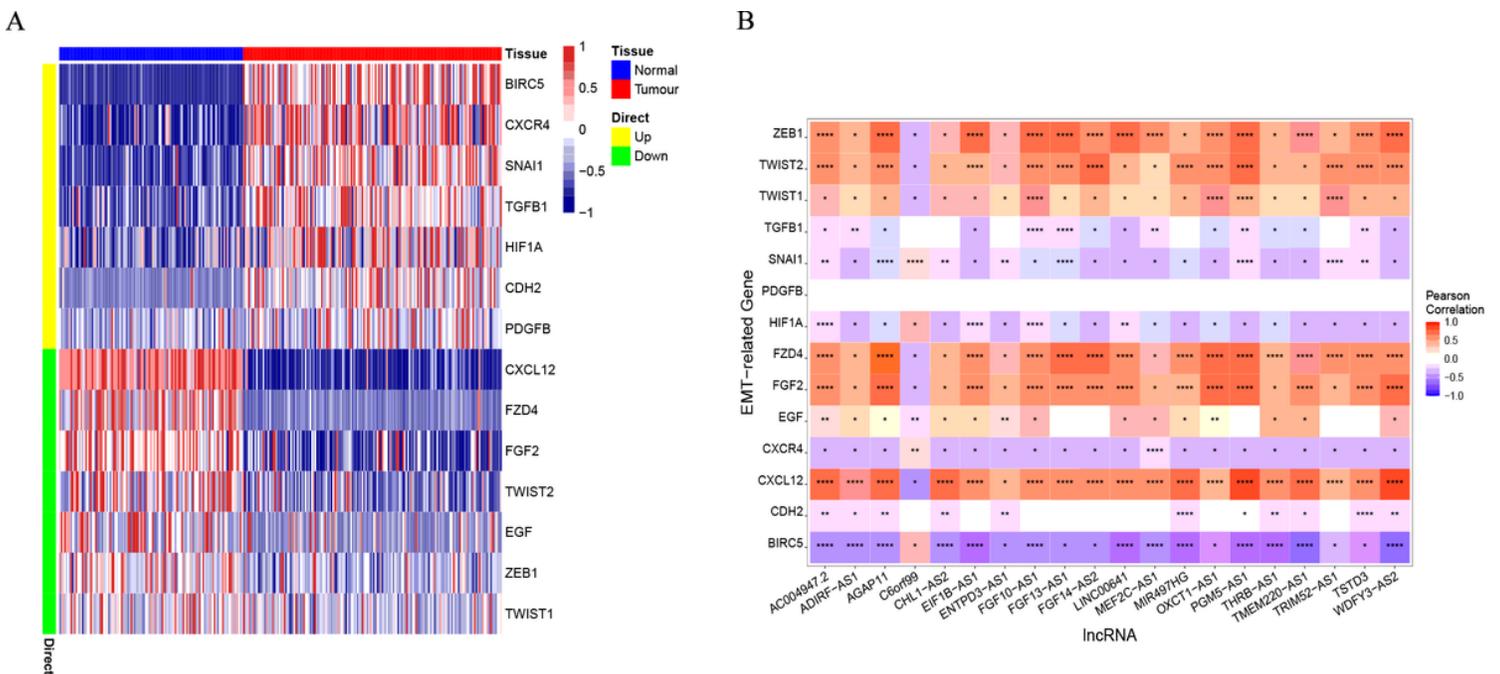
1. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. *N Engl J Med*. 2010;363(20):1938–48.
2. Gierach GL, Burke A, Anderson WF. Epidemiology of triple negative breast cancers. *Breast Dis*. 2010;32(1–2):5–24.
3. Kong X, Liu W, Kong Y. Roles and expression profiles of long non-coding RNAs in triple-negative breast cancers. *J Cell Mol Med*. 2018;22(1):390–4.
4. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. *Clin Cancer Res*. 2007;13(15 Pt 1):4429–34.
5. Dent R, Hanna WM, Trudeau M, Rawlinson E, Sun P, Narod SA. Pattern of metastatic spread in triple-negative breast cancer. *Breast Cancer Res Treat*. 2009;115(2):423–8.
6. Park JH, Ahn JH, Kim SB. How shall we treat early triple-negative breast cancer (TNBC): from the current standard to upcoming immuno-molecular strategies. *ESMO Open*. 2018;3(Suppl 1):e000357.
7. Wein L, Loi S. Mechanisms of resistance of chemotherapy in early-stage triple negative breast cancer (TNBC). *Breast*. 2017;34(Suppl 1):27–30.
8. Mediratta K, El-Sahli S, D'Costa V, Wang L. Current Progresses and Challenges of Immunotherapy in Triple-Negative Breast Cancer. *Cancers (Basel)*. 2020;12(12).

