

# Tumor microenvironmental prognosis analysis and molecular labeling of immune-related genes in Luminal B type breast cancer

**Fengjiao Gan**

Breast and Thyroid Disease Medical Center, Affiliated Hospital of Zunyi Medical University

**Kang Hu**

Breast and Thyroid Disease Medical Center, Affiliated Hospital of Zunyi Medical University

**Xiaomei Li**

Department of Cancer Research Laboratory, the Affiliated Hospital of Zunyi Medical University

**Ni Jiang**

Department of cancer research laboratory, the affiliated hospital of zunyi medical university

**Qiaoyuan Wu**

Department of cancer research laboratory, the affiliated hospital of zunyi medical university

**Yi Li**

Department of cancer research laboratory, the affiliated hospital of zunyi medical university

**Mengxi Xu**

Breast and thyroid disease medical center, affiliated hospital of zunyi medical university

**Suhong Sun**

Breast and thyroid disease medical center, affiliated hospital of zunyi medical university

**Qing Luo** (✉ [zlsysluoqing@163.com](mailto:zlsysluoqing@163.com))

Department of Cancer Research Laboratory, the Affiliated Hospital of Zunyi Medical University

---

## Research

**Keywords:** Tumor microenvironment, luminal B breast cancer, TME score, bioinformatics, TCGA

**Posted Date:** December 18th, 2019

**DOI:** <https://doi.org/10.21203/rs.2.19160/v1>

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

[Read Full License](#)

---

# Abstract

**Background:** Tumor microenvironment (TME) cells is one of the important elements, constitute the tumor tissues and in predicting tumor clinical results and treatment effect has the important significance of clinical pathology. However, a detailed map for prognostic landscape of TME in Luminal B breast cancer is still lacking.

**Methods:** downloaded the expression profile and clinical follow-up information of luminal B breast cancer from TCGA and GEO, and used CIBERSORT to evaluate the infiltration mode of TME in 209 patients, and constructed the molecular subtype of luminal B breast cancer based on TME, further evaluated the relationship between molecular subtype and prognosis. DESeq2 was used to analyze the differentially expressed genes among molecular subtypes of luminal B breast cancer, and cox multivariate regression analysis was used for feature selection to construct TME signature.

**Results:** used on the median value of TMEscore, the samples were divided into High TMEscore and Low TMEscore, and the relationship with prognosis, clinical features, immune gene expression and genomic variation was further evaluated. Different TME cells have significant correlation in luminal B breast cancer. These TME cells stratified different clinical results of luminal B breast cancer, and further TMEscore was established by Cox regression analysis. High TME score is associated with poor prognosis. Meanwhile, this study showed that CXCL10, CXCL9, GZMB and other immune activation genes and PDCD1LG2 and other immune checkpoint genes have higher expression levels in high TME score, and the mutation frequency of TP53 gene was lower in risk-H than that in risk-L.

**Conclusion:** In this study, we systematically evaluated the TME infiltration patterns of 209 TCGA luminal B breast cancer patients and developed the TME score approach. It was found that TME score is a powerful prognostic biomarker and predictor of response to immunotherapy and provides a new strategy for luminal B breast cancer immunotherapy.

## Introduction

Breast cancer is one of the most common malignant tumors in the female population, and its incidence has leapt to the top of female malignant tumors. The latest global cancer data estimate that in 2018, there were 8.6 million new cancer cases and 4.2 million deaths among women worldwide, including 2.1 million new cases of breast cancer, accounting for 24.2% of the total, and 630,000 deaths, accounting for 15% of the total, both ranked first (1). The continuous advancement of gene detection technology has promoted the research on the molecular level of breast cancer. In 2000, Perou et al. first proposed the concept of molecular typing (2), Molecular typing is determined by the intrinsic gene expression of breast cancer (2, 3). Currently, breast cancer is roughly divided into four subtypes, namely luminal type A, luminal type B, HER-2 overexpression type and basal type (4, 5). Although most women with luminal B-like breast cancer use endocrine therapy alone, some of them have fatal recurrences, so understanding

the development of luminal B-like breast cancer and developing new prognostic factors is an urgent problem to be solved.

Different tumor microenvironments (TMEs) are formed in different stages of tumor development, and have multiple capabilities to induce adverse and beneficial consequences of tumorigenesis, immune cells can be activated to promote tumor growth and progression, most likely to be affected by the tumor microenvironment (6). More and more research shows that TME plays a vital role in cancer progression and therapeutic response (7, 8), For example, differences in the compositions of resident cell types within the TME, including cytotoxic T cells, helper T cells, dendritic cells (DCs), tumor-associated macrophages, mesenchymal stem cells, and associated inflammatory pathways, have been reported in patients with cancer (9–12). Detection of TME at the time of diagnosis can reflect immune response and chemotherapy benefits (13), and changes in the number of infiltrating CD8 + T cells, CD4 + T cells, macrophages, and fibroblasts in TME are associated with clinical and prognosis of multiple malignancies (12, 14–16). There are several algorithms that can estimate the abundance of immune cells and other cells in TME (17–19). Although some studies have used these methods to explore the relationship between TME infiltration and clinical outcomes of tumors (20, 21), a comprehensive landscape of TME infiltration has not been explored in luminal B breast cancer.

In this study, based on the clinical sample annotation of luminal B breast cancer gene expression profile, CIBERSORT algorithm was used to estimate 22 types of immune cells and the proportion of cancer associated fibroblasts, 209 TME infiltration mode of luminal B breast cancer patients were assessed, and systematically the TME phenotype and luminal B sample characteristics and clinical pathologic features associated with breast cancer genome. Therefore, we established a method to quantify the penetration pattern of TME (TME score). TME score was found to be a powerful prognostic biomarker and predictor of response to immunotherapy inhibitors.

