Novel Immune-Related Gene Signature for Risk Stratification and Prognosis of Survival in Estrogen Receptor (ER) or Progesterone Receptor (PR) Positive and Human Epidermal Growth Factor Receptor 2 (HER2) Negative Breast Cancer

Wei Wang (wangwei1981@sxmu.edu.cn)
Second Hospital of Shanxi Medical University
Hongnan Jiang
Second Hospital of Shanxi Medical University
Yanrong Gao
Second Hospital of Shanxi Medical University

Research

Keywords: Breast cancer, Immune gene signature,Prognosis, Immune cell, cancer phenotypes

Posted Date: May 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-489056/v1

Abstract

Background: Although intrinsic molecular subtype has been extensively used, the risk stratification have not been fully elucidated in estrogen receptor (ER) or progesterone receptor (PR) positive and human epidermal growth factor receptor 2 (HER2) negative breast cancer.

Methods: RNA transcriptional data from The Cancer Genome Atlas (TCGA), METABRIC and GEO were used. Immune-related genes were obtained from the datasets and literature search. Univariate, lasso regression and multivariate cox regression were employed to identify prognostic immune-related genes and establish the risk signature. Relationships between the risk signature and clinical parameters, tumor-infiltrating immune cell abundances and cancer phenotypes were further evaluated.

Results: Noted, 102 immune-related prognostic genes were identified in METABRIC dataset by univariate cox analysis. Consecutively, 7 immune genes (SHMT2, AGA, COL17A1, FLT3, SLC7A2, ATP6AP1 and CCL19) were selected as risk signature by lasso regression and multivariate cox analysis. Its performance was further verified in TCGA,GSE20685 and GSE9195 datasets. Multivariate Cox regression indicated that the risk signature was an independent predictor. The prognostic signature showed significant correlation with intrinsic molecular subtypes, 70-gene signature and tamoxifen resistance signature. The CIBERSORT algorithm revealed that CD4+ memory T cells were significant higher in low-risk group. Conversely, M0-type macrophages were significant higher in high-risk group in both TCGA and METABRIC cohorts, which may have effect on the prognosis. Furthermore, we found that low-risk group may be associated with immune-related pathway and high-risk group was with cell cycle-related pathway, which also showed impact on the prognosis.

Conclusion: The present study constructed a robust seven immune-related gene signature and established an effective method in risk stratification and prediction of clinical outcome in ER or PR positive and HER2 negative breast cancer.

Introduction

Breast cancer ranks as the first in incidence rate among female malignant tumors and significantly affects women's health. Breast cancer now is considered a heterogeneous disease with different clinical and prognostic features. Though breast cancer has been classified into four subtypes with distinct molecular features and clinical outcome[1, 2], great heterogeneities were still found in each subgroup, especially in luminal breast cancer. [3, 4]

The most commonly used luminal A/B classification does not fully reveal the heterogeneity in luminal breast cancers and therefore failed to properly address risk stratification.[5, 6] Thus, it is necessary to explore new subtyping for prognostic prediction to guide individualized treatment beyond the existing molecular subtyping.

Gene expression analysis has emerged as a powerful tool to reveal molecular diversity of breast cancer.Recent studies showed that the immune-related gene signature were associated with prognosis in various types of cancer.[6] A recent study identified substantial heterogeneity in immune profiles across and within cancer types in The Cancer Genome Atlas (TCGA) Pan-Cancer datasets. [7]

In this study, we profiled immune gene expression of estrogen receptor (ER) or progesterone receptor (PR) positive and human epidermal growth factor receptor 2 (HER2) negative breast cancer in the TCGA, METABRIC and GEO datasets, for whom extensive clinical and transcriptional expression profile data were collected. We constructed a novel prognostic model, which was then evaluated and validated in independent cohorts for the prognostic robustness. Furthermore, we explored the association between the model and clinical parameter as well as the potential mechanism.

Materials And Methods

Data processing

We collected publicly transcriptional expression data of breast cancer cohorts from Gene Expression Omnibus ,UCSC xena and TCGA data portals. For METABRIC and TCGA database,we selected those patients with ER or PR positive and HER2 negative breast cancer.

The GSE20685 cohort was derived from a study of gene expression profiling conducted on fresh frozen breast cancer tissue collected from 327 patients in conjunction with thoroughly documented clinical data. All clinical and microarray data of these patients can be publicly downloaded at the GEO website (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20685) . The GSE9195 cohort was derived from a study of gene expression profiling including 77 ER or PR positive and HER2 negative breast cancer patients.All clinical and microarray data of these patients can be publicly downloaded at the GEO website (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE9195)

Bioinformatic analysis

R package genefu was used to classify the PAM50 molecular subtypes and calculate the 70-gene signature and the tamoxifen resistance signature (TAMR13) score of each case based on the gene expression data.[8] Gene set enrichment analysis (GSEA) was conducted using Molecular Signatures Database (MSigDB) collections by the R package clusterProfier[9] To determine the optimal number of groups, we used the Nbclust and ConsensusClusterPlus R package.[10, 11] The CIBERSORT algorithm was performed on transcriptional expression data to estimate the proportions of twenty-two types of immune cells in each case.[12] The ESTIMATE algorithms were used to characterize immune cell composition in a given sample.[13] Gene set analysis was carried out using the GSVA Bioconductor package.[14] We selected gene sets for various immune and cell cycle-related pathways. For each sample, enrichment score of selected gene set was obtained. Limma R package was used to identify significantly differential expressed genes (DEGs).[15] The Heat map of the representative DEGs were generated using the package ComplexHeatmap in R version 3.6.1.[16]

Statistical analyses

Statistical analyses were performed using SPSS version 23.0 (IBM, USA), GraphPad Prism version 8.00 (GraphPad Software, USA) and R version 3.6.1 (R Core Team, Vienna, Austria) . Pearson's chi-square test and Fisher's exact test were used to compare the categorical variables and ordered categorical variables Pearson correlation analysis was used to evaluate the association between two continuous variable. Mann–Whitney U tests were performed to evaluate the statistical significance within boxplots.Survival analysis was implemented by log-rank test. The Lasso Cox regression model was analysed using the glmnet package. Univariate and multivariate regression analyses were performed with the Cox proportional hazards regression model to determine the parameters that were significantly correlated with prognosis.