9. Denkert C, Liedtke C, Tutt A, von Minckwitz G. Molecular alterations in triple-negative breast cancer—the road to new treatment strategies. *Lancet*. 2017;389(10087):2430–42.
10. Jia H, Truica CI, Wang B, Wang Y, Ren X, Harvey HA, et al. Immunotherapy for triple-negative breast cancer: Existing challenges and exciting prospects. *Drug Resist Updat*. 2017;32:1–15.
11. Aiello NM, Kang Y. Context-dependent EMT programs in cancer metastasis. *J Exp Med*. 2019;216(5):1016–26.
12. Zhang H, Wang J, Yin Y, Meng Q, Lyu Y. The role of EMT-related lncRNA in the process of triple-negative breast cancer metastasis. *Biosci Rep*. 2021;41(2).
13. Felipe Lima J, Nofech-Mozes S, Bayani J, Bartlett JM. EMT in Breast Carcinoma—A Review. *J Clin Med*. 2016;5(7).
14. Kopp F, Mendell JT. Functional Classification and Experimental Dissection of Long Noncoding RNAs. *Cell*. 2018;172(3):393–407.
15. Zhang W, Guan X, Tang J. The long non-coding RNA landscape in triple-negative breast cancer. *Cell Prolif*. 2021;54(2):e12966.
16. Iser IC, Pereira MB, Lenz G, Wink MR. The Epithelial-to-Mesenchymal Transition-Like Process in Glioblastoma: An Updated Systematic Review and In Silico Investigation. *Med Res Rev*. 2017;37(2):271–313.
17. Meel MH, Schaper SA, Kaspers GJL, Hulleman E. Signaling pathways and mesenchymal transition in pediatric high-grade glioma. *Cell Mol Life Sci*. 2018;75(5):871–87.
18. Gong Y, Ji P, Yang YS, Xie S, Yu TJ, Xiao Y, et al. Metabolic-Pathway-Based Subtyping of Triple-Negative Breast Cancer Reveals Potential Therapeutic Targets. *Cell Metab*. 2021;33(1):51–64 e9.
19. Khaled N, Bidet Y. New Insights into the Implication of Epigenetic Alterations in the EMT of Triple Negative Breast Cancer. *Cancers (Basel)*. 2019;11(4).
20. Neelakantan D, Zhou H, Oliphant MUJ, Zhang X, Simon LM, Henke DM, et al. EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. *Nature communications*. 2017;8:15773.
21. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016 Cell. 2016;166(1):21–45.
22. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. *Nat Rev Cancer*. 2018;18(2):128–34.
23. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. *Nature*. 2018;556(7702):463–8.
24. Cheng JT, Wang L, Wang H, Tang FR, Cai WQ, Sethi G, et al. Insights into Biological Role of lncRNAs in Epithelial-Mesenchymal Transition. *Cells*. 2019;8(10).
25. Gugnoni M, Ciarrocchi A. Long Noncoding RNA and Epithelial Mesenchymal Transition in Cancer. *Int J Mol Sci*. 2019;20(8).
26. Ge X, Li GY, Jiang L, Jia L, Zhang Z, Li X, et al. Long noncoding RNA CAR10 promotes lung adenocarcinoma metastasis via miR-203/30/SNAI axis. *Oncogene*. 2019;38(16):3061–76.

27. Han Q, Xu L, Lin W, Yao X, Jiang M, Zhou R, et al. Long noncoding RNA CRCMSL suppresses tumor invasive and metastasis in colorectal carcinoma through nucleocytoplasmic shuttling of HMGB2. *Oncogene*. 2019;38(16):3019–32.
28. Luo L, Tang H, Ling L, Li N, Jia X, Zhang Z, et al. LINC01638 lncRNA activates MTDH-Twist1 signaling by preventing SPOP-mediated c-Myc degradation in triple-negative breast cancer. *Oncogene*. 2018.
29. Li S, Zhou J, Wang Z, Wang P, Gao X, Wang Y. Long noncoding RNA GAS5 suppresses triple negative breast cancer progression through inhibition of proliferation and invasion by competitively binding miR-196a-5p. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2018;104:451-7.
30. Li GY, Wang W, Sun JY, Xin B, Zhang X, Wang T, et al. Long non-coding RNAs AC026904.1 and UCA1: a "one-two punch" for TGF- $\beta$ -induced SNAI2 activation and epithelial-mesenchymal transition in breast cancer. *Theranostics*. 2018;8(10):2846–61.
31. Yang F, Shen Y, Zhang W, Jin J, Huang D, Fang H, et al. An androgen receptor negatively induced long non-coding RNA ARNILA binding to miR-204 promotes the invasion and metastasis of triple-negative breast cancer. *Cell death differentiation*. 2018;25(12):2209–20.
32. Li Y, Lv M, Song Z, Lou Z, Wang R, Zhuang M. Long non-coding RNA NNT-AS1 affects progression of breast cancer through miR-142-3p/ZEB1 axis. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2018;103:939 – 46.
33. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N Engl J Med*. 2018;379(2):111–21.
34. Lin W, Wu S, Chen X, Ye Y, Weng Y, Pan Y, et al. Characterization of Hypoxia Signature to Evaluate the Tumor Immune Microenvironment and Predict Prognosis in Glioma Groups. *Frontiers in Oncology*. 2020;10(796).
35. Tu Z, Wu L, Wang P, Hu Q, Tao C, Li K, et al. N6-Methyladenosine-Related lncRNAs Are Potential Biomarkers for Predicting the Overall Survival of Lower-Grade Glioma Patients. *Frontiers in Cell and Developmental Biology*. 2020;8(642).
36. Hong W, Liang L, Gu Y, Qi Z, Qiu H, Yang X, et al. Immune-Related lncRNA to Construct Novel Signature and Predict the Immune Landscape of Human Hepatocellular Carcinoma. *Molecular therapy Nucleic acids*. 2020;22:937–47.
37. Xu TP, Liu XX, Xia R, Yin L, Kong R, Chen WM, et al. SP1-induced upregulation of the long noncoding RNA TINCR regulates cell proliferation and apoptosis by affecting KLF2 mRNA stability in gastric cancer. *Oncogene*. 2015.
38. Xu TP, Wang YF, Xiong WL, Ma P, Wang WY, Chen WM, et al. E2F1 induces TINCR transcriptional activity and accelerates gastric cancer progression via activation of TINCR/STAU1/CDKN2B signaling axis. *Cell Death Dis*. 2017;8(6):e2837.
39. Wang Q, Liu J, You Z, Yin Y, Liu L, Kang Y, et al. lncRNA TINCR favors tumorigenesis via STAT3-TINCR-EGFR-feedback loop by recruiting DNMT1 and acting as a competing endogenous RNA in human breast cancer. *Cell Death Dis*. 2021;12(1):83.