## Materials And Methods

### Data collection and processing

We download expression profile data of TCGA Affymetrix HT Human Genome U133a microarray platform from GDC API database and the prognostic information on August 14, 2019. The samples without clinical data and follow-up time less than 30 days were removed, and 209 luminal B breast cancer samples were included in the final study as training set. Similarly, we downloaded two sets of chip data sets GSE1456 and GSE21653 from the Gene Expression Omnibus (GEO) database(22), which integrates the expression data and clinical follow-up information of 65 luminal B breast cancer samples from Affymetrix Human Genome U133A Array, Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Human Genome U95 Version 2 Array, with a follow-up time of more than 30 days. The sample informations are shown in Table 1. For the probe data, the R package hgu133plus2.db was used to map the probe to the GeneSymbol, and the median was obtained when the multiple probes correspond to the expression of one gene, and the probes that match multiple genes were removed.

Table 1  
Clinical information of the three sets of data sets after pretreatment.

Characteristic		TCGA datasets	GSE1456(n = 23)	GSE21653(n = 42)
Age(years)	<=60	109	-	28
	> 60	89	-	14
Survival Status	Living	165	16	23
	Dead	33	7	19
pathologic_T	T 1	37	-	-
	T 2	127	-	-
	T 3	24	-	-
	T 4	10	-	-
pathologic_N	N 0	77	-	-
	N 1	74	-	-
	N 2	33	-	-
	N 3	10	-	-
pathologic_M	M 0	168	-	-
	M 1/ M X	28	-	-
Tumor Stage	Stage I	23	-	41
	Stage II	112	-	31
	Stage III	57	-	19
	Stage IV	4	-	3
Pathology	IDC	121	-	34
	ILC	5	-	2
	MIX	24	-	3
	Others	25	-	1

#### Infiltrating cell score in TME

CIBERSORT is a deconvolution algorithm that uses a set of reference gene-expression values (a signature with 547 genes) considered a minimal representation for each cell type and, based on those values, infers cell type proportions in data from bulk tumor samples with mixed cell types using support vector regression, CIBERSORT is able to identify 22 kinds of human immune cells according to the gene

expression data with high specificity and sensitivity, including B cells, T cells, NK cells, macrophages, DCs and myeloid subset cells. In order to quantify the proportion of immune cells in breast cancer samples, with LM22 gene signature as a reference, we used the CIBERSORT algorithm (17) calculation scored 22 kinds of immune cells in TCGA and GSE13041 database. In particular, the gene expression data were uploaded to CIBERSORT website (<http://cibersort.stanford.edu/>), the LM22 signatures was setted up and 1000 times permutation obtained scored of 22 kinds immune cells.

#### Consensus clustering to obtain molecular subtypes associated with TME-infiltrating cells

Consensus clustering was performed using the ConsensusClusterPlus (23) package in R to determine subgroups of GBM based on the TME-infiltrating cells. As zhang et al. (24) evaluated the optimal number of clusters between  $k = 2-10$ , repeated 1000 times to ensure the stability of the results, and used the R software package pheatmap for visualization.

#### Differentially expressed genes (DEG) associated with the TME phenotype

To identify genes associated with TME cell-infiltrating patterns, the linear model was used to analyze the gene expression differences among phenotypic related TME subgroups. Specifically, the R software package DESeq2 (25) was used to calculate the differential genes, and  $FDR < 0.05$  was selected. In order to include more candidate genes, Foldchange was not restricted.

#### TME phenotype related differential gene re-clustering

Nonnegative matrix factorization (NMF) is an unsupervised clustering method that is widely used in the discovery of genomics-based tumor molecular subtypes (26, 27). In order to further observe the relationship between the expression of TME phenotype-related differential genes and phenotype, we used the NMF method to re-clusters the samples based on the expression profiles of TME phenotype-related differential genes, and analyzed the clinical features of the re-clustered samples. NMF The method selects the standard "brunet" for 50 iterations, the number of clusters  $k$  is set to 2 to 10, and the average profile width of the common member matrix is calculated using the R package NMF (28), and the minimum member of each subclass is set to 10.

#### Dimension reduction and generation of TME gene signatures

In order to obtain robust TME gene signatures, the random forest algorithm is further used to assess the importance of these DEGs. Specifically, the coph function of survival of R software package was used for univariate survival analysis, and the selection threshold was 0.05, and the genes with significant prognosis were included for randomForest feature selection using randomForest of R software package, and mtry of each segmentation set as 1-235 and ntree = 500, and the mtry value with the lowest error rate was selected as the optimal mtry value of the randomForest algorithm. Then ntree = 100 was selected according to the error rate of random forest. Finally, each DEG was ranked according to its importance, and 95% DEGs with cumulative importance  $> 95\%$  were selected as candidate feature genes. k-means (29) was used to divide these genes into 5 categories. Psych R package was used for principal

component analysis of the expression profiles of 5 categories of genes, and the first principal component was extracted as signature score after 100 iterations. This approach has the advantage of focusing the score on the set with the largest block of well-correlated (or anticorrelated) genes in the set, while downweighting contributions from genes that do not track with other set members. For gene type  $j$ , signature score formula of samples is as follows:

$$S_j = \sum_{i=1}^{n_j} Pc1_i * Exp_i$$

$j$  represents the  $j$ th class of the five types of genes, wherein  $n_j$  represents the number of genes of the  $j$ th gene,  $Pc1_i$  represents the first principal component coefficient of the  $i$ th gene of the  $j$ th gene, and  $Exp_i$  indicates the first The  $i$ th gene expression level of the  $j$ -type gene.

PCA analysis was conducted on signature G1, signature G2 and signature G3 using R's psych package. For each gene signature, 100 iterations were conducted to obtain the optimal number of principal components (PCs), and then the respective PCs scores were calculated. PC1 scores of G1, G2 and G3 were selected as the final score. Cox multivariate regression was used to obtain the signature score risk coefficients of each G1, G2 and G3. For any sample, the TMEscore formula is as follows:

$$\text{TME score} = \sum PC1 * \beta$$

Where,  $\beta$  is the multivariate regression coefficient of each signature, and PC1 is the PC1 score of each signature.