Results

Identification of prognostic immune-related genes

The work flow of this study is delineated in Fig. 1. A total of 2600 genes were identified from datasets and literature search. And then, the gene expression profiles from 1369 patients with ER or PR positive and HER2 negative breast cancer identified in the METABRIC dataset were used to perform the univariate cox analysis. As a result, 102 genes were discovered to be significantly associated with overall survival.

Then we performed the Lasso Cox regression analysis to eliminate the redundant collinearity and further validate the robustness. As a result, 7 genes were identified in Lasso regression from the 102 genes. Then we did stepwise multivariate Cox regression analysis of the 7 genes in the METABRIC dataset and ultimately, a prognostic signature comprising these 7 genes, including SHMT2, AGA, COL17A1, FLT3, SLC7A2, ATP6AP1, and CCL19, were selected to construct a prediction model. As shown in the forest plot, SHMT2 and ATP6AP1 were risk factors, whereas the other five genes were protective factors.

The comprehensive risk score (RS) was imputed as follows: $h0(t)*\exp(\beta1X1 + \beta2X2 +.... + \beta nXn)$. The Nbclust and ConsensusClusterPlus analysis showed that the optimal number of clusters was two. Therefore, the median risk score was set as the cut-off value, patients were categorized as two groups, RS-low and RS-high groups. **(**Fig. 2**)**

Performance of Risk Signature in Luminal breast cancer from METABRIC and TCGA

In order to validate the prognostic predicting role of the RS-based group in different datasets, METABRIC, TCGA, GSE20685 and GSE 9195 cohorts were included. The RS performed consistently in predicting long-term outcomes, including overall survival, disease free survival or distant-metastasis free survival. **(**Fig. 3**)**

In subgroup analysis, the RS also showed excellent predicting ability in luminal A, luminal B, stage II, with or without lymph node, and different age groups in both METABRIC and TCGA datasets. **(**Fig. 3**)**

Multivariate Cox analysis showed that the risk signature was the independent prognostic factor in both METABRIC and TCGA cohorts, after adjusting for established prognostic variables including TNM stage, age and tumor grade.**(Table.1–2)**

The Association Between Risk Signature And Clinicopathological Parameters

Subsequently, we analysed the relationship between the RS and clinicopathological parameters. It is shown that the RS-high group accounted for the highest percentage (> 80%) in the HER2-positive disease, followed by Luminal-B and Basal subtype in both METABRIC and TCGA cohort.**(**Fig. 4**)**

In order to explore the relationship between other gene prognostic score and our risk signature. 70-gene prognostic score and tamoxifen resistance signature (TAMR13) were calculated by genefu package. The results demonstrated that both 70-gene prognostic score and TAMR13 were significantly higher in the RS-high compared with RS-low group.**(**Fig. 4**)**

Correlation Of The Risk Signature With Tumor Microenvironment

To investigate tumor immunity relevance of the risk signature, the association of the RS with tumor purity and the presence of infiltrating stromal/immune cells in tumor tissues were evaluated by applying the ESTIMATE and CIBERSORT algorithm. As showed in Fig. 5, the stromal- and immune- score were both significantly higher in the RS-low group. RS was also strongly correlated with several important immune-related genes such as IL33 and TGFBR2. **(**Fig. 5**)**

By applying the CIBERSORT algorithm, the relative proportions of 22 immune cell subsets of ER or PR negative and HER2 positive breast cancer in the TCGA and METABRIC datasets were estimated. Consecutively, compared with RS-high group, the infiltration levels of CD4 memory resting T cells was increased, while the level of macrophage M0 was decreased significantly in RS-low group.**(**Fig. 5**)**

We further investigated the prognostic values of the abundance of infiltrative immune cells. The high abundance of CD4 memory resting T cell was significantly associated with favourable prognosis ($p < 0.001$), while the high abundance of macrophage M0 was associated with unfavourable survival ($p = 0.021$).**(**Fig. 5**)**

Risk signature is associated with immune and cell cycle related phenotypes

In order to further characterize the phenotype contributing to the worse prognosis in the high-risk group, we firstly performed differential expressed gene (DEGs) analysis of RS-high versus low group in the METABRIC and TCGA dataset. Then we performed GSEA using the collection of the MsigDB for these

DEGs. As a result,we found that in the RS-low group, the immune-related pathway such as GSE22886_NAIVE_BCELL_VS_NEUTROPHIL_UP and KEGG_CYTOKINE-CYTOKINE RECEPTOR_INTERACTION were significantly enriched, while in the RS-high group, the cell cycle-related pathway such as KONG_E2F3_TARGETS and ISHIDA_E2F_TARGETS were enriched. (Fig. 6)

The heatmap displayed the expression of the core genes that contribute to pathway enrichment between the two groups. Notably, those cell cycle-related genes were predominantly up-regulated in the high-risk group, while the immune-related genes were significantly up-regulated in the low risk group.(Fig. 6)

In order to explore the relationship between the enrichment of pathway and survival, we selected gene sets involving immune and cell cycle pathway. A total of 10 gene sets from the MsigDB were selected and the gene set enrichment score of each sample was calculated by the GSVA method. The activity of each gene set in each sample was estimated by the enrichment score. Then the univariate cox analysis were performed. As showed in Fig. 7, the enrichment of cell cycle-related pathway was significantly associated with worse survival, and the enrichment of immune-related pathway was significantly related with favourable survival.

Discussion

In this study, we provided a novel immune-related 7 gene signature with relevance for the prognosis in the ER or PR-positive and HER2-negative breast cancer. The risk signature was independent of other prognostic factors, including PAM50 and TNM stage. Besides, the group were characterized by specific composition of immune infiltrate and molecular phenotype, which may contribute to the clinical outcome. Our finding of this novel prognostic model demonstrated a association between tumor phenotype and immune gene expression.

Through dimensionality reduction, survival analysis and unsupervised consensus clustering, we identified two groups of patients with substantial prognostic difference in ER or PR-positive and HER2-negative breast cancer. This model (1) provided independent prognostic information (2) associated with specific immune microenvironment and phenotype.

Our findings added to the emerging body of evidence that the expression of immune genes could provide additional prognostic information. Xavier Tekpli et al [17] identified three groups of patients with distinct levels of immune infiltration. The intermediate cluster is correlated with poor clinical outcome and lower response to therapy. Additionally, proliferation and epithelial mesenchymal transition phenotypes were associated with the immune gene-based model. On the other hand, Bin Zhu et al[18] stratified the luminal breast cancer into three subtypes based on immune gene expression in an Asian population. One subtype was characterized by higher level of infiltrating lymphocytes, activation of immune checkpoint genes and higher tumor mutation burden.