40. Xu Y, Ge Z, Zhang E, Zuo Q, Huang S, Yang N, et al. The lncRNA TUG1 modulates proliferation in trophoblast cells via epigenetic suppression of RND3. *Cell Death Dis.* 2017;8(10):e3104.
41. Sun J, Hu J, Wang G, Yang Z, Zhao C, Zhang X, et al. LncRNA TUG1 promoted KIAA1199 expression via miR-600 to accelerate cell metastasis and epithelial-mesenchymal transition in colorectal cancer. *J Exp Clin Cancer Res.* 2018;37(1):106.
42. Shen X, Hu X, Mao J, Wu Y, Liu H, Shen J, et al. The long noncoding RNA TUG1 is required for TGF-beta/TWIST1/EMT-mediated metastasis in colorectal cancer cells. *Cell Death Dis.* 2020;11(1):65.
43. Wang X, Li S, Xiao H, Deng X. Serum lncRNA TINCR Serve as a Novel Biomarker for Predicting the Prognosis in Triple-Negative Breast Cancer. *Technology in cancer research treatment.* 2020;19:1533033820965574.
44. Tang T, Cheng Y, She Q, Jiang Y, Chen Y, Yang W, et al. Long non-coding RNA TUG1 sponges miR-197 to enhance cisplatin sensitivity in triple negative breast cancer. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie.* 2018;107:338 – 46.
45. Naorem LD, Prakash VS, Muthaiyan M, Venkatesan A. Comprehensive analysis of dysregulated lncRNAs and their competing endogenous RNA network in triple-negative breast cancer. *Int J Biol Macromol.* 2020;145:429–36.

