# Identification of Prognostic Immune-Related lncRNAs Pairs in Immune Infiltration of Breast Cancer

**Qianhui Xu**
Yuying Children's Hospital of Wenzhou Medical University
**Shaohuai Chen**
Yuying Children's Hospital of Wenzhou Medical University
**Yuanbo Hu**
Yuying Children's Hospital of Wenzhou Medical University
**Wen Huang** ( qq2627897941@126.com )
Yuying Children's Hospital of Wenzhou Medical University

**Research Article**

**Keywords:** immune-related long noncoding RNAs, breast cancer, prognostic value, tumor immune microenvironment, immunotherapy

**Posted Date:** August 9th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-753945/v1

# Abstract

**Background:** Accumulating evidence has supported that long non-coding ribonucleic acids (lncRNAs) could act as essential regulators in cancer immunity. An immune-related lncRNAs (irlncRNAs) risk signature without specific expression value was established to accurately predict prognosis in patients with breast cancer (BRCA).

**Methods:** First, irlncRNAs were identified using co-expression analysis and differential expressed irlncRNAs (DEirlncRNAs) were recognized by “Limma” package. Then, DEirlncRNA pairs were determined using an iteration loop and a 0-or-1 screening matrix. Next, single factor test and Lasso algorithm followed by multivariate Cox regression were employed to establish risk signature. Besides, the Akaike information criterion (AIC) values were calculated to recognize the optimal cut-off point of low- or high-risk groups. Additionally, the potential role of risk score was explored in terms of overall survival, clinical variables, tumor immune microenvironment features, immunotherapeutic targets, and TMB (tumor mutation board) statues.

**Results:** A total of 946 irlncRNAs were determined and 188 DEirlncRNAs were identified. 15 DEirlncRNA pairs were introduced into establishment of prognostic signature, which harbored powerful and independent predictive prognostic ability. Taking advantage of AIC values, risk model was demonstrated to accurately distinguish samples from the viewpoint of clinical outcome, clinicopathological features, infiltrating immune cells, and immunosuppressive biomarkers. Besides, comprehensive prognostic nomogram was constructed to quantitatively estimate risk. Finally, synergistic effect of risk score with TMB value was corroborated.

**Conclusions:** The DEirlncRNA pairs risk signature without specific expression value possessed excellent prognostic performance and may provide direction for immunotherapy in BRCA.


# 1. Background

The morbidity of breast cancer (BRCA), one of the most malignant cancer, are reported to be 2.1 million in over 100 countries. Additionally, The disease was also the leading cause of women tumor-related death in 2018 worldwide(1). Despite considerable improvements in clinical comprehensive treatments, the cruel reality is that prognosis of breast cancer is still unsatisfactory, mainly caused by metastasis(2, 3). Additionally, BRCA was well characterized with high intra- and inter-tumor heterogeneity and of which pathological manifestations and medicinal outcome vary from person to person(4). It was, therefore, of great urgency to identify the molecular indicators underlying tumor progression to contribute insights into improving prognosis for advance therapeutic purposes.

The recent clinical success of immune checkpoint therapy in multiple type of malignancies presented its potential in lifesaving(5). Nevertheless, immunotherapy has yet to realize its full potential in BRCA, the most formidable challenge against immunotherapy was its immunologically quiescence. Given immunotherapy harnessed the immune system to battle tumor, recent researches have placed emphasis

on tumor infiltrating immune cells (TIICs) as well as their functional programs in progression and recurrence in BRCA(6). Some studies have shown that patients with early-stage triple-negative breast cancers harboring greater numbers of T cells have a substantially improved prognosis and longer overall survival(7, 8). These results suggested the pivotal roles of the tumor immune microenvironment (TIME) in shaping the progression of tumor and efficacy immunotherapy.

The long non-coding RNAs (lncRNAs) are characterized with transcripts of the length of 200 nucleotides or more and do not get translated into proteins(9). The role of lncRNAs playing in the process of cancer immunity is being explained step-by-step. So far, a considerable amount of studies have reached a consensus that lncRNAs play an extensive regulatory role in cancer immunity, such as antigen release, immune cell migration and in filtration, antigen presentation, and immune activation (10, 11). Mineo Marco et al reported that lNCR1 knockdown sensitizes tumor cells to cytotoxic T cell-mediated killing, improving CAR T cell therapy (12). Another research found that interference lncRNA SNHG1 could inhibit the differentiation of Treg cells via regulating miR-448/IDO, thereby impeding the immune escape of breast cancer (BRCA)(13). Moreover, there were accumulating establishment of prognostic signatures based on multi genes and corresponding gene expression in prognosis prediction, and recurrence monitoring of cancer (14, 15). Notably, combinations of two-indicator harbored superiority of prognostic accuracy compared with simple genes(16). However, few risk signatures have revealed the function of lncRNAs in this setting, especially in BRCA.

Herein, a novel and robust modeling algorithm, paring, and iteration were employed to establish an immune-related lncRNAs signature without requirement of any specific expression levels. The prognostic performance of risk signature was validated using comprehensive analyses. Additionally, prognostic risk-clinical nomogram was developed to facilitate clinical practice. Furthermore, the potential role of risk signature in TIME characterization and immunotherapy was investigated. Finally, the synergistic effect of risk score with TMB was demonstrated. In summary, immune-related lncRNAs risk score was explored to serve as robust predictive indicator and prognostic biomarker, contributing insight into immunotherapeutic treatment for BRCA.

## 2. Materials And Methods

# 2.1 Collection of RNaseq Data and Somatic Mutation Data

mRNA expression profiling for BRCA samples compared with normal tissues were obtained from The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/). Their corresponding clinical information were also downloaded from the TCGA-BRCA project. The mutation files which were obtained through the “varscan variant aggregation and masking” platform for subsequent analysis. We prepared the Mutation Annotation Format (MAF) of somatic variants and implemented the “maftools”(17) R package. There was no necessity to obtain Ethics Committee approval since all information were publicly available and open-access.

# 2.2 Identification of Immune-Related lncRNAs(irlncRNAs)

Annotation files were downloaded from Ensembl (http://asia.ensembl.org) to identify the protein-coding genes and non-coding genes for subsequent analysis. Briefly, genes were distinguished as mRNAs or lncRNAs based on their Refseq IDs or Ensembl IDs, and only the long non-coding genes in NetAffx Annotation files were retained. The immune-related genes (ir-genes) was obtained through the ImmPort data portal (http://www.immport.org). And a list of 2,483 immune-related genes were employed to determine irlncRNAs by using correlation analysis between all lncRNAs and ir-genes. The lncRNAs with square of correlation coefficient $|R| > 0.4$ and $P < 0.001$ was recognized as irlncRNAs.

## 2.3 Distinguish of DEirlncRNAs (Differential Expressed Immune-Related lncRNAs)

Taking advantage of the “Limma” package with $|log2FC| > 1$ and False Discovery Rate (FDR) < 0.05, the DEirlncRNAs between tumor tissues and normal samples were screened. With the help of package “pheatmap”, heatmap was plotted to present the expression difference.

## 2.4 Establishment of DEirlncRNAs Pairs

The DEirlncRNAs were cyclically singly paired, and a 0-or-1 matrix was constructed assuming C is equal to lncRNA A plus lncRNA B; C is defined as 1 if the expression level of lncRNA A is higher than lncRNA B, otherwise C is defined as 0. Then, the constructed 0-or-1 matrix was further screened. No relationship was considered between pairs and prognosis if the expression quantity of lncRNA pairs was 0 or 1 because pairs without a certain rank could not properly predict patient survival outcome. When the amounts of lncRNA-pairs of which expression quantity was 0 or 1 accounted for more than 20% of total pairs, it was considered a valid match.