Our findings are consistent with prior studies showing that SHMT2 was an important risk factors that contributed to poor clinical outcomes. Proteomics data verified that high SHMT2 protein expression was

significantly correlated with poor OS.[19, 20] Besides, findings from our analysis added more evidence on the protective effect of COL17A1, SLC7A2 and CCL19 expression in breast cancer.[21–23] On the other hand, the association between ATP6AP1 ,AGA and FLT3 expression and prognosis in breast cancer were firstly reported.

Conclusion

In summary, we developed and validated a 7-gene prognostic signature based on immune gene expression in ER or PR positive and HER2 negative breast cancer, displaying distinct patterns of prognosis and genomic features. If confirmed, these findings may have important clinical implications in risk stratification for precision oncology treatment in this population.

Abbreviations

ER:estrogen receptor; PR: progesterone receptor;HER2:human epithelial growth factor receptor 2; TCGA: The Cancer Genome Atlas;TAMR13: tamoxifen resistance signature ; GSEA: Gene set enrichment analysis; MSigDB: Molecular Signatures Database;DEGs: differential expressed genes;RS: risk score

Declarations

Authors' contributions

Conception and design: Wei Wang; Acquisition and analysis of data (provided tissue microarray, statistical analysis, biostatistics, etc.): Wei Wang; Hongnan Jiang; Yanrong Gao. Writing, review, and/or revision of the manuscript: Wei Wang; Hongnan Jiang; Yanrong Gao. Study supervision: Wei Wang. The authors read and approved the final manuscript.

Funding

The authors declared no funding

Availability of data and materials

Three public datasets were obtained from TCGA (https://portal.gdc.cancer.gov/), METABRIC (http://www.cbioportal.org/) and GEO (https://www.ncbi.nlm.nih.gov/geo)

Ethics approval and consent to participate

This article was approved by the medical ethics committee of The second hospital of Shanxi Medical University

Consent for publication

Written informed consent was obtained from all participants.

Competing interests

The authors declare that they have no conflicts of interest.

Acknowledgements

Not applicable

References

1. *Sørlie 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*: *10869-74*.
2. *Perou CM*, *Sørlie T*, *Eisen MB*, *van de Rijn M*, *Jeffrey SS*, *Rees CA*, et al *Molecular portraits of human breast tumours*. *Nature*. *2000*; *406*: *747 – 52*.
3. *Cancer Genome Atlas Network*. Comprehensive molecular portraits of human breast tumours. Nature. *2012*; 490: 61–70.
4. *Curtis C*, *Shah SP*, *Chin SF*, *Turashvili G*, *Rueda OM*, *Dunning MJ*, et al *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups*. *Nature*. *2012*; *486*: *346 – 52*.
5. *Gatza ML*, *Silva GO*, *Parker JS*, *Fan C*, *Perou CM*. *An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer*. *Nat Genet*. *2014*; *46*: *1051-9*.
6. *Netanely D*, *Avraham A*, *Ben-Baruch A*, *Evron E*, *Shamir R*. *Expression and methylation patterns partition luminal-A breast tumors into distinct prognostic subgroups*. *Breast Cancer Res*. *2016*; *18*: *74*.
7. *Thorsson V*, *Gibbs DL*, *Brown SD*, *Wolf D*, *Bortone DS*, *Ou Yang TH*, et al *The Immune Landscape of Cancer*. *Immunity*. *2018*; *48*: *812 – 30.e14*.
8. *Parker JS*, *Mullins M*, *Cheang MC*, *Leung S*, *Voduc D*, *Vickery T*, et al *Supervised risk predictor of breast cancer based on intrinsic subtypes*. *J Clin Oncol*. *2009*; *27*: *1160-7*.
9. *Yu G*, *Wang LG*, *Han Y*, *He QY*. *clusterProfiler: an R package for comparing biological themes among gene clusters*. *OMICS*. *2012*; *16*: *284-7*.
10. *Wilkerson MD*, *Hayes DN*. *ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking*. *Bioinformatics*. *2010*; *26*: *1572-3*.
11. *Malika Charrad NG*, *Veronique Boiteau AN*. *NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set*. *Neoplasma*. *2014*; *61*: *1–36*.
12. *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*: *453-7*.
13. *Yoshihara K*, *Shahmoradgoli M*, *Martínez 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*.

14. *Hänzelmann S*, *Castelo R*, *Guinney J*. *GSVA: gene set variation analysis for microarray and RNA-seq data*. *BMC Bioinformatics*. *2013*; *14*: *7*.
15. *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*: *e47*.
16. *Gu Z*, *Eils R*, *Schlesner M*. *Complex heatmaps reveal patterns and correlations in multidimensional genomic data*. *Bioinformatics*. *2016*; *32*: *2847-9*.
17. *Tekpli X*, *Lien T*, *Røssevold AH*, *Nebdal D*, *Borgen E*, *Ohnstad HO*, et al *An independent poor-prognosis subtype of breast cancer defined by a distinct tumor immune microenvironment*. *Nat Commun*. *2019*; *10*: *5499*.
18. *Zhu B*, *Tse LA*, *Wang D*, *Koka H*, *Zhang T*, *Abubakar M*, et al *Immune gene expression profiling reveals heterogeneity in luminal breast tumors*. *Breast Cancer Res*. *2019*; *21*: *147*.
19. *Bernhardt S*, *Bayerlová M*, *Vetter M*, *Wachter A*, *Mitra D*, *Hanf V*, et al *Proteomic profiling of breast cancer metabolism identifies SHMT2 and ASCT2 as prognostic factors*. *Breast Cancer Res*. *2017*; *19*: *112*.
20. *Wang CY*, *Chiao CC*, *Phan NN*, *Li CY*, *Sun ZD*, *Jiang JZ*, et al *Gene signatures and potential therapeutic targets of amino acid metabolism in estrogen receptor-positive breast cancer*. *Am J Cancer Res*. *2020*; *10*: *95–113*.
21. *Shi W*, *Hu D*, *Lin S*, *Zhuo R*. *Five-mRNA Signature for the Prognosis of Breast Cancer Based on the ceRNA Network*. *Biomed Res Int*. *2020; 2020: 9081852.*.
22. *Rodriguez-Ruiz ME*, *Buqué A*, *Hensler M*, *Chen J*, *Bloy N*, *Petroni G*, et al *Apoptotic caspases inhibit abscopal responses to radiation and identify a new prognostic biomarker for breast cancer patients*. *Oncoimmunology*. *2019*; *8*: *e1655964*.
23. *Cui Q*, *Tang J*, *Zhang D*, *Kong D*, *Liao X*, *Ren J*, et al *A prognostic eight-gene expression signature for patients with breast cancer receiving adjuvant chemotherapy*. *J Cell Biochem*. *2019*.