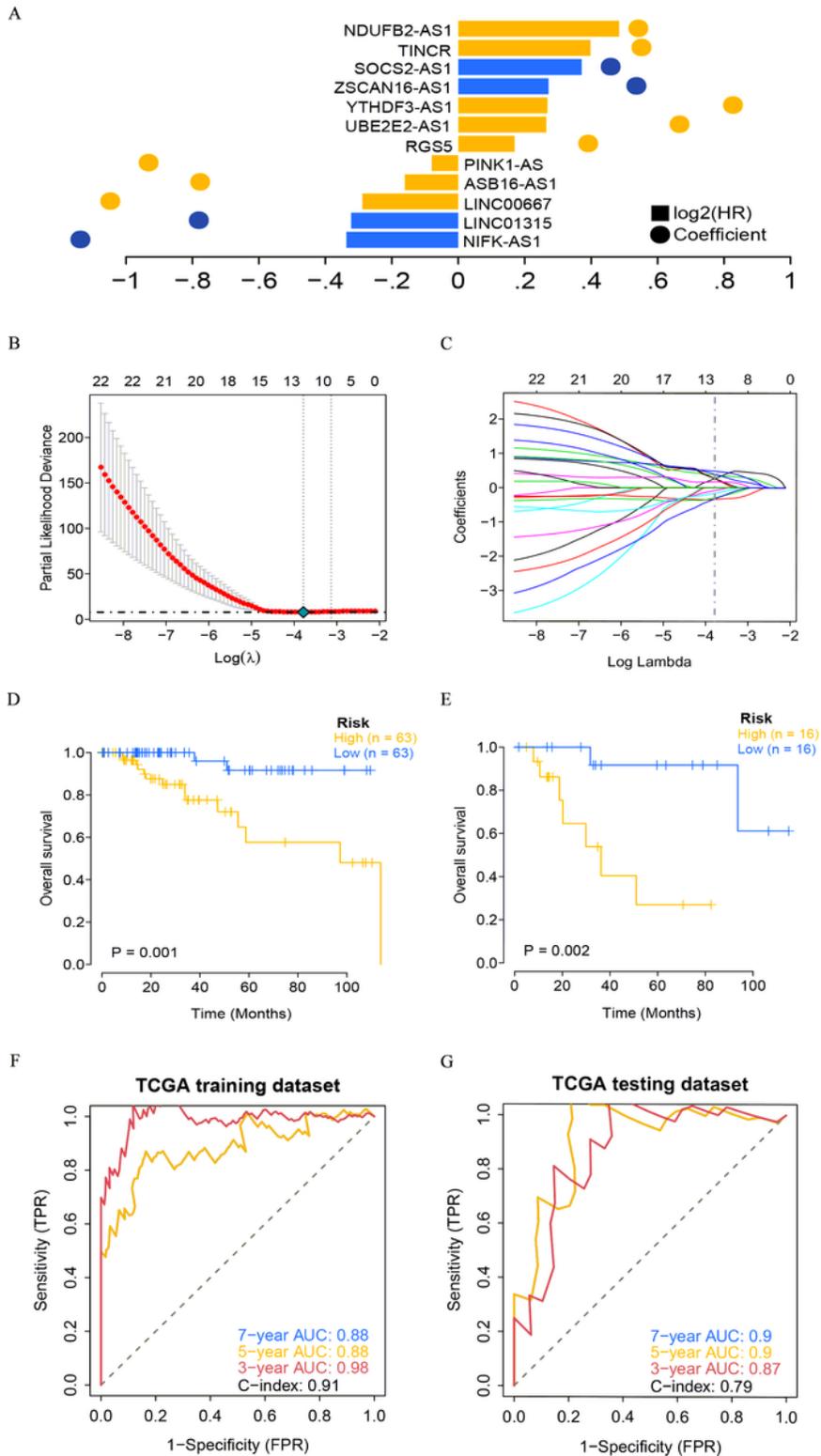
## Figures



**Figure 1**

Identification of candidate EMT-related lncRNAs in TNBC. A. Identification of 14 differential EMT genes in the expression profile between tumor group and normal group by differential analysis ( $\log_{2}FC > 1$  &  $\log_{2}FC$

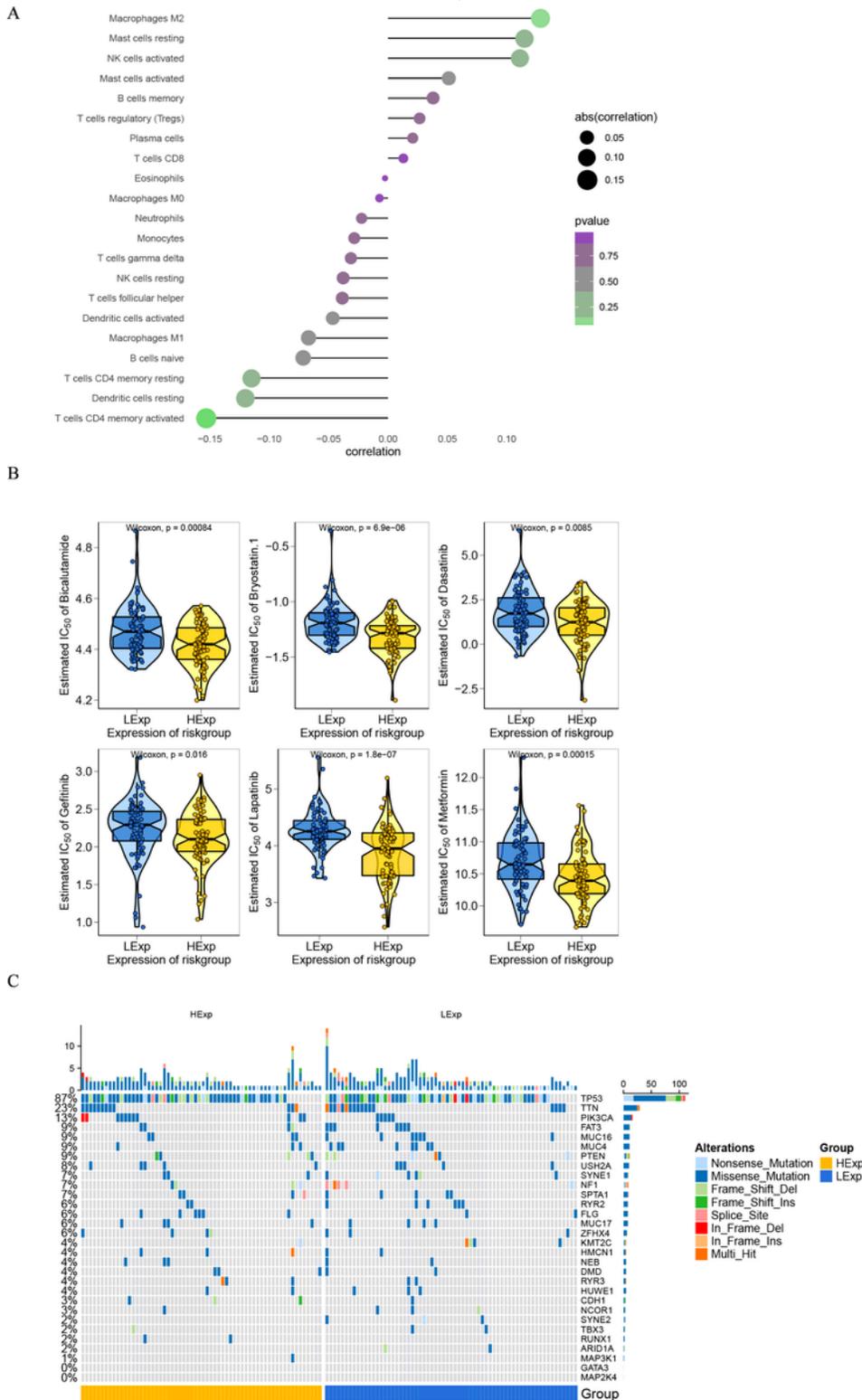
< -1, and  $p < 0.05$ ), including 7 up-regulated genes and 7 down-regulated genes. B. Correlation between 20 randomly selected LncRNAs and 14 EMT-related genes.