## 2.5 Construction of Risk Signature

Firstly, univariate Cox regression analysis was implemented, followed by least absolute shrinkage and selection operator (LASSO) regression using the R “glmnet” package. Next, the frequency of each DEirlncRNAs pair in the 1,000-times-repeated Lasso regression was recorded and pairs with frequency more than 100 times were introduced into Cox proportional hazard regression analysis then establishment of risk signature. The AUC value of each risk signature was also computed and presented in ROC curve. The calculation procedure was terminated until ROC curve obtained the maximum AUC value which suggesting risk signature the best candidate. The prognostic ability of risk score for 1/2/3-year overall survival was evaluated by plotting ROC curves and calculating the AUC values. The riskScore based on constructed risk signature was calculated as the following formula: $\text{RiskScore} = \sum ki = 1 \beta i Si$. The AIC values of every points of the 1-year OS ROC curve were counted to determine the maximum inflection point, which was employed as the cut-off point to stratified samples into high- or low-risk groups.

## 2.6 Validation of Risk Signature

First, K–M survival analyses were performed to recognize the overall survival distinction between low- or high-risk samples. Furthermore, univariate and multivariate Cox regression were employed for prognostic

validity of risk score as an independent indicator. The R packages “survival”, “survminer”, and “pHeatmap” were employed in above analysis.

## 2.7 Risk Score with Clinical Characteristics

To elucidate the clinical significance of risk score, chi-square test was performed to reveal correlation of risk score with such main clinicopathological variables as gender, age, pathological staging, and TNM categories was performed. Wilcoxon signed-rank test was used to compare the riskScore among distinct subgroups based on clinicopathological features. To further validate whether constructed signature remained great prognostic validity when BRCA samples assigned into distinct subgroups according to clinical characteristics, stratification survival analysis was conducted. The R packages “ComplexHeatmap”, “limma”, “ggpubr”, “survival”, and “survminer” were employed in above analysis.

## 2.8 Development of Prognostic Nomogram

To construct a quantitative risk model to predicting overall survival rate, a nomogram including risk score and other clinical variables to predict 1/2/3-OS probability. Subsequently, the calibration curve which shown the prognostic value of as-constructed nomogram was developed.

## 2.9 Risk Score with TIME characterization

To uncover the correlation between the risk score and tumor infiltrating immune cells, we implemented the seven methods including XCELL, TIMER, QUANTISEQ, MCPcounter, EPIC, CIBERSORT, and CIBERSORT-ABS to evaluate the immune infiltrating situation. Wilcoxon signed-rank test was employed to compare the distinction of immune infiltrating cell content between low- and high-risk groups. Spearman correlation was analyzed to explore the relevance between risk score and the immune infiltration statues. The R packages “scales”, “limma”, “ggtext”, “ggplot2”, and “ggpubr” were employed in above analysis.

## 2.10 Role of Risk Score in Immune Checkpoint Blockade Treatment

According to previous research, expression patterns of immune checkpoint blockade-related hub targets might contribute into efficacy of immunotherapy administration[24].

In this study, six hub genes of immunotherapy were fetched: programmed death ligand 1 (PD-L1, also known as CD274), programmed death 1 (PD-1, also known as PDCD1), programmed death ligand 2 (PD-L2, also known as PDCD1LG2), cytotoxic T-lymphocyte antigen 4 (CTLA-4), T-cell immunoglobulin domain and mucin domain-containing molecule-3 (TIM-3, also known as HAVCR2), and indoleamine 2,3-dioxygenase 1 (IDO1) in PDAC[25–27]. To further explore the potential role of risk signature in immunotherapy, correlation of prognostic signature with expression value of six ICB hub genes was analyzed. The R packages “corrplot”, “ggplot2”, “ggpubr”, and “ggExtra” were employed in above analysis.

## 2.11 Statistical analysis

Wilcoxon rank-sum test was a non-parametric statistical hypothesis test mainly used for comparisons between two groups and Kruskal-Wallis test was suitable for two or more categories. Overall survival (OS) refers to the interval from the date of diagnosis to the date of death. Survival curves were plotted via the Kaplan-Meier log rank test. CIBERSORT algorithm results with $p >= 0.05$ were rejected for further analysis. Univariate and multivariate analyses were performed via Cox regression models to validate the independent prognosis predictive performance of risk signature. The prognostic value for 1-, 2- and 3-year OS was assessed with the ROC curves. $p < 0.05$ deemed statistical significance. R software (version 4.0.2) was utilized for all statistical analyses.

# 3. Results

## 3.1 Identification of DEirlncRNAs

As descripted previously, a total of 13,162 lncRNAs were identified and annotated in TCGA-BRCA project. Furthermore, 2,483 immune-related genes were obtained from the ImmPort data portal. A co-expression network between lncRNAs and ir-gene was constructed to distinguish irlncRNAs. Finally, 946 irlncRNAs were determined ($|R| > 0.4$ and $P < 0.001$) after correlation analysis (Table S1). Taking advantage of differentially expressed analysis, 188 DEirlncRNAs (48 down-regulated and 140 up-regulated) were determined as described previously (Fig. 1A, Table S2). The distribution of top 50 DEirlncRNAs were visualized and presented in the heatmap (Fig. 1B).

## 3.2 Establishment of DEirlncRNA Pairs

With the help of an iteration loop and a 0-or-1 screening matrix, 11,933 DEirlncRNA pairs were determined. After univariate COX regression combined with Lasso algorithm analysis, 29 pairs of DEirlncRNAs were identified (Fig. 2A and 2B, Table S3). Next, a stepwise multivariate Cox regression analysis was performed, and 15 lncRNAs pairs finally identified as the predictors of OS in BRCA samples (Fig. 2C).

## 3.3 Development of Risk Model

Then, the value of AUC for each ROC curve was calculated and the result showed that the maximum AUC value was 0.817 to construct the most ideal risk model (Fig. 3A). Additionally, the 1-, 2-, and 3-year ROC curves were drawn and all AUC values were over 0.8 (Fig. 3B), suggesting validity of risk model. To demonstrate risk score was the best prognostic predictor, age, gender, clinical stage and TNM status were listed as the candidate indicators. These clinical variables were introduced into the AUC analysis for 1/2/3- OS and risk signature were observed to obtain the most AUC value (Figs. 3C, 3D and 3E). The maximum inflection point was recognized as the cut-off point according to the Akaike information criterion (AIC) values (Fig. 3F). BRCA samples were classified as low- or high-risk groups based on cut-off point.

## 3.4 Clinical Evaluation of Risk Model

In total, 203 samples were recognized as the high-risk group and 866 as the low-risk group. Figures 4A, 4B respectively displayed the distributions of risk score and survival time of each patient, suggesting that samples with low-risk group exhibited prognosis advantage compared with high-risk patients. Kaplan-Meier survival curve showed that clinical outcome of low-risk samples was superior to that of high-risk group ($P < 0.001$; Fig. 4C). Besides, the hazard ratio (HR) for risk score in univariate Cox proportional hazards regression was 1.341 ($p < 0.001$, 95% CI [1.280–1.406]; Fig. 4D). Consistent results were observed on multivariate Cox proportional hazards regression ($p < 0.001$, HR = 1.321, 95% CI: 1.252 – 1.393; Fig. 4E), indicating risk score could serve as independent prognostic indicator.

# 3.5 Clinical significance of risk score

Firstly, the distribution of clinicopathological features subtypes in different risk groups was explored and visualized (Fig. 5A). According to the result of a series of chi-square tests, with increased age ($P = 0.037$, Fig. 5B), advanced clinical stage (4 out of 6 $P < 0.05$, Fig. 5C), higher T status (3 out of 6 $P < 0.05$, Fig. 5D), higher N status (4 out of 6 $P < 0.05$, Fig. 5E), and higher M status ($P = 0.0064$, Fig. 5F), risk score significantly escalated.

Furthermore, to confirm whether prognostic signature remained robust prognosis prediction validity in patients subdivided into different subtypes according to clinicopathological variables, stratification analysis was performed. Compared with low-risk samples, BRCA patients with high-risk had lower overall survival rate in both the young ( < = 65) and old (> 65) groups (Figures S1A and S1B). Likewise, risk score suggested prognostic difference well for samples in female gender or male gender (Figures S1C and S1D), samples with early- or late-stage (Figures S1E and S1F), samples in T1-2 or T3-4 category (Figures S1G and S1H), samples in N0 or N1 status (Figures S1I and S1J) and samples in M0 or M1 category (Figures S1K and S1L). These results demonstrated that risk score was an outstanding prognostic predictor which independent of clinical variables.