#### Relationship between TME score and clinical features

To observe the relationship between TMEscore and clinical phenotype, the samples were divided into two groups based on the median TMEscore of the samples, which compared the prognostic differences between high TMEscore and low TMEscore, respectively. Similarly, the relationship between high and low TMEscore and age and gender is analyzed.

#### Relationship between TME Score and immune related gene expression

In order to observe the relationship between TMEscore and immune-related genes, three types of immune-related genes were first collected, 1) immune activation genes CXCL10, CXCL9, GZMA, GZMB, PRF1, IFNG, TBX2, TNF, CD8A) (30). 2) Immunological checkpoint genes PDCD1, CTLA4, LAG3, IDO1, CD274, PDCD1LG2, HAVCR2) (31). 3) TGF/EMT pathway genes VIM, ACTA2, COL4A1, TGFBR2, ZEB1, CLDN3, SMAD9, TWIST1 (32). The expression profiles of these genes were extracted to further analyze the differential expression of these three types of genes in high TME Score and low TME Score.

## Relationship between TME Score and tumor genomic variation

To observe differences in genomic variation between high TMEScore and low TMEScore samples, we downloaded SNP data from TCGA, removed intron and silent mutations, and used Fisher's exact test to analyze the difference in the mutations between the two samples, selected the threshold  $p < 0.05$ .

## Statistical Analysis

The normality of the variables was tested by the Shapiro-Wilk normality test (33), unless otherwise specified. For the comparison of the two groups, the statistical significance of the normal distribution variables was estimated by the unpaired Student t test, and the non-normal distribution variables were analyzed by the Mann-Whitney U test. For comparisons between the two groups, the Kruskal-Wallis test and the one-way ANOVA were used as non-parametric and parametric methods, respectively (34). Correlation coefficients were calculated by Spearman and distance correlation analysis. The two-sided Fisher exact test was used to analyze the contingency table, and we used the Benjamini-Hochberg method to convert the P value to FDR. The Kaplan-Meier method was used to generate survival curves for the subgroups in each data set, and the logrank test was used to determine the statistical significance of the differences, with significance being defined as  $P < 0.05$ . All of these analyses were performed in R 3.4.3, and all analyses were based on default parameters unless otherwise stated.

# Results

## TME landscape in Luminal B breast cancer

To investigate the association of various immune cells in TME in Luminal B breast cancer, we used CIBERSORT to calculate the correlation of 22 immune cell scores in 209 Luminal B breast cancer samples, there is significant positive correlation in four types such as signs of immune activation cell, directly reflecting the immune cells have a particular mode of communication (Fig. 1A). Univariate cox regression was used to analyze the relationship between 22 immune cell scores and prognosis, and found that the scores of Mast cells activated and Macrophages M2 were significantly associated with poor prognosis (log rank  $p < 0.05$ ,  $HR > 1$ ), while T cells CD4 memory The scores of activated and Macrophages M1 are associated with a better prognosis (log rank  $p < 0.05$ ,  $HR < 1$ ) (Fig. 1B). First, the scores of 6 immune cells with significant correlation with prognosis were selected for consensus clusterPlus, and the optimal number of clusters between  $k = 2-10$  was evaluated. 1000 times was repeated, and  $k = 3$  was selected as the optimal number of clusters according to the CDF value and Delta area (S1\_Figure). TME scores are grouped into 3 categories, defined as TMEC1-TMEC3, According to the clustering results, immune cells such as B cells naive and T cells regulatory (Tregs) scored significantly higher in TMEC1, while T cells CD4 resting and Macrophages M1 scored significantly higher in TMEC2. Macrophages M2 and NK cells resting have relatively high scores in TMEC3 (Fig. 1C). The prognosis analysis of OS between TMEC showed that there was a significant difference in OS prognosis between TMEC (log rank  $p < 0.0001$ ), and TMEC3 had the worst prognosis, TMEC2 had the best prognosis, and TMEC1 was between the two (Fig. 1D). Furthermore, we found that there were significant prognostic

differences between TMEC3 and TMEC1 and TMEC2, respectively, while there were no significant differences between TMEC1 and TMEC2 (S2\_Figure). We also analyzed the prognostic differences of the 22 immune cell scores among the three types of samples, 12 of which (54.54%) had significant differences (Fig. 1E). In conclusion, TME score may be closely related to the development of luminal type B breast cancer.

### Differential expression and functional analysis of genes between TMEC

To investigate differences in gene expression patterns between different TMECs, TCGA data was selected for TMEC differential expression analysis. The differentially expressed genes (DEGs) between TMEC3 and TMEC1, TMEC2, which had the worst prognosis, were analyzed using the DESeq2 tool, and 241 DEGs shared between TMEC3/TMEC1, TMEC3/TMEC2 for subsequent analysis (Fig. 2A). Based on a total of 241 DEGs, we first filtered out the genes with a value of 0 in 50% of the samples, and finally obtained 235 genes for subsequent analysis. NMF was used for reclassification analysis of TCGA samples to obtain two stable samples, which were defined as GeneC1 and GeneC2, respectively (Fig. 2B). OS prognostic analysis showed that there was a significant difference in total survival between GeneC1 and GeneC2 (Fig. 2C). Comparing the distribution of GeneC1 and GeneC1 on 22 immune cell scores, GeneC1 and GeneC1 were significantly different in various TMEs. For example, GeneC1 with the worst prognosis scored significantly higher on Macrophages M0, Macrophages M1 and Macrophages M2 than GeneC2 (Fig. 2D).

### Construction of the TME signature

To further screen the 235 DEGs shared between TMECs, we evaluated the importance of these 235 DEGs using the R package random Forest. According to the plot of the random forest, ntree = 100 was selected (S3\_FigureA), and 24 candidate genes were selected by selecting DEGs with cumulative importance  $\geq$  95% (S3\_FigureB, S3\_FigureC), and these 24 genes are mainly enriched in T cell activation, NF - kappa B signaling pathway, Primary immunodeficiency and other immune-related GO and KEGG Pathway (Fig. 3A, B). K-means algorithm was used to cluster them, and the number of clustering was 3 (Fig. 3C), defined as signature G1, signature G2, signature G3, and contains 10, 6 and 8 genes, respectively (Fig. 3C), among G3 was in the high expression group, G2 was in the low expression group, and G1 was in the middle (Fig. 3D).