**Figure 2**

Prognostic genes used for model construction. A-C. Prognostic lncRNAs were screened out from TNBC data by A, C. Coefficient by lasso regression analysis, and B. Cox univariate regression analysis. D, E. The overall survival (OS) of high-risk group in both sets was significantly lower than low-risk group analyzed

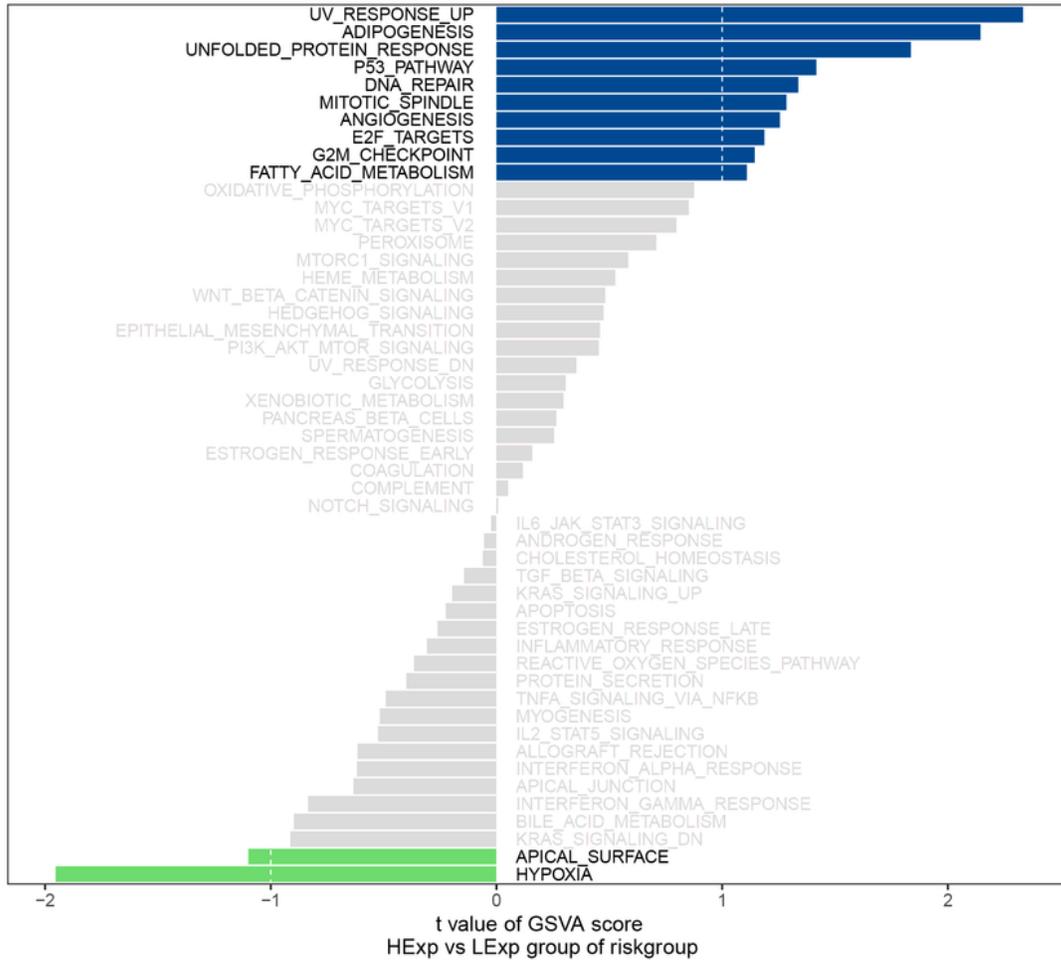
by Kaplan-Meier curve. (Left: TCGA training dataset; Right: TCGA testing dataset). F, G. The model's efficiency evaluated by ROC curve. (Left: TCGA training dataset; Right: TCGA testing dataset).