# 3.6 Development of Prognostic Nomogram

To develop quantitative risk manner for clinical practice, prognostic nomogram integrating risk score, age and clinical stage was delineated to predict overall survival rate (Fig. 6A). Gender was excluded out of the nomogram given its AUC value was less than 0.6. Calibrate curves were approximately diagonal, supporting great prognostic predictive validity of overall survival rate in risk-clinical nomogram (Figs. 6B-D).

# 3.7 Role of Risk Score in context of TIME

Given risk score was derived from irlncRNAs, the potential contribution of as-constructed signature in diversity and complexity of TIME were further explored. The results showed that high risk score was significantly and negatively correlated with abundance of memory B cell, M1 Macrophages, resting Myeloid dendritic cells, activated NK cell, CD8 + T cells, follicular helper T cells, gamma delta T cells and regulatory T cells, whereas positively related with infiltration of Cancer associated fibroblast, Endothelial cell, M2 Macrophages, resting Mast cells, resting NK cell, and Neutrophil (Figures S2-S4). Furthermore,

Spearman correlation analysis was further performed (Fig. 7A, Table S4) and the detailed results were provided in Table S5. These findings suggested that low-risk group characteristic with immune response activated condition, which may contribute to anti-tumor effect.

Besides, we attempted to elucidate the potential function of risk score in immunotherapy. First, the correlation of immunotherapy key targets (PDCD1, CD274, PDCD1LG2, CTLA-4, HAVCR2, and IDO1) [25–27] with risk score was performed (Fig. 7B). And we observed that risk score was positively and significantly correlated with CTLA4 (r= -0.085; P = 0.0054; Figs. 7C) and PDCD1 (r= -0.14; P = 5.1e-06; Figs. 7D), indicating risk score might serve as a pivotal player in the prediction of clinical outcome of immunotherapy.

## 3.8 Correlation of Risk Signature with TMB

Existing studies have contributed strong evidence to demonstrate that high tumor mutation burden (TMB) was correlated with increasement of infiltrating CD8 + T cells, which recognized tumor neoantigens then resulted in intense tumor-killing effects to annihilate tumor cells (18–20).Thus, we speculated that TMB might act as a nonnegligible prognostic factor of responsiveness to antitumor immunotherapy and aimed to investigate the potential interaction between risk score and TMB to uncover the hereditary variations of risk score subtype. Firstly, the patients were assigned into distinct subtypes based on the TMB immune set point. Survival curve demonstrated that high TMB value significantly suggested shorter overall survival time (p = 0.001, Fig. 8A). Subsequent correlation analysis further validated that the TMB was positively and significantly related with the risk score (R = 0.15, p = 2.5e-06; Fig. 8B). To further explore the validity of consistent prognostic significance of risk score and TMB, we validated the cooperative effect of two indicators in prognostic prediction As demonstrated in stratified survival curve, there was no interference of TMB status with risk score prognostic predictive performance. Risk score subgroups exhibited evident prognosis distinctions in both low and high TMB status subtypes (p < 0.001; Fig. 8C). In summary, these results suggested that risk score might act as independent prognostic predictor and hold the potential to evaluate the clinical outcome of antitumor immunological treatment.

Besides, we explored and visualized the distribution of gene mutation in both the high-and low-risk score subtypes. The comprehensive landscape of somatic variants visualized the mutation patterns and clinical features of the top 20 driver genes with the most frequent alteration (Figs. 8D and 8E). These findings might contribute novel insight into the intrinsic connection of irlncRNAs and somatic variants in immunotherapy.

# 4. Discussion

Globally, breast cancer is a primary cause of death due to cancer in women and is the most frequently diagnosed form of cancer in a large number of countries(1). Despite striking progress have been made in the early diagnosis, therapeutic process monitoring, and prognostic evaluation, clinical outcome of BRCA still remains poor. The high heterogeneity of breast cancer exists not only in the genotypes and phenotypes of tumor cells but also in the tumor immune microenvironment, resulting in patients with the

same clinical stage presented distinct responses to treatments and clinical outcomes. Traditionally clinical and pathological classification methods, such as tumor size, regional lymph node metastasis status, and distant metastasis, are too general to accurately determinate the prognosis of an individual patient(21). Thus, the novel prognostic biomarkers based on risk stratification may be the most effective strategy for precise prognostic prediction, contributing into optimal tailored treatment. In the past decades, accumulating studies suggested that lncRNAs participate in tumorigenesis, progression, metastasis and the prognosis of breast cancer via variety of ways(22, 23). Also, lncRNAs were reported as key regulators in regulating cancer immunity(11). Emerging evidence has supported that immune-related lncRNAs may serve as new therapeutic targets and disease molecular biomarkers for cancer clinic management and possess predictive value for survival prognosis(24–26).

However, most of previous researches aimed at prognostic signatures based on noncoding RNAs, focusing on prognostic prediction of cancer patients, are constructed by adopting the exact expression levels(27–29). This research was inspired by gene pairing and designed to develop a novel and reliable risk signature based on combinations of two-lncRNAs, without requirement of specific expression levels.

In our study, raw data of lncRNAs from TCGA-BRCA project was introduced into differential co-expression analysis to determine differential expressed immune-related lncRNAs (DEirlncRNAs), and lncRNA-pairs were validated by improved algorithm of cyclically single pairing along with a 0-or-1 matrix. Then, univariate and multivariate regression analysis followed by Lasso penalized regression to identify DEirlncRNAs pairs. Next, each AUC value of ROC curve was counted to obtain the ideal risk signature and the AIC value of each point on the AUC curve was calculated to determine the best cut-off point to stratify BRCA patients into the low- or high-risk-group. Additionally, prognostic value of risk model was validated using survival analysis, ROC curve, and univariable and multivariable regression analysis. Additionally, prognostic nomogram was constructed and confirmed to facilitate clinical extension. Furthermore, the potential role of risk score in TIME characterization and immunotherapy was investigated. Finally, the synergistic effect of risk score with TMB in term of prognostic prediction was demonstrated.

It was well established that lncRNAs with high abundance functioned as pivotal players in tumorigenesis, progression, and prognosis of cancer(30). This algorithm was designed to determination of DEirlncRNAs followed by establishment the most significant pairing of irlncRNA. As such, irlncRNA pairs with lower or higher expression value rather than each lncRNA expression level had to be examined. It was worthwhile to mentioned that as-constructed risk model harbored significant superiority of analysis cost relative to prognostic signature dependent on specific expression value of genes.

To enhance the efficacy and accuracy of prognostic prediction, the AIC values were employed to obtain the best cut-off point for risk stratification instead of simple median value. Additionally, AUC value of 1-/2-/3-year OS ROC curve was calculated then compared with various clinical variables to exhibit prognostic advantage of risk model. The excellent prognostic performance of risk model was validated by K-M analysis. Furthermore, risk signature was demonstrated to perform well as an independent prognostic predictor in both univariable and multivariable regression analysis. Besides, risk signature

remained powerful prognostic ability in clinical variables stratified survival curves. Finally, risk-clinical nomogram that integrated risk score, age and clinical stage was established for clinical transformation.

Since risk model was established on irlncRNAs, this risk model was potentially mediated in modeling of TIME or suppression of immune-relevant cells. The results of TIME context indicated that risk score was negatively related with activated immune cell (i.e., M1 Macrophages, activated NK cell, CD8 + T cells, etc.,), whereas positively correlated with immunosuppressive cells (i.e., M2 Macrophages, etc.,), implying subjects with high risk was well characterized as immune suppressive phenotype, which was coincident with lower risk score suggested longer overall survival time.

Immunotherapy was treatment employing immune system fight against cancer cells, thus, infiltrating immune cells could affect clinical outcome of immune checkpoint blockades administration. The results showed that risk score was significantly and negatively correlated with the immunotherapy hub targets (i.e., PDCD1, etc.,), suggesting samples with low-risk score might be more affected by immune checkpoint blockade pathways, then inhibited anti-tumor immune activation and deteriorate prognosis accordingly. Since no immunotherapy data in BRCA cohort, it was unable to further explore the correlation of risk score with response of immunotherapy.