As described in the Methods section, TMEScore is built to calculate the TMEScore for each sample in the training set. Comparing TME score of Gene C, we found that the worst prognosis of GeneC1 was significantly higher than the best prognosis of GeneC2 (Fig. 3E, F). With the median of TME score (0.3657), the samples were divided into two categories: risk-h and risk-l. The prognosis of OS in the risk-h group and the risk-l group was significantly different (log rank p = 0.0024, HR = 2.928) (Fig. 3G).

### Relationship between clinical characteristics of TMEScore and immune gene expression

Furthermore, we assessed the relationship between TME scores and clinical characteristics, and found no significant difference (Figure S4). In order to study the relationship between different TME and immune state, the relationship between the expression of immune activation genes (CXCL10, CXCL9, GZMA, GZMB, PRF1, IFNG, TBX2, TNF, CD8A) and TMEC, GeneC and TMEScore was analyzed, and there were different expression patterns in different GeneC, TMEScore and TMEC, in which the expression of CXCL10 and CXCL9 in the risk-H group with poor prognosis was significantly higher than that in the risk-L group (Fig. 4A). The relationship between the expression of immune checkpoint genes (PDCD1, CTLA4, LAG3, IDO1, CD274, PDCD1LG2 and HAVCR2) and TMEC, GeneC and TMEScore was analyzed, although the expression level was not high, different GeneC, TMEScore and TMEC also had different expression trends, among which the expression of HAVCR2 and IDO1 in the risk-H group with poor prognosis was significantly higher than that in the risk-L group (Fig. 4B). The expression levels of TGF/EMT pathway genes VIM, ACTA2, COL4A1, TGFBR2, ZEB1, CLDN3, SMAD9 and TWIST1 in TMEC, GeneC and TMEScore were analyzed, and found that there was little difference in the expression patterns of different GeneC, TMEScore and TMEC, only TGFBR2 was significantly higher in the risk-h group with poor prognosis than in the risk-l group (Fig. 4C). Similarly, the same phenomenon was observed in three groups of TME samples (Figure S5, S6). In conclusion, different TMEScore is closely related to the expression of immune genes.

#### TME characteristics of tumor genome

Patients were divided into risk-h and risk-l groups according to TME score. We compared the relationship between TME score and genomic variation and screened out a group of important genes related to TME score. Fisher test was used to compare the genes with significant difference in mutation frequency between risk-h and risk-l groups (intron and silent mutations were removed), and a total of 14 genes with prominent mutation were obtained (Fig. 5). The results showed that the mutation frequency of TP53 gene in risk-h was lower than that in risk-l, which might indicate that these genes were significantly correlated with TME in luminal B breast cancer.

#### Comparison of risk models with other models

Three prognostic-related risk models were selected: 4-gene signature (Luthra) (35), 4-gene signature (Li) (36), 4-gene signature (Xie) (37) and our TMEScore model was compared. In order to make the models comparable, according to the corresponding genes in the three models, the same method was used to calculate risk scores in the TCGA Luminal B breast cancer samples, and the ROC of each model was evaluated and based on the median risk score. The samples were divided into the Risk-H and Risk-L groups, and the difference in OS prognosis between the two groups was calculated. Five-year AUC of TMEScore is 0.82, and the ROC results of the other three models are not very good. Only the Li et al. model has a 3-year AUC greater than 0.61 (Fig. 6A-D). Similarly, the prognostic KM curves of the three models were observed, and found that no significant differences in the Risk-H and Risk-L group samples (Fig. 6E-G). Finally, in order to compare the predictive performance of these models on Luminal B breast samples, the rms package in R was used to calculate the concordance index (C-index) of the three models

and the TMEscore model. The TMEscore model of the four models has the highest C-index (Fig. 6H), indicating that the overall performance of the model is better than the other three.

## Discussion

Up to now, malignant tumor is still a major threat to human health. In the past, research has tended to focus on the tumor cells themselves. In 1989, Paget(38) put forward the theory of "seed and soil", which received widespread attention once it was put forward, and thus the tumor microenvironment entered the researchers' vision. With the development of the research, it is found that the tumor is a complete tissue with its own homeostasis (39), which makes the treatment plan of the tumor gradually shift from focusing solely on the tumor itself to simultaneously intervening in the tumor microenvironment. In recent years, immunotherapy have shown significant efficacy in the treatment of some advanced cancer patients (40). Tumor infiltrating immune cells and immune responses in tumor microenvironment have attracted great attention from researchers and become promising therapeutic targets. In this study, a systematic analysis of the TME landscape in Luminal B breast cancer was conducted. Among the 22 types of immune cells, a variety of them were correlated with the poor prognosis of Luminal B breast cancer, and they had the potential for molecular stratification of Luminal B breast cancer, which suggested that the immune characteristics of Luminal B breast cancer in different stages were different.

In the tumor microenvironment, T cells are regulated by both stimulating and inhibitory signal, and immune checkpoint is the "switch" of inhibition. Schadendorf et al. (41) analyzed the clinical data of 1 861 patients with advanced melanoma and found that anti-CTLA-4 antibody (Ipilimumab) significantly increased the long-term survival rate of patients. Recent studies have shown that anti-PD-1 antibody (Nivolumab) is not only superior in melanoma, but also effective in the treatment of non-small cell lung cancer, metastatic renal cell carcinoma, advanced squamous cell carcinoma and other tumors (42). Therefore, it is important to develop biomarkers to predict the benefits of checkpoint immunotherapy, and the latest data support the view that TME plays a key role in checkpoint inhibitor immunotherapy (43). The role of PD-L1 blockade is enhanced by regulating the immune microenvironment in small cell lung cancer (44). Multi-functional nano-regulator reshapes immune microenvironment and enhances immune memory of tumor immunotherapy (45). Therefore, the development of biomarkers to predict checkpoint immunotherapy is extremely important. Here, we have elucidate the comprehensive landscape of interactions between Luminal B breast cancer clinical characteristics and infiltrating TME cells, and with the help of several computational algorithms, we have established a method to quantify the infiltrating pattern of TME -TMEscore.