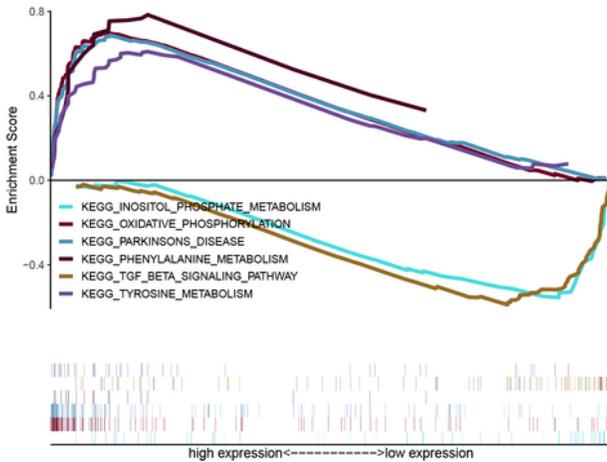
**Figure 3**

Clinical Predictive Value of the Model. A. Relationship between the model and tumor immune infiltration. B. Relationship between the model and sensitivity of common chemotherapy drugs. C. Mutation spectrum of high/low-risk groups (Left: High-risk group; Right: Low-risk group).

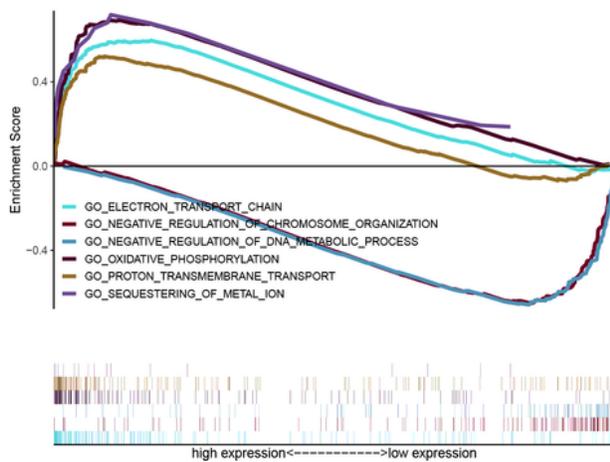
A



B