Currently, several clinical data pointed out a correlation between genetic alternations with responsiveness to immunological treatment (31, 32). We calculated and determined the TMB, which is a predictive indicator of sensitivity to immunological treatment, increased significantly with risk score elevated. Subsequent stratified survival curve demonstrated that risks score held prognostic predictive capability which was independent of TMB, suggesting that TMB and risk score represent different aspects of immunobiology. Besides, risk score together with mutation data revealed the significant distinction of genes variant frequency between high and low risk score group from the level of transcriptome.

# 5. Conclusions

In summary, the landscape of irlncRNAs was systematically delineated by employing co-expression network and differential expressed analysis. Notably, novel and robust prognostic signature without specific lncRNAs expression values was constructed and validated to contribute into clinical outcome stratification and TIME heterogeneity, presenting valuable clinical applications and prognostic target in BRCA antitumor immunotherapy. Furthermore, the synergistic effect of risk score and TMB value in terms of prognostic prediction was proposed. Even though, further experimental and clinical validation were required for these findings at different centers and larger cohort.

# Abbreviations

AIC: Akaike information criterion

AUC: area under the curve

BRCA: breast cancer

CTLA-4: cytotoxic T-lymphocyte antigen 4

CI: confidence interval

CD274: Also known as PD-L1

DEirlncRNAs: differential expressed irlncRNAs

FDR: false discovery rate

HR: hazard ratio

HAVCR2: Also known as TIM3

IDO1: indoleamine 2,3-dioxygenase 1

ICB: immune checkpoint blockade

irlncRNAs: immune-related lncRNAs

K–M: Kaplan-Meier

LASSO: least absolute shrinkage and selection operator

lncRNAs: long non-coding ribonucleic acids

MAF: Mutation Annotation Format

OS: overall survival

PD-1: Programmed Cell Death 1

PD-L1: Programmed Cell Death-Ligand 1

PD-L2: Programmed Cell Death-Ligand 2

PDCD1: Also known as PD-1

PDCD1LG2: Also known as PD-L2

RNA: Ribonucleic Acid

ROC: receiver operating characteristic

TCGA: The Cancer Genome Atlas

TIICs: tumor-infiltrating immune cells

TIME: tumor immune microenvironment

TIMER: tumor immune estimation resource

TIM-3: T-cell immunoglobulin domain and mucin domain-containing molecule-3

TMB: tumor mutation board

TNM: Tumor Node Metastasis

## Declarations

# Ethics approval and consent to participate

Not applicable

# Consent for publication

Not applicable.

# Availability of Data and Materials

The datasets generated for this study can be found in the TCGA database (https://portal.gdc.cancer.gov). The original data was too large to upload in the system, so they were uploaded in the JianguoYun. The download link was as follow: https://www.jianguoyun.com/p/DQp56owQ4L_ICRj7h_8D.

# Competing interests

The authors declare that they have no competing interests.

# Funding

This study was supported by Funding of Wenzhou Municipal Science and Technology Bureau (Grant No.Y2020971). The funder had no role in the data collection, analysis, or interpretation of results.

# Authors' contributions

HW designed the overall study and revised the paper, XQH performed public data interpretation and drafted manuscript. XQH and HYB participated in data collection, XQH and CSH contributed to data analysis. All authors read and approved the final manuscript.

*Qianhui Xu, Shaohuai Chen, and Yuanbo Hu contributed equally to this paper.

# Acknowledgements

The authors would like to give our sincere appreciation to the reviewers for their helpful comments on this article and research groups for the TCGA, which provided data for this collection.

# References

1. Bray F, Ferlay J, Soerjomataram I, Siegel R, Torre L, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018;68(6):394-424.
2. Saad ED, Katz A, Buyse M. Overall survival and post-progression survival in advanced breast cancer: a review of recent randomized clinical trials. J Clin Oncol. 2010;28(11):1958-62.
3. Karaman S, Detmar M. Mechanisms of lymphatic metastasis. J Clin Invest. 2014;124(3):922-8.
4. Koren S, Bentires-Alj M. Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. Molecular cell. 2015;60(4):537-46.
5. Couzin-Frankel J. Breakthrough of the year 2013. Cancer immunotherapy. Science. 2013;342(6165):1432-3.
6. Baxevanis C, Fortis S, Perez S. The balance between breast cancer and the immune system: Challenges for prognosis and clinical benefit from immunotherapies. Seminars in cancer biology. 2019.
7. Byrne A, Savas P, Sant S, Li R, Virassamy B, Luen S, et al. Tissue-resident memory T cells in breast cancer control and immunotherapy responses. Nature reviews Clinical oncology. 2020;17(6):341-8.
8. Harao M, Forget M, Roszik J, Gao H, Babiera G, Krishnamurthy S, et al. 4-1BB-Enhanced Expansion of CD8 TIL from Triple-Negative Breast Cancer Unveils Mutation-Specific CD8 T Cells. Cancer immunology research. 2017;5(6):439-45.
9. Prensner J, Chinnaiyan A. The emergence of lncRNAs in cancer biology. Cancer discovery. 2011;1(5):391-407.
10. Carpenter S, Fitzgerald K. Cytokines and Long Noncoding RNAs. Cold Spring Harbor perspectives in biology. 2018;10(6).
11. Denaro N, Merlano M, Lo Nigro C. Long noncoding RNAs as regulators of cancer immunity. Molecular oncology. 2019;13(1):61-73.

12. Mineo M, Lyons S, Zdioruk M, von Spreckelsen N, Ferrer-Luna R, Ito H, et al. Tumor Interferon Signaling Is Regulated by a lncRNA INCR1 Transcribed from the PD-L1 Locus. Molecular cell. 2020;78(6):1207-23.e8.
13. Pei X, Wang X, Li H. LncRNA SNHG1 regulates the differentiation of Treg cells and affects the immune escape of breast cancer via regulating miR-448/IDO. International journal of biological macromolecules. 2018;118:24-30.
14. Liu GM, Zeng HD, Zhang CY, Xu JW. Identification of a six-gene signature predicting overall survival for hepatocellular carcinoma. Cancer Cell Int. 2019;19:138.
15. Speers C, Chang SL, Pesch A, Ritter C, Olsen E, Chandler B, et al. A Signature That May Be Predictive of Early Versus Late Recurrence After Radiation Treatment for Breast Cancer That May Inform the Biology of Early, Aggressive Recurrences. Int J Radiat Oncol Biol Phys. 2020;108(3):686-96.
16. Lv Y, Lin S, Hu F, Ye Z, Zhang Q, Wang Y, et al. Landscape of cancer diagnostic biomarkers from specifically expressed genes. Briefings in bioinformatics. 2020;21(6):2175-84.
17. Mayakonda A, Lin D, Assenov Y, Plass C, Koeffler H. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome research. 2018;28(11):1747-56.
18. Chan T, Yarchoan M, Jaffee E, Swanton C, Quezada S, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. Annals of oncology : official journal of the European Society for Medical Oncology. 2019;30(1):44-56.
19. Rizvi N, Hellmann M, Snyder A, Kvistborg P, Makarov V, Havel J, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science (New York, NY). 2015;348(6230):124-8.
20. McGranahan N, Furness A, Rosenthal R, Ramskov S, Lyngaa R, Saini S, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science (New York, NY). 2016;351(6280):1463-9.
21. Sui Y, Ju C, Shao B. A lymph node metastasis-related protein-coding genes combining with long noncoding RNA signature for breast cancer survival prediction. Journal of cellular physiology. 2019;234(11):20036-45.
22. Huang Q, Liu G, Qian X, Tang L, Huang Q, Xiong L. Long Non-Coding RNA: Dual Effects on Breast Cancer Metastasis and Clinical Applications. Cancers. 2019;11(11).
23. Kim J, Piao H, Kim B, Yao F, Han Z, Wang Y, et al. Long noncoding RNA MALAT1 suppresses breast cancer metastasis. Nature genetics. 2018;50(12):1705-15.
24. Xu J, Meng Q, Li X, Yang H, Xu J, Gao N, et al. Long Noncoding RNA MIR17HG Promotes Colorectal Cancer Progression via miR-17-5p. Cancer research. 2019;79(19):4882-95.
25. Zhou Y, Zhu Y, Xie Y, Ma X. The Role of Long Non-coding RNAs in Immunotherapy Resistance. Frontiers in oncology. 2019;9:1292.
26. Wei C, Liang Q, Li X, Li H, Liu Y, Huang X, et al. Bioinformatics profiling utilized a nine immune-related long noncoding RNA signature as a prognostic target for pancreatic cancer. J Cell Biochem. 2019;120(9):14916-27.