Comprehensive analysis showed that TMEscore was a prognostic biomarker for Luminal type B breast cancer, and high TMEscore was associated with poor prognosis. An immune-related gene expression pattern was identified in liver tissues of patients with early-stage HCC that associates with risk of HCC development (46). Zemek RM et al. found that activation of a STAT1/NK axis in the tumor microenvironment could predicts response to immune checkpoint blockade, which suggest a biomarker-driven approach to patient management to establish whether a patient would benefit from treatment with

sensitizing therapeutics before immune checkpoint blockade (47). In this study, it was found that CXCL10, CXCL9, GZMB and other immune activation genes as well as PDCD1, CTLA4 and other immune checkpoint genes were highly expressed in high TMEScore, which indicated the potential of TMEScore as a marker for the benefit of immune checkpoint blocking therapy. TP53 mutation is one of the factors of breast cancer and plays an important role in clinical treatment and prognosis (48). The frequency of TP53 mutations in high TMEScore samples was significantly lower than that in low TMEScore samples. These data can provide a new perspective for studying the mechanism of TME formation and explore the role of individual mutations in Luminal B breast cancer immunity and immunotherapy.

Although we identified potential candidate immune gene markers involved in Luminal type B breast cancer in a large sample using bioinformatics techniques, further validation should be performed in a prospective cohort of immunotherapies to more fully define cut-off values to be used. Second, given the heterogeneity of different tumor area, on the edge of the core and invade the tumor to evaluate the immune system infiltration is appropriate, because not all of the patients with high TMEScore immunotherapy has greater benefits, so it should be more clinical factors into the model to improve accuracy, in the end, only the results obtained through bioinformatics analysis is inadequate, experimental verification is needed to confirm these results. Therefore, further genetic and experimental studies with a larger sample size and experimental validation are needed.

## Conclusions

In summary, in this study, we systematically evaluated TME infiltration patterns from 209 Luminal B breast cancer patients and developed methods for TME infiltration patterns. TME score was found to be a strong prognostic biomarker and predictor of response to immunotherapy inhibitors.

## Abbreviations

Tumor microenvironment (TME); dendritic cells (DCs); Gene Expression Omnibus (GEO); Differentially expressed genes (DEG); Nonnegative matrix factorization (NMF);

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

All authors agree with publication.

### Data Availability

The data used for supporting the results of the study are included within the article.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Fund

This work was supported by grants from National Nature Science Foundation of China (NCFS No.81860469).

## Authors' contributions

QL designed the study. FJG and KH contributed to the literature search. XML and NJ contributed to date acquisition; QYW and YL analyzed data; MGX and SHS interpreted data; QL was responsible for writing the initial draft of the manuscript. The manuscript has been read and approved by all authors.

## Acknowledgement

None.

## References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406(6797):747-52.
3. Prat A, Perou CM. Deconstructing the molecular portraits of breast cancer. *Molecular oncology*. 2011;5(1):5-23.
4. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Annals of oncology : official journal of the European Society for Medical Oncology*. 2011;22(8):1736-47.
5. Blows FM, Driver KE, Schmidt MK, Broeks A, van Leeuwen FE, Wesseling J, et al. Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS medicine*. 2010;7(5):e1000279.
6. Kim J, Bae JS. Tumor-Associated Macrophages and Neutrophils in Tumor Microenvironment. *Mediators Inflamm*. 2016;2016:6058147.
7. Zeng D, Zhou R, Yu Y, Luo Y, Zhang J, Sun H, et al. Gene expression profiles for a prognostic immunoscore in gastric cancer. *Br J Surg*. 2018;105(10):1338-48.
8. Jiang Y, Zhang Q, Hu Y, Li T, Yu J, Zhao L, et al. ImmunoScore Signature: A Prognostic and Predictive Tool in Gastric Cancer. *Ann Surg*. 2018;267(3):504-13.
9. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as

treatment targets in oncology. *Nat Rev Clin Oncol*. 2017;14(7):399-416.

10. Kalluri R. The biology and function of fibroblasts in cancer. *Nat Rev Cancer*. 2016;16(9):582-98.
11. Fridman WH, Zitvogel L, Sautes-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol*. 2017;14(12):717-34.
12. Turley SJ, Cremasco V, Astarita JL. Immunological hallmarks of stromal cells in the tumour microenvironment. *Nat Rev Immunol*. 2015;15(11):669-82.
13. Rosenberg JE, Hoffman-Censits J, Powles T, van der Heijden MS, Balar AV, Necchi A, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet*. 2016;387(10031):1909-20.
14. Nishino M, Ramaiya NH, Hatabu H, Hodi FS. Monitoring immune-checkpoint blockade: response evaluation and biomarker development. *Nat Rev Clin Oncol*. 2017;14(11):655-68.
15. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFbeta attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*. 2018;554(7693):544-8.
16. Lee K, Hwang H, Nam KT. Immune response and the tumor microenvironment: how they communicate to regulate gastric cancer. *Gut Liver*. 2014;8(2):131-9.
17. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453-7.
18. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol*. 2016;17(1):218.
19. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
20. Fu H, Zhu Y, Wang Y, Liu Z, Zhang J, Xie H, et al. Identification and Validation of Stromal Immunity Predict Survival and Benefit from Adjuvant Chemotherapy in Patients with Muscle-Invasive Bladder Cancer. *Clin Cancer Res*. 2018;24(13):3069-78.
21. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, et al. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol Res*. 2019;7(5):737-50.
22. Lee Y, Scheck AC, Cloughesy TF, Lai A, Dong J, Farooqi HK, et al. Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. *BMC Med Genomics*. 2008;1:52.
23. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869-74.
24. Zhang S, Wang Y, Gu Y, Zhu J, Ci C, Guo Z, et al. Specific breast cancer prognosis-subtype distinctions based on DNA methylation patterns. *Mol Oncol*. 2018;12(7):1047-60.
25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
26. Mirzal A. Nonparametric Tikhonov Regularized NMF and Its Application in Cancer Clustering.