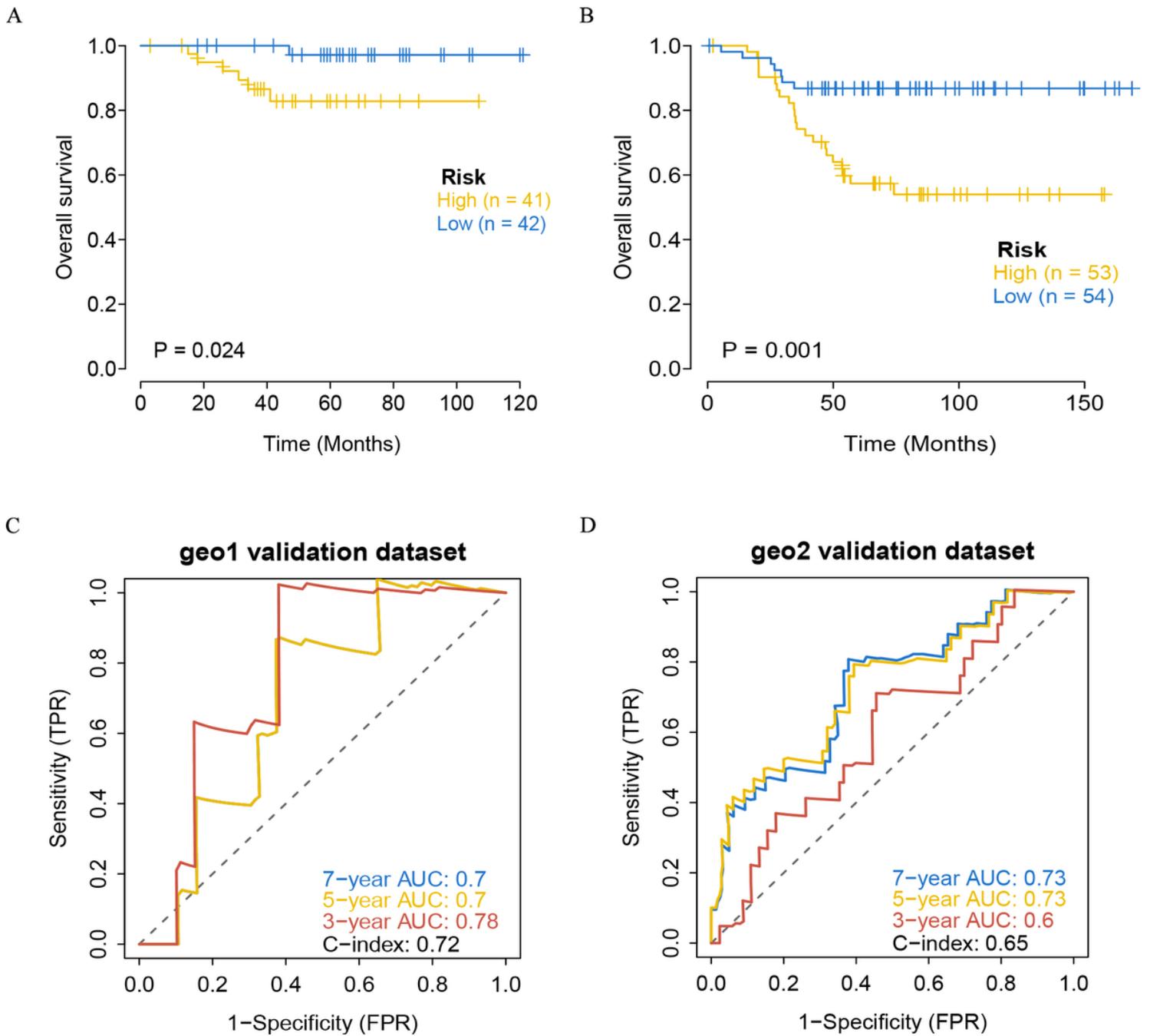


C



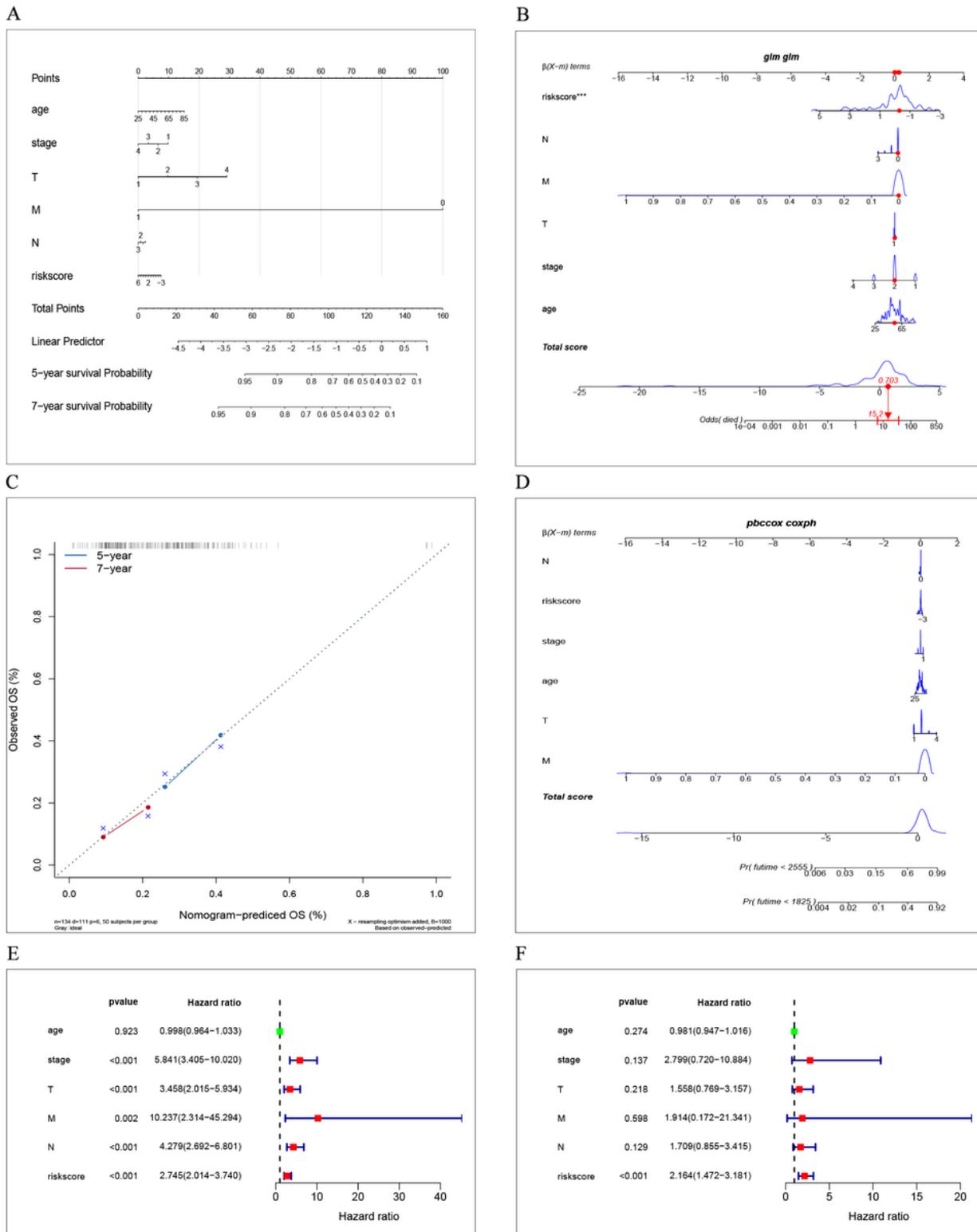
**Figure 4**

Prognostic model associated signaling pathways. A. Results of GSVA showed the differential pathways of high/low-risk groups. B. Results of GSEA showed the significant enrichments in various related pathways by KEGG. C. Results of GSEA showed the significant enrichments in various related pathways by GO.