Tables

Table.1 Multivariate cox analysis of prognostic factors in METABRIC cohort					
		p value	HR	95%LI	95%UI
Tumor Grade		0.753	1.02	0.88	1.19
TNM stage		< 0.001	1.37	1.17	1.61
Risk Group		< 0.001	1.43	1.18	1.73
PAM50					
	Basal	Reference	-	-	-
	HER2	0.123	2.14	0.82	5.61
	Luminal-A	0.089	2.18	0.89	5.36
	Lumnal-B	0.09	2.09	0.89	4.89
	Normal	0.161	2.18	0.73	6.45
Number of positive node		< 0.001	1.07	1.04	1.10
Age		< 0.001	1.04	1.04	1.05
70-gene score		0.106	1.76	0.89	3.48

Table.2 Multivariate cox analysis of prognostic factors in TCGA cohort					
		P value	HR	95%LI	95%UI
Age		0.998	1.00	0.97	1.03
TNM stage		< 0.001	2.86	1.83	4.48
PAM50					
	Basal	Reference	-	-	-
	HER2	0.978	0.00	0.00	.
	Luminal A	0.744	0.73	0.11	4.76
	Luminal B	0.914	1.09	0.24	4.86
	Normal	0.865	0.80	0.06	11.28
Risk group		0.002	3.25	1.54	6.90
70- gene score		0.631	0.57	0.06	5.77