IEEE/ACM Trans Comput Biol Bioinform. 2014;11(6):1208-17.

27. Yu N, Gao YL, Liu JX, Shang J, Zhu R, Dai LY. Co-differential Gene Selection and Clustering Based on Graph Regularized Multi-View NMF in Cancer Genomic Data. *Genes (Basel)*. 2018;9(12).
28. Ye C, Toyoda K, Ohtsuki T. Blind Source Separation on Non-contact Heartbeat Detection by Non-negative Matrix Factorization Algorithms. *IEEE Trans Biomed Eng*. 2019.
29. Boillaud E, Molina G. Are judgments a form of data clustering? Reexamining contrast effects with the k-means algorithm. *J Exp Psychol Hum Percept Perform*. 2015;41(2):415-30.
30. Shi W, Zou R, Yang M, Mai L, Ren J, Wen J, et al. Analysis of Genes Involved in Ulcerative Colitis Activity and Tumorigenesis Through Systematic Mining of Gene Co-expression Networks. *Frontiers in physiology*. 2019;10:662.
31. Huffaker TB, Lee SH, Tang WW, Wallace JA, Alexander M, Runtsch MC, et al. Antitumor immunity is defective in T cell-specific microRNA-155-deficient mice and is rescued by immune checkpoint blockade. *The Journal of biological chemistry*. 2017;292(45):18530-41.
32. Shao M, Wen ZB, Yang HH, Zhang CY, Xiong JB, Guan XX, et al. Exogenous angiotensin (1-7) directly inhibits epithelial-mesenchymal transformation induced by transforming growth factor-beta1 in alveolar epithelial cells. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2019;117:109193.
33. Ghasemi A, Zahediasl S. Normality tests for statistical analysis: a guide for non-statisticians. *Int J Endocrinol Metab*. 2012;10(2):486-9.
34. Hazra A, Gogtay N. Biostatistics Series Module 3: Comparing Groups: Numerical Variables. *Indian J Dermatol*. 2016;61(3):251-60.
35. Luthra S, Chandran U, Diergaarde B, Becich M, Lee AV, Neumann CA. Expression of reactive species related genes is associated with patient survival in luminal B breast cancer. *Free radical biology & medicine*. 2018;120:170-80.
36. Li Z, Zhang Y, Zhang Z, Zhao Z, Lv Q. A four-gene signature predicts the efficacy of paclitaxel-based neoadjuvant therapy in human epidermal growth factor receptor 2-negative breast cancer. *Journal of cellular biochemistry*. 2019;120(4):6046-56.
37. Xie X, Wang J, Shi D, Zou Y, Xiong Z, Li X, et al. Identification of a 4-mRNA metastasis-related prognostic signature for patients with breast cancer. *Journal of cellular and molecular medicine*. 2019;23(2):1439-47.
38. Paget S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer metastasis reviews*. 1989;8(2):98-101.
39. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501(7467):346-54.
40. Miller JF, Sadelain M. The journey from discoveries in fundamental immunology to cancer immunotherapy. *Cancer cell*. 2015;27(4):439-49.
41. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell*. 2017;168(4):707-23.
42. Sharpe AH, Pauken KE. The diverse functions of the PD1 inhibitory pathway. *Nature reviews Immunology*. 2018;18(3):153-67.
43. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers

for PD-1 checkpoint blockade-based immunotherapy. *Science*. 2018;362(6411).

44. Sen T, Della Corte CM, Milutinovic S, Cardnell RJ, Diao L, Ramkumar K, et al. Combination treatment of the oral CHK1 inhibitor, SRA737 and low dose gemcitabine, enhances the effect of PD-L1 blockade by modulating the immune microenvironment in small cell lung cancer. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2019.

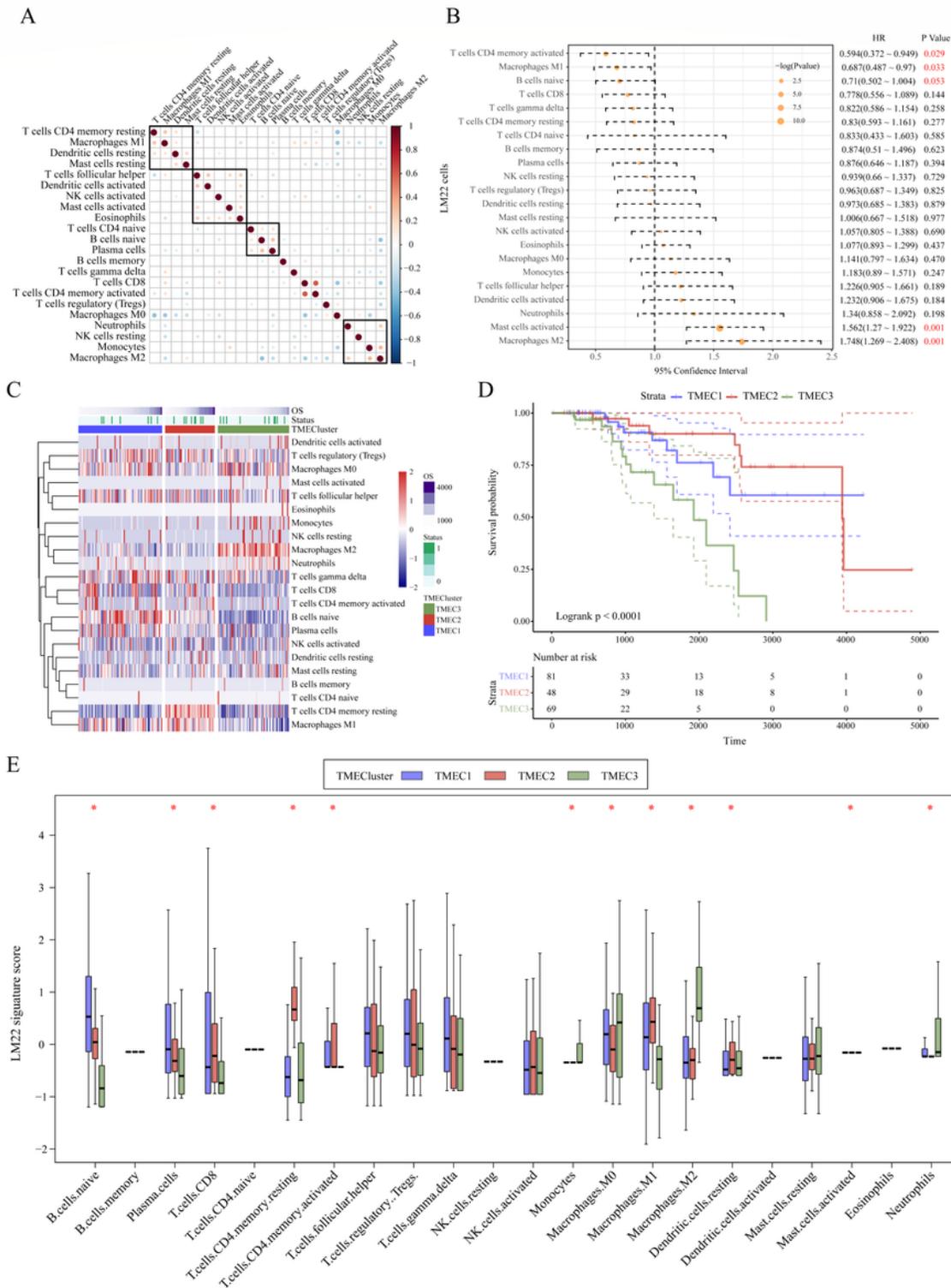
45. Yu M, Duan X, Cai Y, Zhang F, Jiang S, Han S, et al. Multifunctional Nanoregulator Reshapes Immune Microenvironment and Enhances Immune Memory for Tumor Immunotherapy. *Advanced science (Weinheim, Baden-Wurtemberg, Germany)*. 2019;6(16):1900037.