27. Xu Q, Wang Y, Huang W. Identification of immune-related lncRNA signature for predicting immune checkpoint blockade and prognosis in hepatocellular carcinoma. Int Immunopharmacol. 2021;92:107333.
28. zhang L, Li P, Liu E, Xing C, Zhu D, Zhang J, et al. Prognostic value of a five-lncRNA signature in esophageal squamous cell carcinoma. Cancer Cell International. 2020;20(1):386.
29. Tian T, Gong Z, Wang M, Hao R, Lin S, Liu K, et al. Identification of long non-coding RNA signatures in triple-negative breast cancer. Cancer Cell International. 2018;18(1):103.
30. Huarte M. The emerging role of lncRNAs in cancer. Nat Med. 2015;21(11):1253-61.
31. George S, Miao D, Demetri G, Adeegbe D, Rodig S, Shukla S, et al. Loss of PTEN Is Associated with Resistance to Anti-PD-1 Checkpoint Blockade Therapy in Metastatic Uterine Leiomyosarcoma. Immunity. 2017;46(2):197-204.
32. Burr M, Sparbier C, Chan Y, Williamson J, Woods K, Beavis P, et al. CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. Nature. 2017;549(7670):101-5.

# Figures

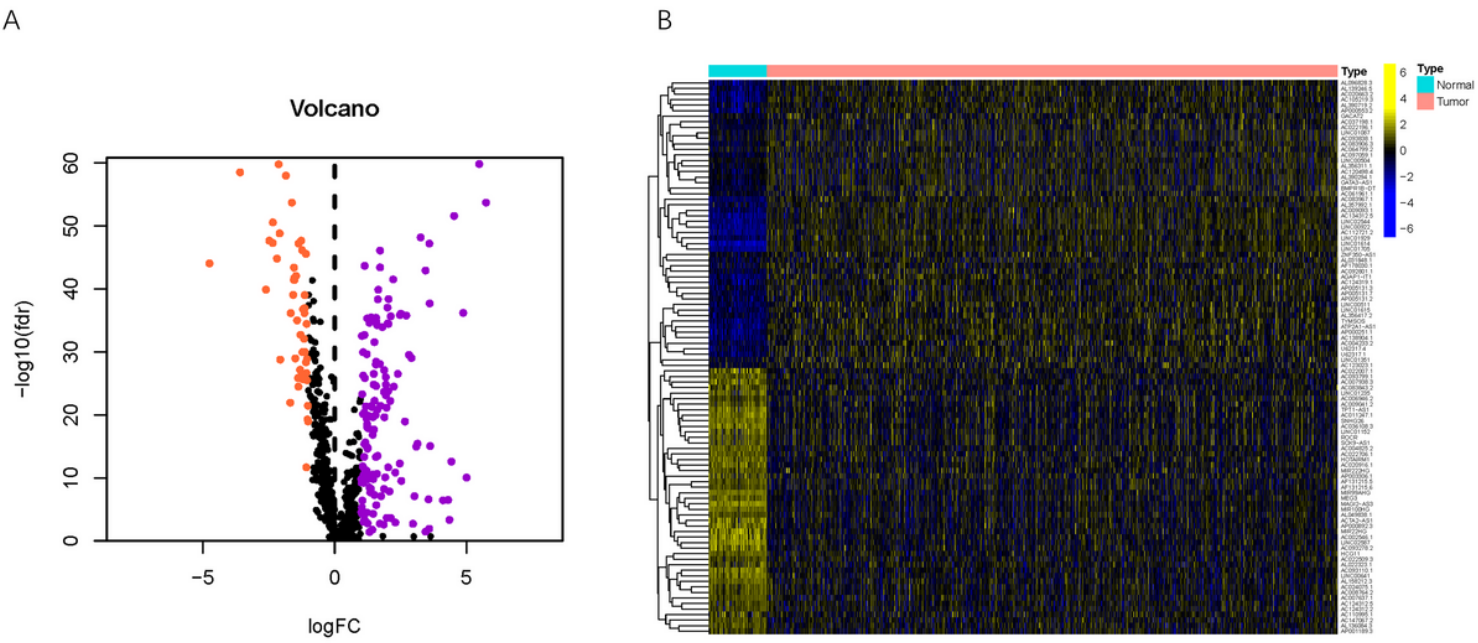
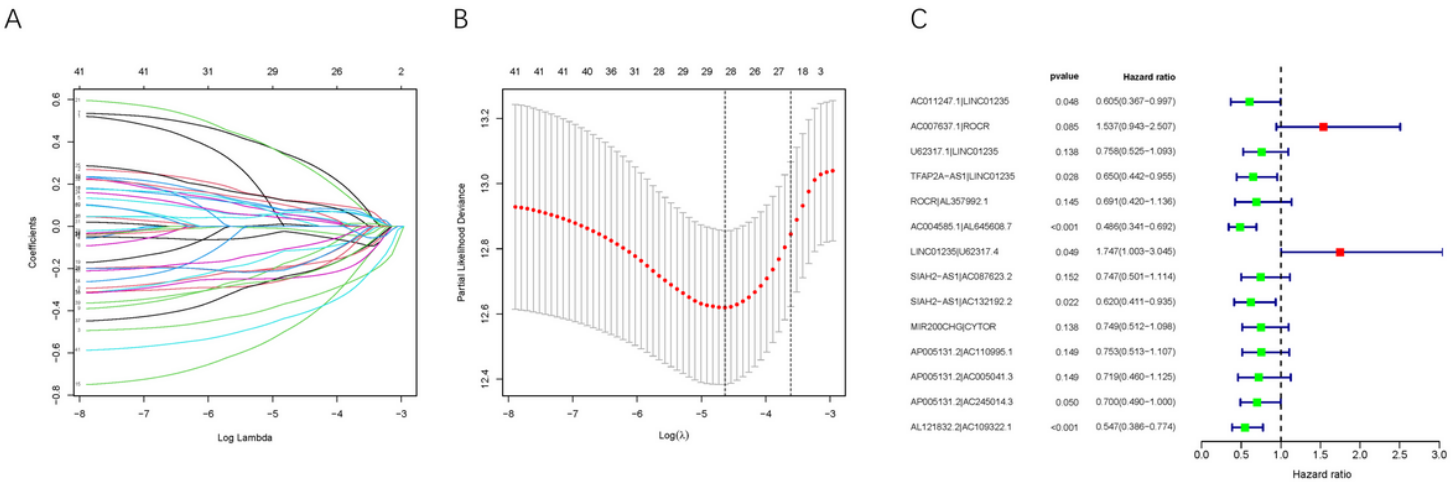