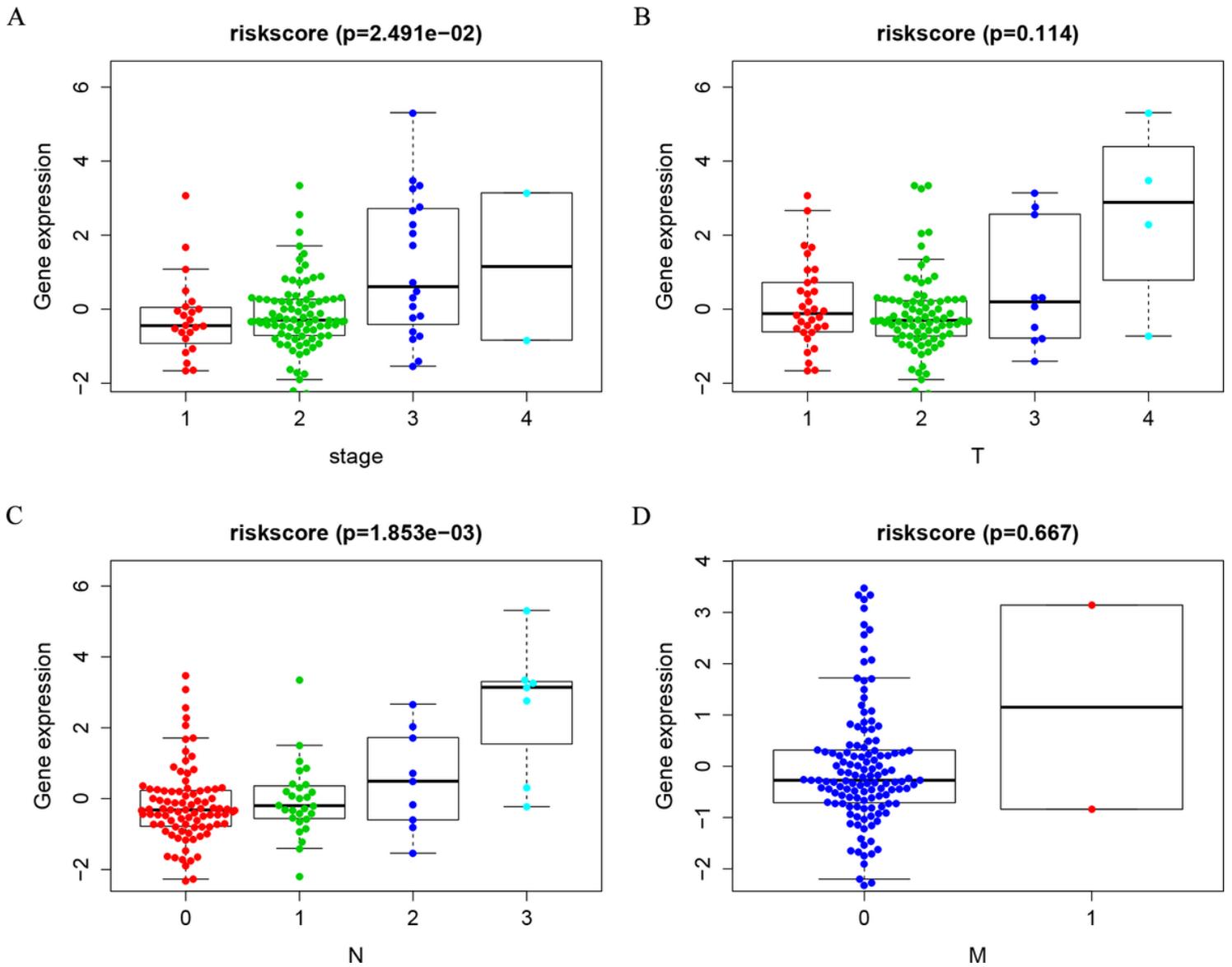
**Figure 5**

Robustness Analysis by External Datasets. A. The survival differences between high/low-risk groups evaluated by Kaplan-Meier analysis in GSE135565. B. The survival differences between high/low-risk groups evaluated by Kaplan-Meier analysis in GSE103091. C. Prediction efficiency of the model verified by ROC curve in GSE135565. D. Prediction efficiency of the model verified by ROC curve in GSE103091.



**Figure 6**

Risk and Independent Prognostic Analysis. A. Logistic regression analysis showed the relationship between TNBC stages and distribution of risk score value. B, D. The effect of distribution of risk score and clinical parameters on TNBC stage scoring analyzed by B. general linear model, and D. Cox proportional hazards model. C. Prediction analyses on the 5-year and 7-year periods. E, F. Risk score as an independent prognostic factor proved by E. univariate, and F. bivariate analysis.



**Figure 7**

Correlation analysis of risk score and clinical parameters A-D. Risk score values grouped by different clinical parameters and analyzed by through Kruskal-Wallis test (A. Stage; B. Tumor; C. Lymph Node; D. Metastasis).

## Supplementary Files

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

- [SupplementaryTable1.xlsx](#)
- [SupplementaryTable2.xlsx](#)