Figures

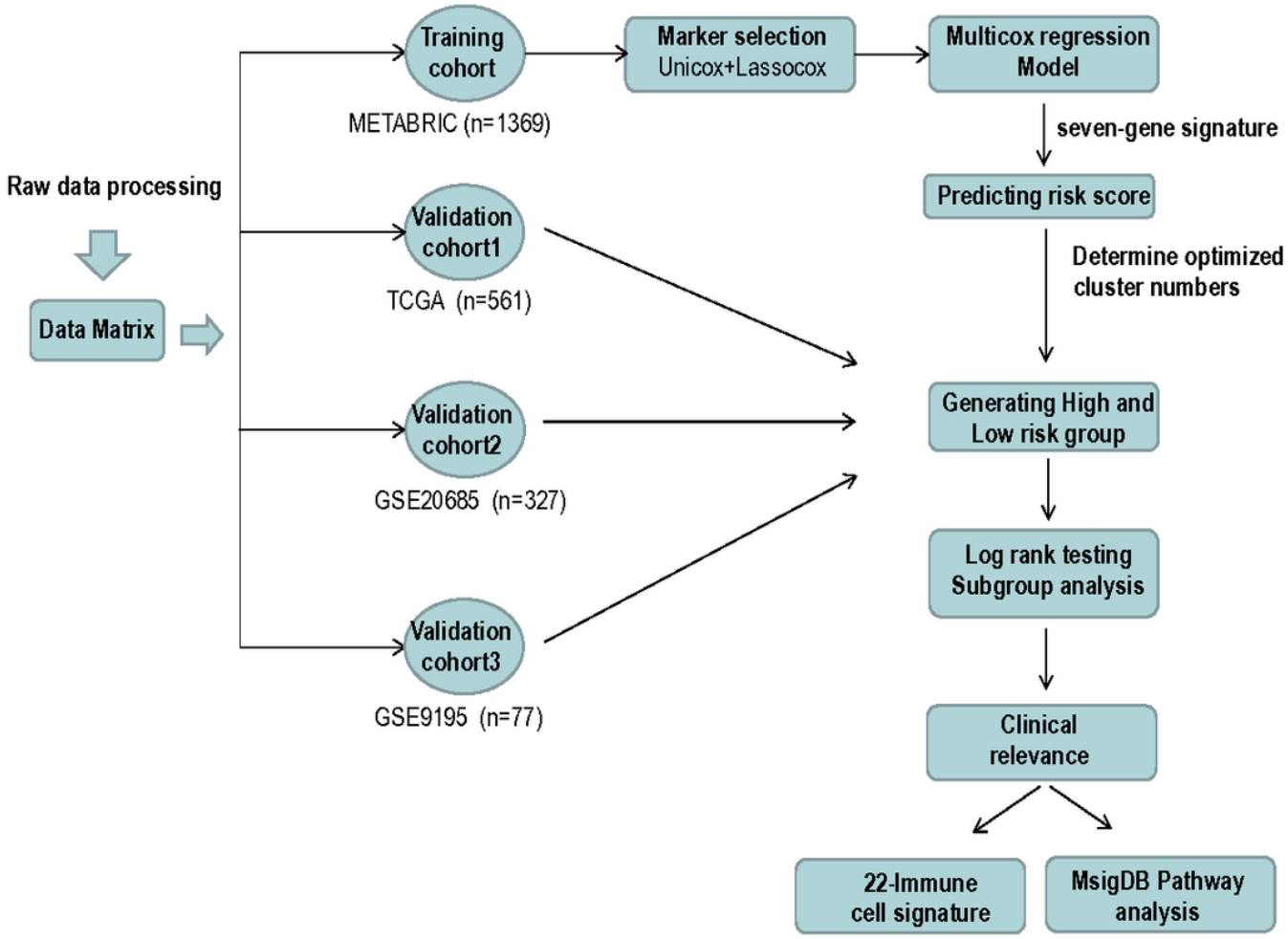

Figure 1

The study flow chart

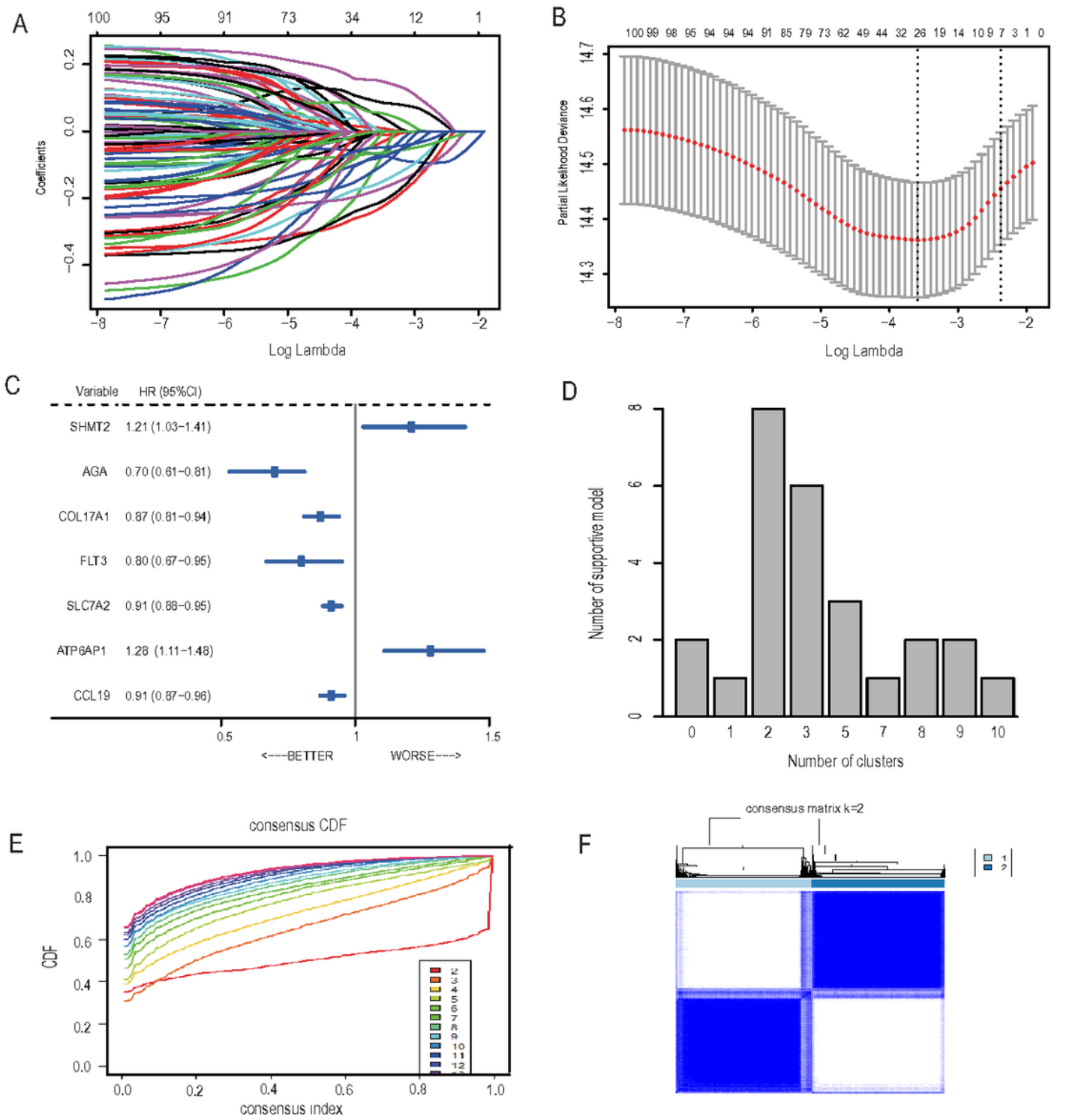

Figure 2

The construction of 7-gene risk signature and the determinant of the optimal number of group (A-B) The processes of LASSO Cox model fitting (C) The forest plot of HR value and confidence interval of the selected seven genes from the multivariate Cox regression analysis in METABRIC cohorts (D) Optimal number of groups determined by R package NbClust (E-F) Optimal number of groups determined by R package ConsensusClusterPlus