46. Moeini A, Torrecilla S, Tovar V, Montironi C, Andreu-Oller C, Peix J, et al. An Immune Gene Expression Signature Associated With Development of Human Hepatocellular Carcinoma Identifies Mice That Respond to Chemopreventive Agents. *Gastroenterology*. 2019.

47. Zemek RM, De Jong E, Chin WL, Schuster IS, Fear VS, Casey TH, et al. Sensitization to immune checkpoint blockade through activation of a STAT1/NK axis in the tumor microenvironment. *Science translational medicine*. 2019;11(501).

48. Wang X, Su D, Qin Z, Chen Z. Identification of FOXP4 as a tumor suppressor of breast carcinogenesis via the activation of TP53 and deactivation of Notch signaling. *Gene*. 2019:144057.

## Figures



**Figure 2**

A: Correlation of 22 immune cells in TME, where dot size and color indicate correlation, blue indicates negative correlation, red indicates positive correlation, and white areas in the figure indicate insignificant. B: Forest map of 22 immune cells in TME. C: heat map of 22 immune cells score in TME, the redder indicates the higher the score, the lower the basket indicates the lower the score. D: KM curve of OS

prognosis for three types of TMEC. E: The distribution box plots of 22 immune cell scores in the three classes of TMEC, red\* indicate significant differences.