**Figure 1**

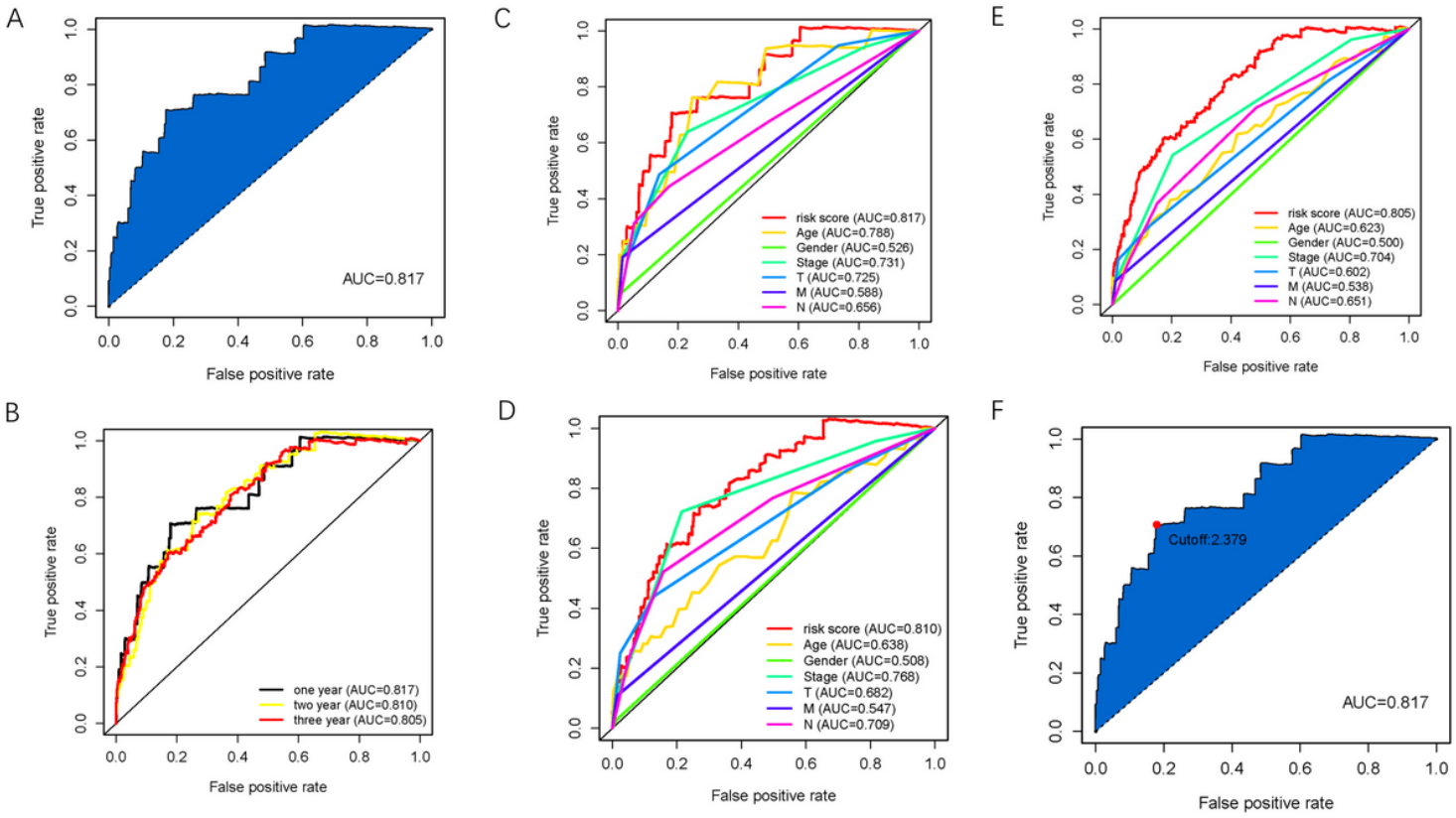
Differential analysis of irlncRNAs expression data in tumor samples and normal tissues. (A) Volcano plot was delineated to visualize the DEirlncRNAs. Purple represented upregulated and orange represented downregulated. (B) Heatmap of top 50 DEirlncRNAs was plotted to reveal different distribution of

expression state, where the colors of yellow to blue represented alterations from high expression to low expression.



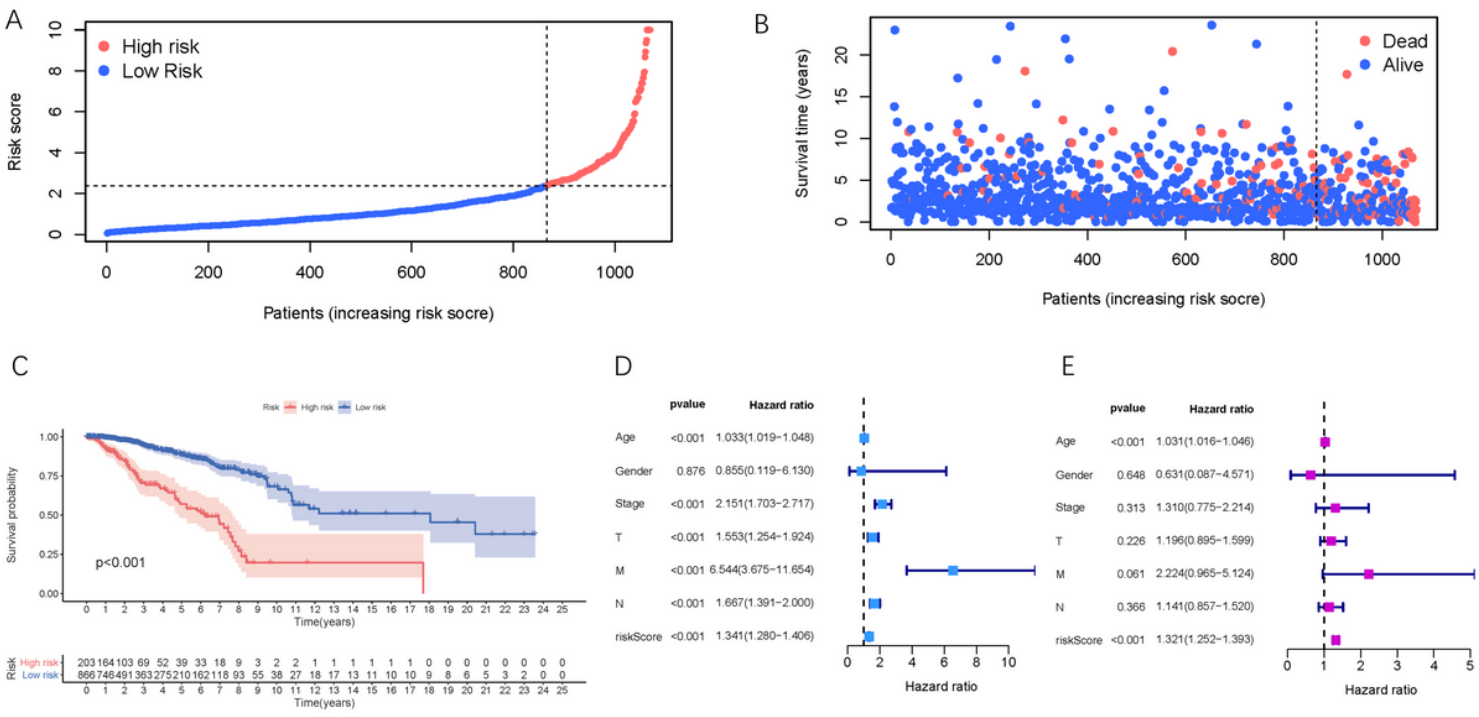

## Figure 2

Establishment of the Prognostic Risk Signature. (A) LASSO coefficient profiles of 41 candidate DEirlncRNAs pairs. A vertical line is drawn at the value chosen by 10-fold cross-validation. (B) Ten-time cross-validation for tuning parameter selection in the lasso regression. The vertical lines are plotted based on the optimal data according to the minimum criteria and 1-standard error criterion. The left vertical line represents the 29 DEirlncRNAs pairs finally identified. (C) A forest map showed 15 DEirlncRNAs pairs identified by Cox proportional hazard regression in the stepwise method.