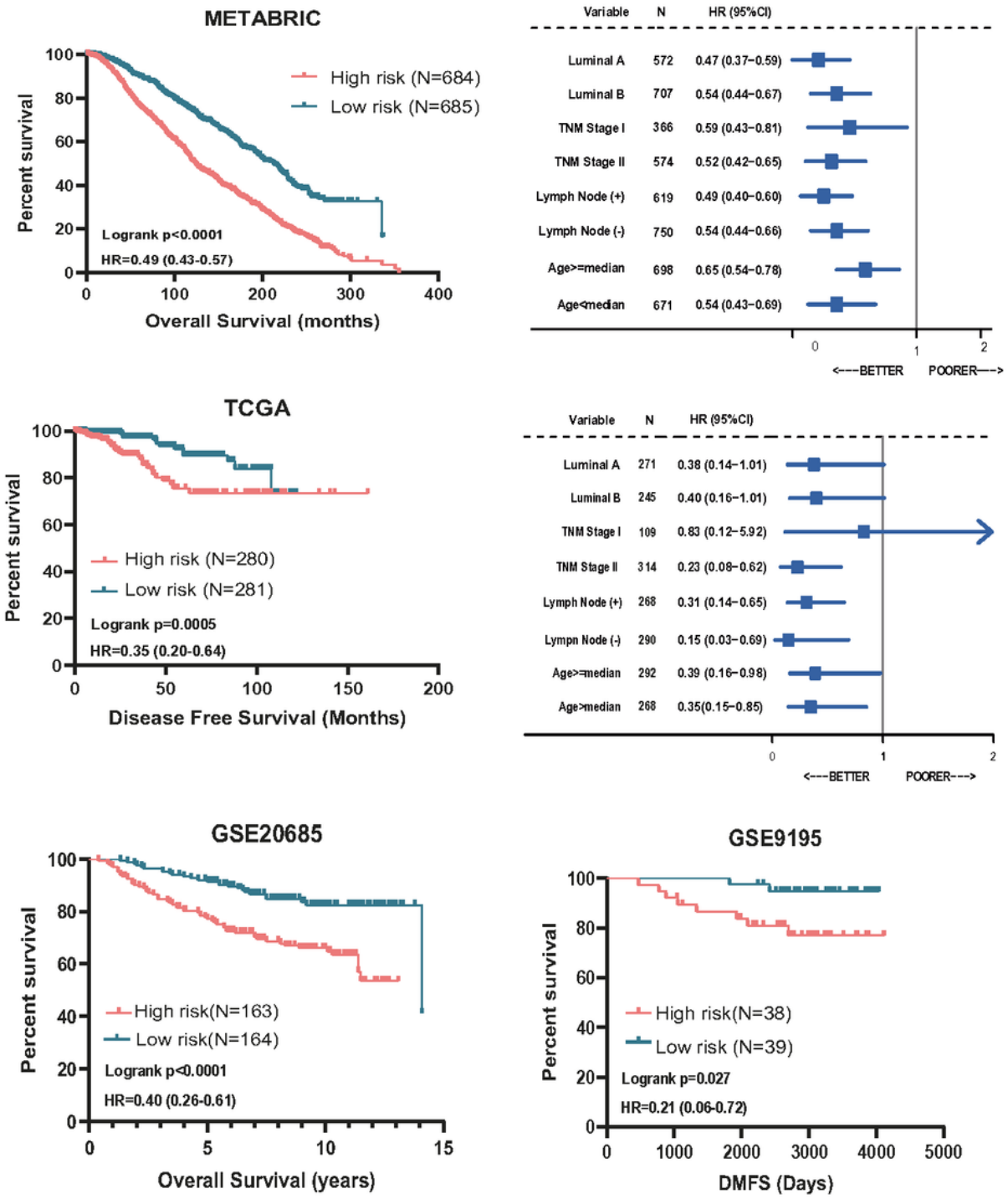

Figure 3

Validation of the prognostic significance of the risk group in different cohorts (A)Kaplan-Meier survival curve for patients in METABRIC cohort (B)The prognostic effect of the risk signature in different subgroups of METABRIC cohort (C)Kaplan-Meier survival curve for patients in TCGA cohort (D)The prognostic effect of the risk signature in different subgroups of TCGA cohort (E)Kaplan-Meier survival curve for patients in GSE20685 cohort (F)Kaplan-Meier survival curve for patients in GSE9195 cohort

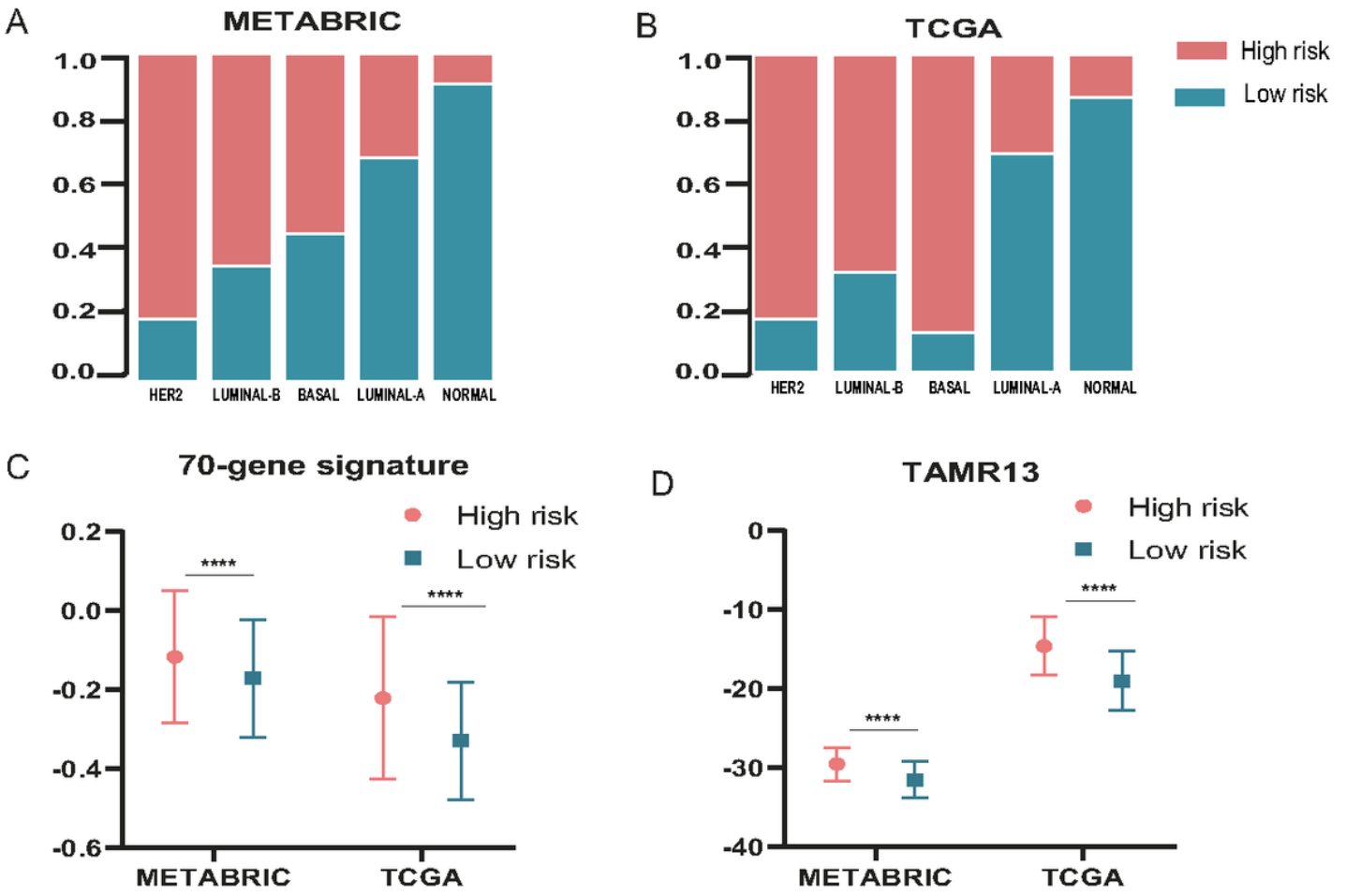

Figure 4

The association of 7-gene risk signature and clinical parameters (A) The distribution of high- and low-risk group by PAM50 subtypes in METABRIC cohort (B) The distribution of high- and low-risk group by PAM50 subtypes in TCGA cohort (C) The comparison of 70-gene score between high-and low-risk group in METABRIC and TCGA cohorts (****: p<0.0001) (D) The comparison of TAMR13 score between high-and low-risk group in METABRIC and TCGA cohorts (****: p<0.0001)