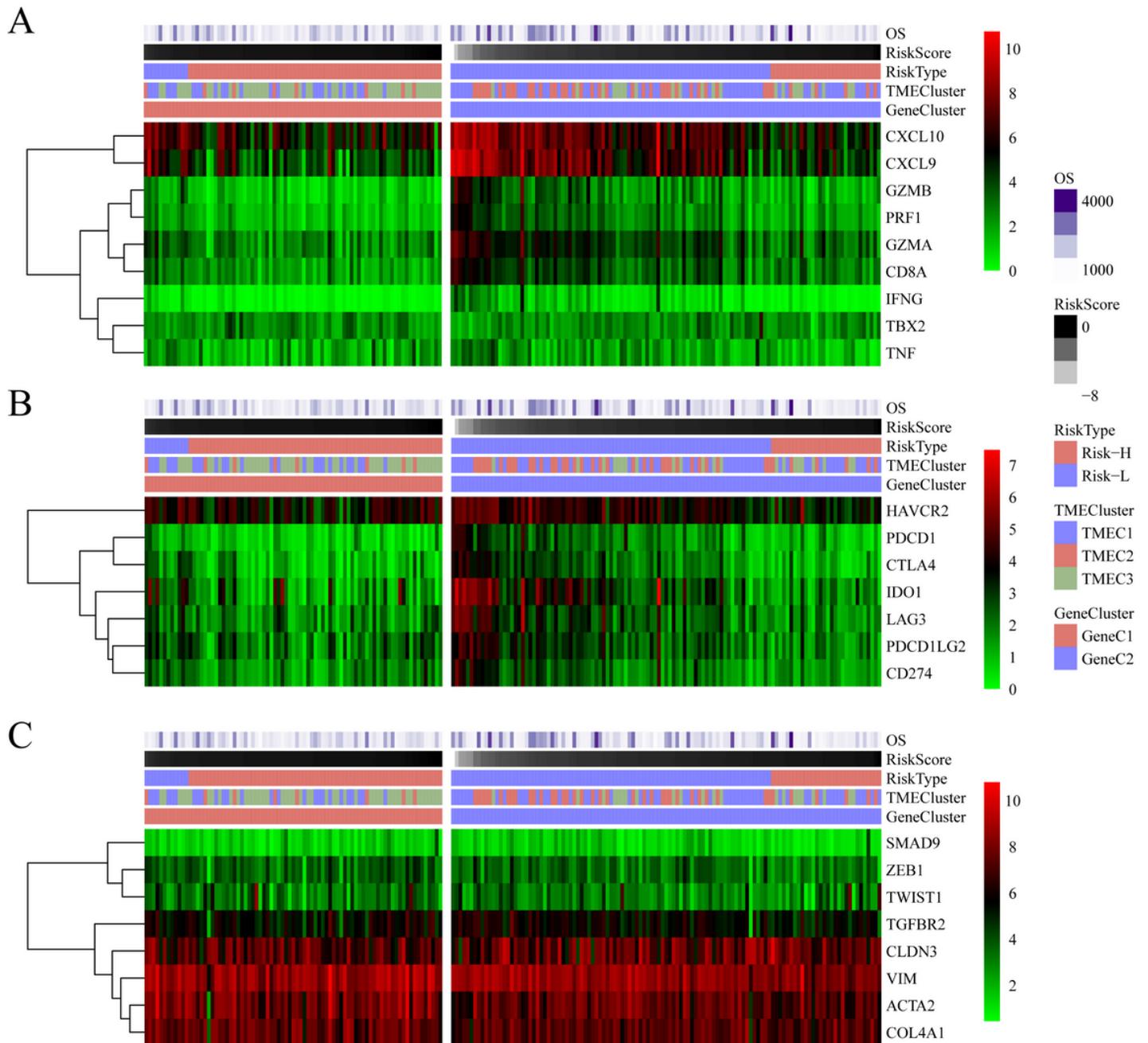
#### Figure 4

A: Venn plots of differentially expressed genes up-regulated by TMEC1, TMEC2 and TMEC3. B: Heatmap of NMF algorithm consistency matrix. C: KM curve of OS prognostic for GeneC1 and GeneC2. D: Box plot of 22 immune cell scores in GeneC1 and GeneC2 samples.



#### Figure 6

A: GO enrichment of 24 genes. B: KEGG enrichment of 24 genes. C: K-means clustering results of 24 genes. D: Heatmaps of 24 gene expressions. E: TME score of GeneC1 and GeneC2. F: TME score distribution of GeneC1 and GeneC2. G: KM curve of OS prognosis in Risk-H and Risk-L groups.



**Figure 8**

A: Heatmap of Immune activation genes were expressed in TCGA samples. B: Heatmap of immune checkpoint genes were expressed in TCGA samples. C: Heatmap of TGF pathway genes were expressed in TCGA samples.



**Figure 10**

The relationship between TME and genomic mutation characteristics. The horizontal axis represents samples, the vertical axis represents genes, the black rectangle represents mutations, and the gray

represents no mutations.

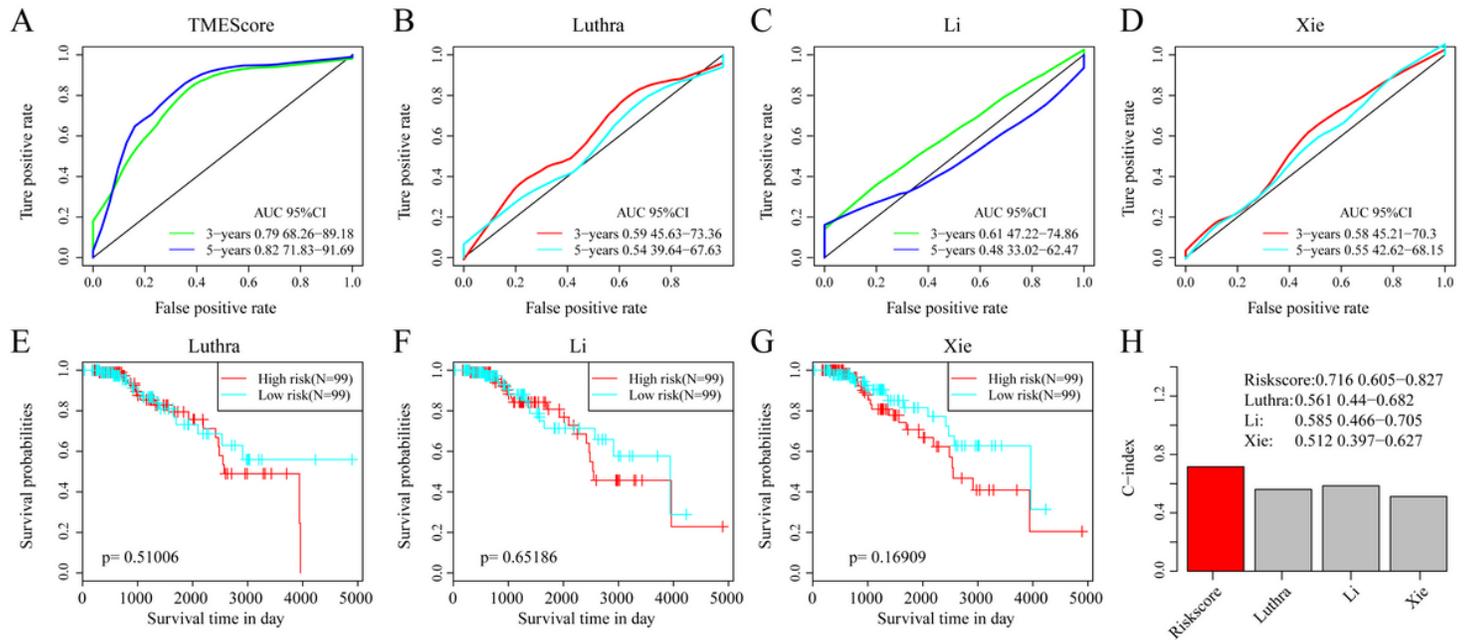


Figure 12

A-D: ROC curve of 4 models. E-G: Survival curve of 3 models. H: C-index of 4 models.

## Supplementary Files

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

- [FigureS1.tif](#)
- [FigureS6.tif](#)
- [FigureS2.tif](#)
- [FigureS1.tif](#)
- [FigureS6.tif](#)
- [FigureS4.tif](#)
- [FigureS5.tif](#)
- [FigureS4.tif](#)
- [FigureS3.tif](#)
- [FigureS2.tif](#)
- [FigureS3.tif](#)
- [FigureS5.tif](#)