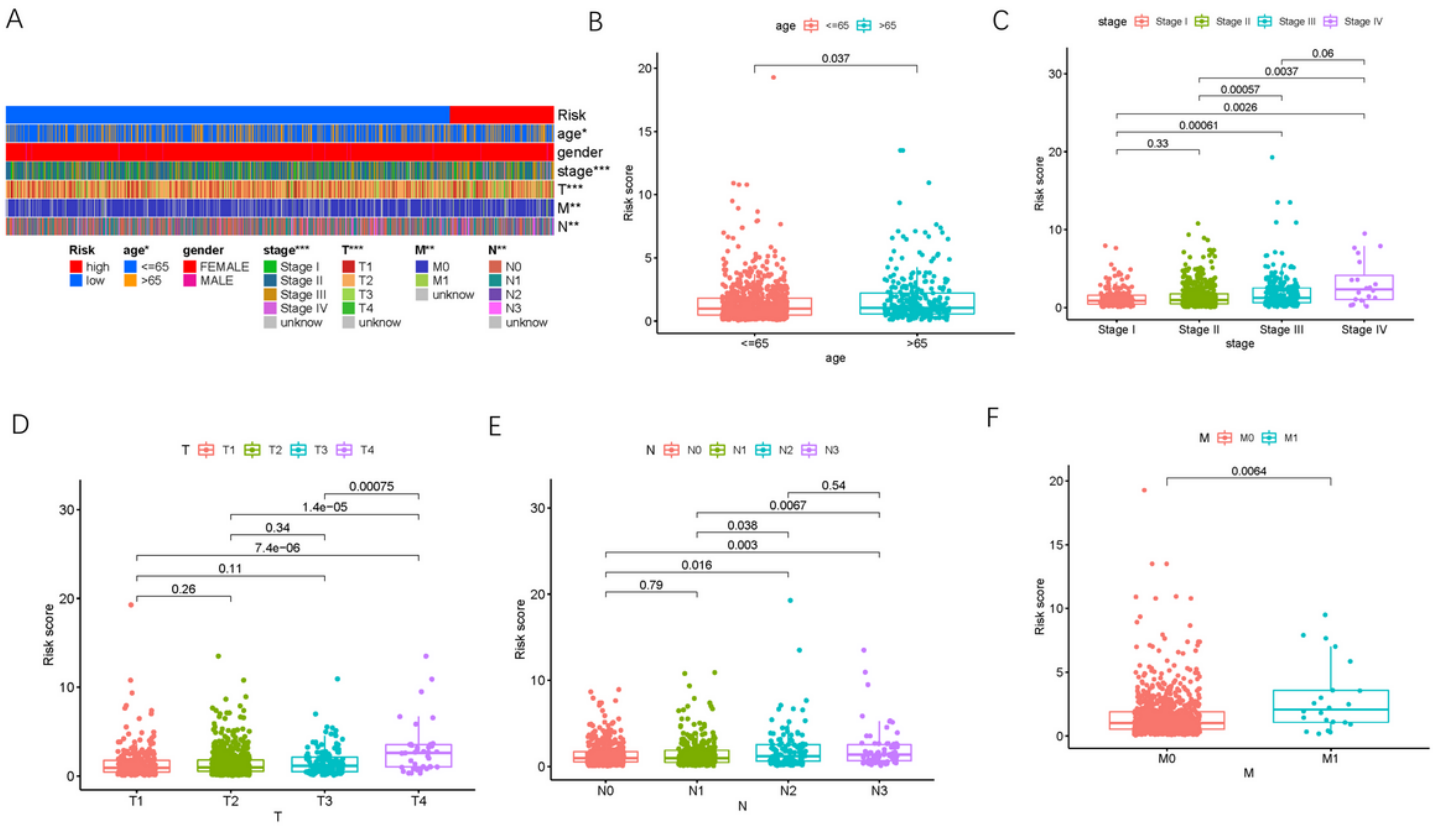

## Figure 3

Identification of Cut-off Point for Risk Signature (A) The ROC of the optimal DEirlncRNA pair models was related to the maximum AUC. (B) The 1/2/3-year ROC of the optimal model. A comparison of 1-year (C), 2-year (D), 3-year (E) ROC curves with other common clinical characteristics. (F) RiskScore for BRCA patients; the maximum inflection point is the cut-off point obtained by the AIC.



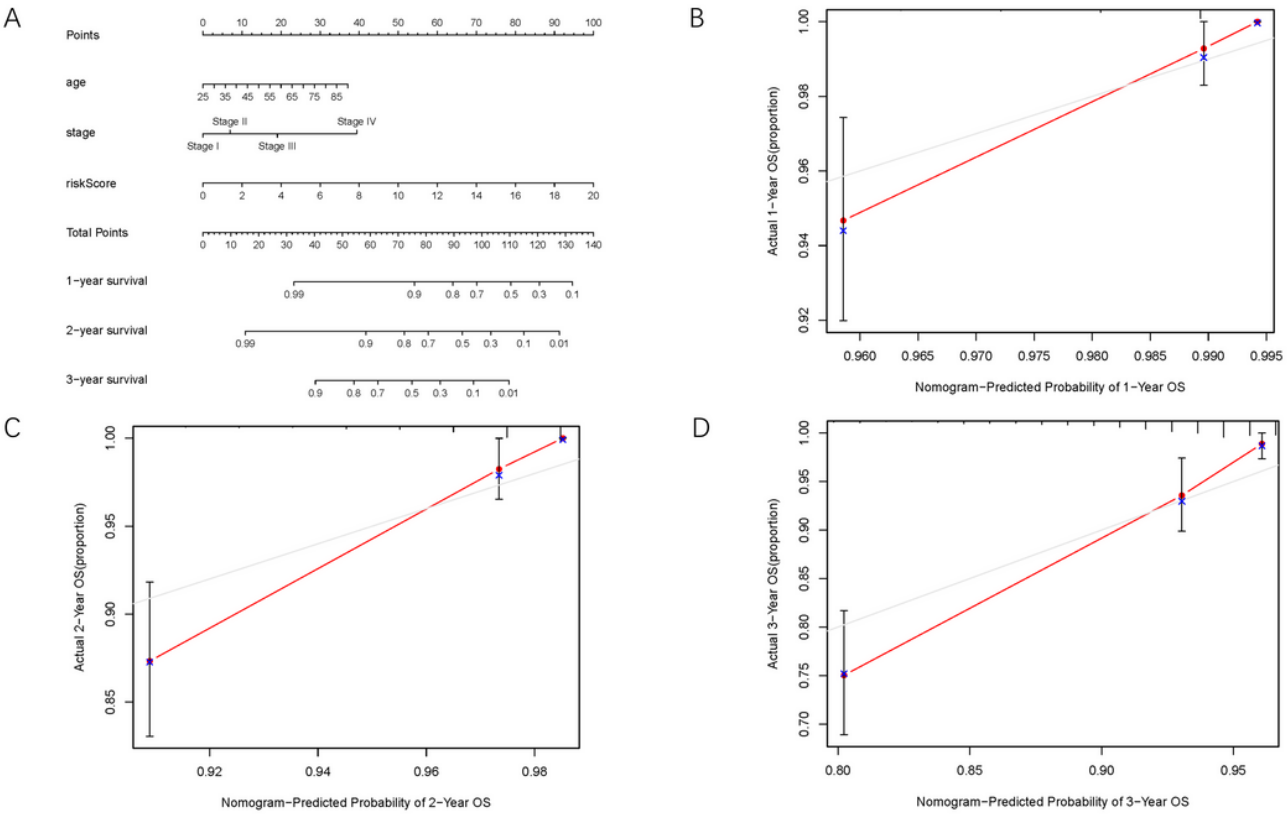

## Figure 4

Validation of the Prognostic Risk Signature (A) Distribution of multi-genes model risk score. (B) The survival status and duration of HCC patients. (C) Kaplan–Meier curve analysis presenting difference of overall survival between the high-risk and low-risk groups. (D) Univariate Cox regression results of overall survival. (E) Multivariate Cox regression results of overall survival.