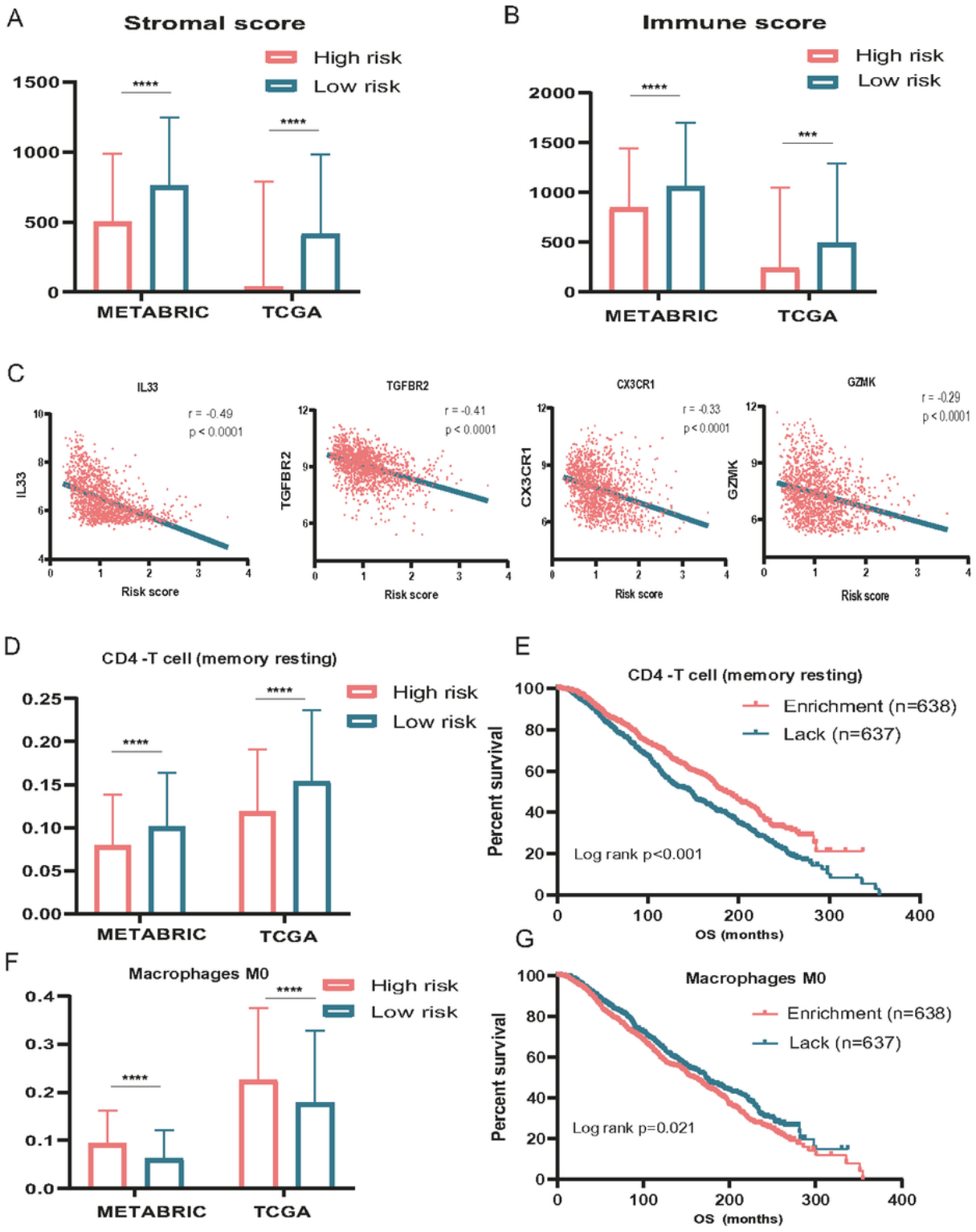

Figure 5

The association of 7-gene risk signature and tumor microenvironment (A) The comparison of stromal score between high-and low-risk group in METABRIC and TCGA cohorts (****: p<0.0001;***:p<0.001) (B) The comparison of immune score between high-and low-risk group in METABRIC and TCGA cohorts (****: p<0.0001;***:p<0.001) (C) Risk signature score was significantly correlated with immune genes (D) The comparison of CD4 memory resting T cell score between high-and low-risk group in METABRIC and TCGA

cohorts (****: p<0.0001;***:p<0.001) (E) Kaplan-Meier survival curve for patients with high or low CD4 memory resting T cell score in METABRIC cohort (F) The comparison of macrophage M0 cell score between high-and low-risk group in METABRIC and TCGA cohorts (****: p<0.0001;***:p<0.001) (G) Kaplan-Meier survival curve for patients with high or low macrophage M0 cell score in METABRIC cohort

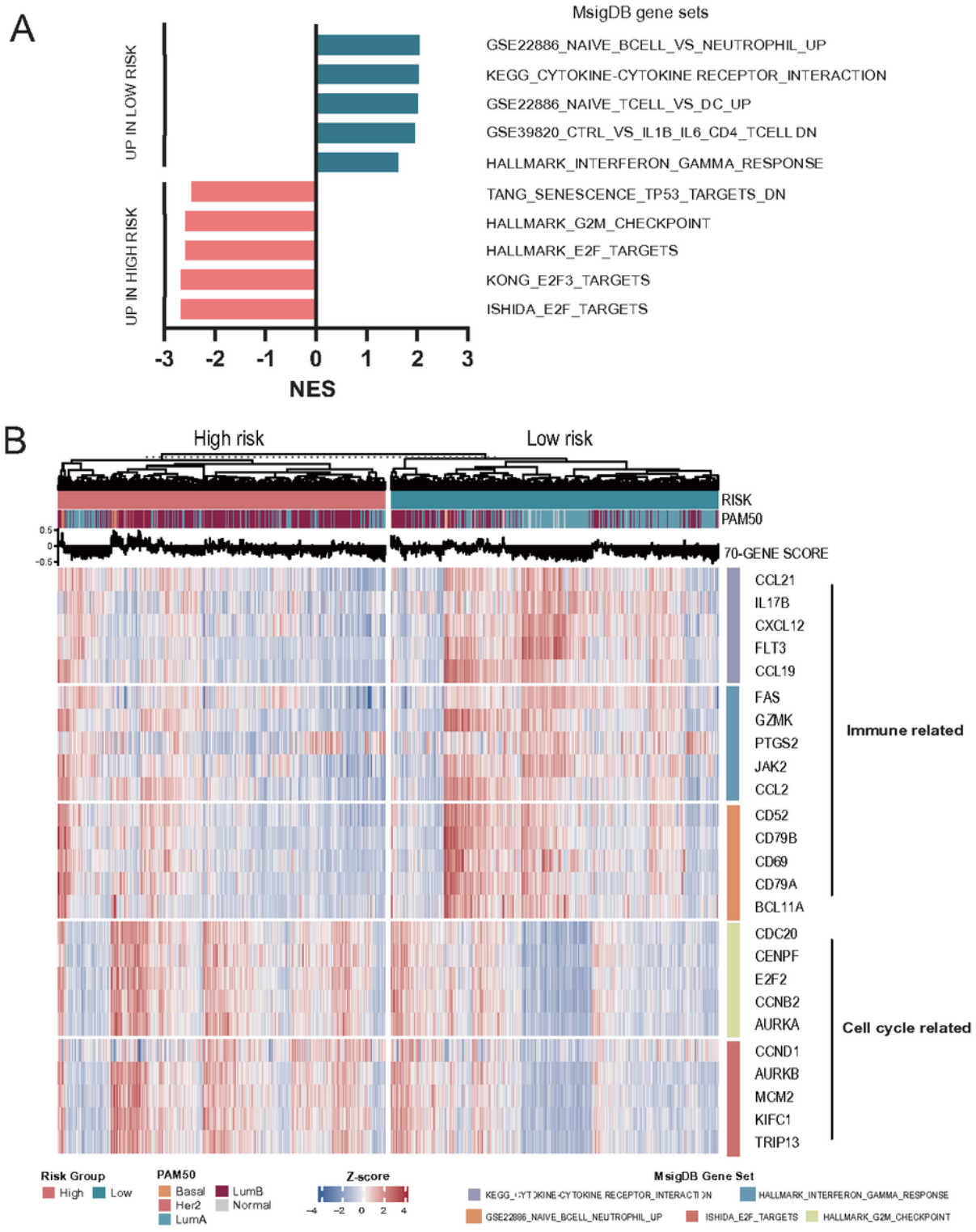

Figure 6

Risk signature is associated with immune and cell cycle related phenotypes in ER or PR positive and HER2 negative breast cancer (A) Genes with significantly higher expression in high or low risk groups were used in a gene set enrichment analysis using the collections of the MsigDB. The five most enriched processes in each collection are denoted. The cell cycle-related or immune-related pathways were significantly up-regulated in high or low risk group,respectively. (B) The heatmap displayed the selected DEGs in representative pathways.

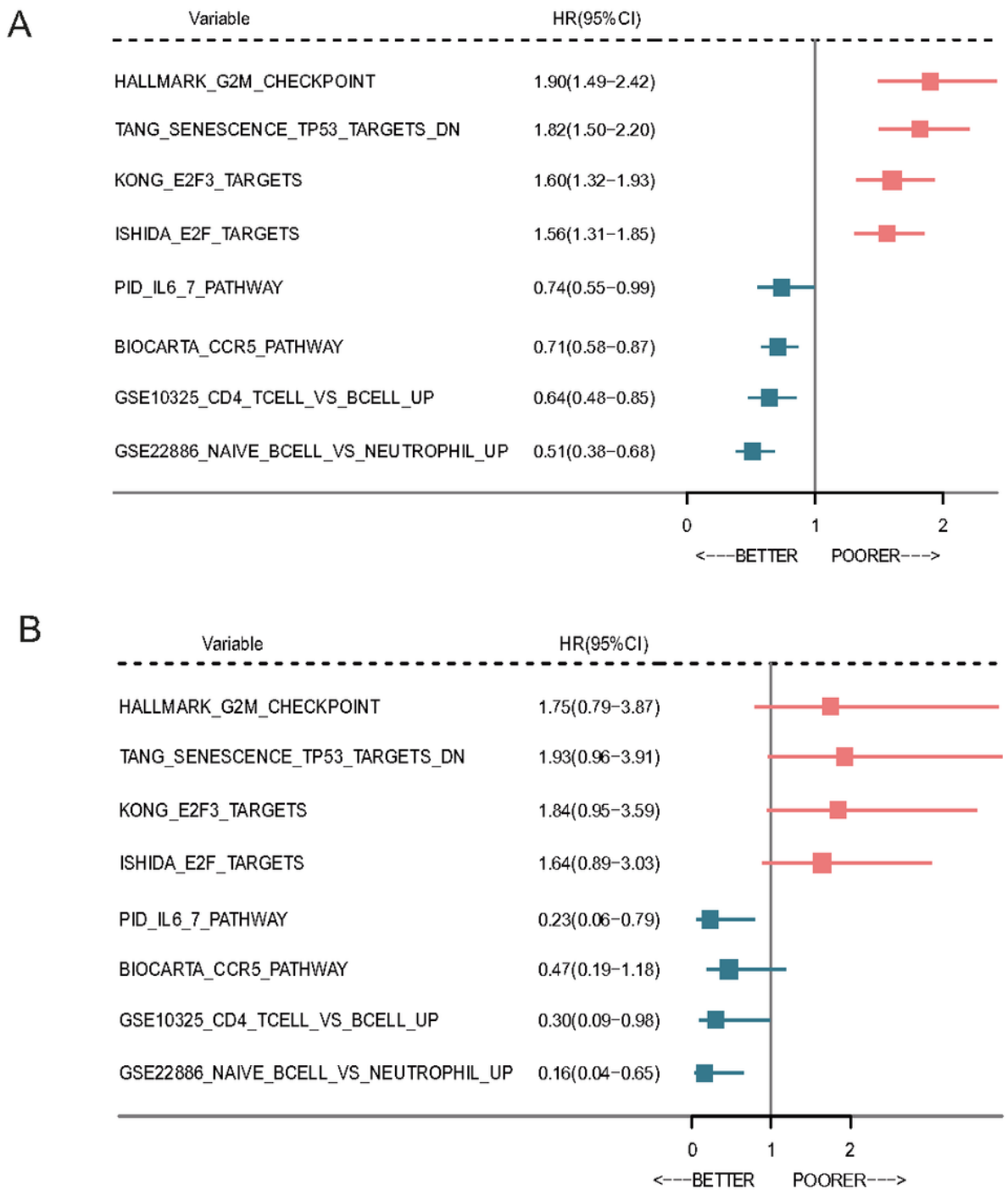

Figure 7

Estimates of univariate cox regression analysis and the 95% confidence interval are illustrated by forest plot to assess the prognostic effect of selected gene set signature scores calculated using GSVA in high risk versus low risk group in METABRIC (A) and TCGA (B) cohorts.