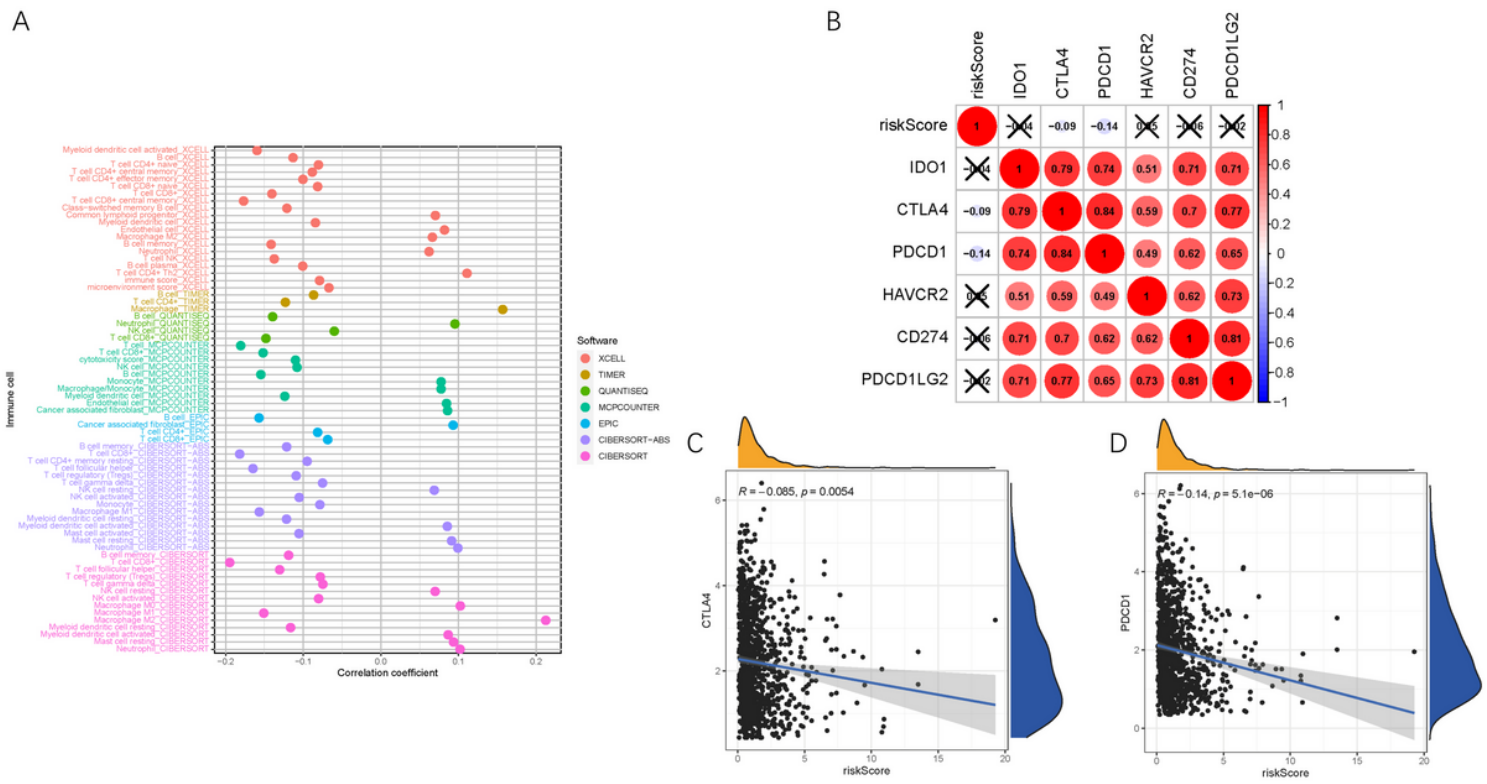

## Figure 5

Clinical significance of the prognostic risk signature. (A) Heatmap presents the distribution of clinical feature and corresponding risk score in each sample. Comparison of risk score among samples from clinical variables subgroups. (B) Age, (C) Clinical stage, (D) T status, (E) N status and (F) M status.




## Figure 6

Establishment of Prognostic Risk-Clinical Nomogram. (A) Nomogram was assembled by age, clinical stage and risk score for predicting overall survival of BRCA patients. (B) One-year nomogram calibration curves. (C) Two-year nomogram calibration curves. (D) Three-year nomogram calibration curves.




## Figure 7

Estimation of Tumor-Infiltrating Cells and Immunotherapy significance. (A) Subjects in the high-risk group were more positively associated with tumor-infiltrating immune cells, as shown by Spearman correlation analysis. Correlation between prognostic risk signature with immune checkpoint hub genes. (B) Correlation analysis between immune checkpoint inhibitors (CD274, PDCD1, PDCD1LG2, CTLA4, HAVCR2, and IDO1) with prognostic risk signature. (C) Correlation between prognostic risk signature and CTLA4. (D) Correlation between prognostic risk signature and PDCD1.

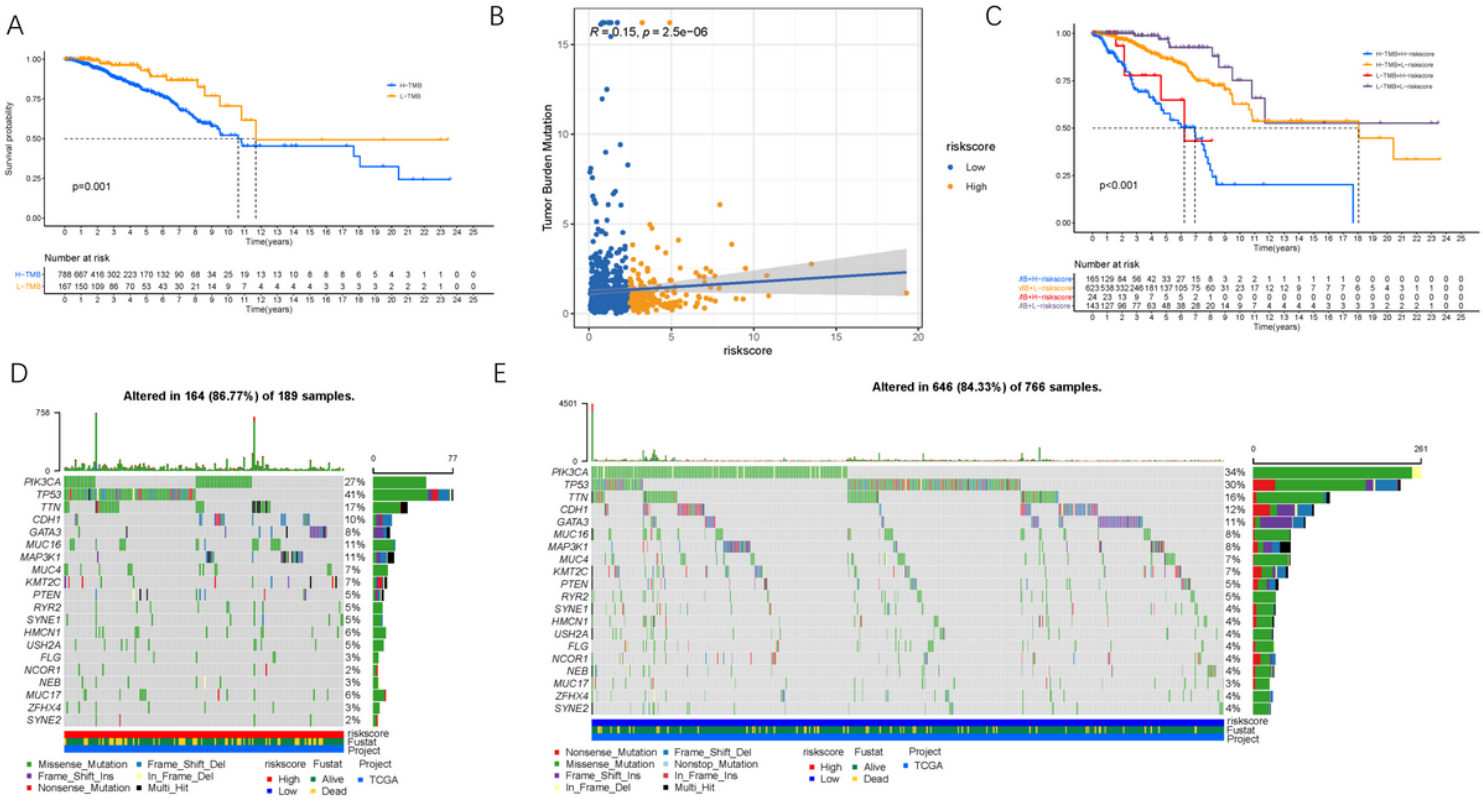



**Figure 8**

The Correlation between the risk Score and TMB (A) Kaplan-Meier curves for high and low TMB groups. (B) Scatterplots depicting the positive correlation between risk scores and TMB. (C) Kaplan-Meier curves for patients stratified by both TMB and risk score. The oncoPrint was constructed using high risk score (D) and low risk score (E).

# Supplementary Files

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

- SupplementaryFigures.pdf
- SupplementaryTables.xlsx