Estrogen receptor - positive breast cancer survival prediction and analysis of resistance - related genes introduction

Chen Shuai [1], Fengyan Yuan [2], Yu Liu [3], Chengchen Wang [3], Jiansong Wang [3], Hongye He [Corresp. 4]

[1] Department of Breast and Thyroid Surgery, Yiyang Central Hospital, Yiyang, Hunan, China

[2] Hunan Normal University of Medicine, Changsha, Hunan, China

[3] Hunan Provincial People's Hospital, Changsha, Hunan, China

[4] Department of General Surgery, The Second Xiangya Hospital, Central South University, Changsha, Hunan, China

Corresponding Author: Hongye He
Email address: hhyhongye@csu.edu.cn

Background: In recent years, ER+ and HER2- breast cancer of adjuvant therapy has made great progress, including chemotherapy and endocrine therapy. We found that the responsiveness of breast cancer treatment was related to the prognosis of patients. However, reliable prognostic signatures based on ER+ and HER2- breast cancer and drug resistance-related prognostic markers have not been well confirmed, This study in amied to establish a drug resistance-related gene signature for risk stratification in ER+ and HER2- breast cancer. **Methods:** We used the data from the Atlas Cancer Genome (TCGA) breast cancer dataset and gene expression database (Gene Expression Omnibus,GEO),constructed a risk profile based on four drug resistance-related genes, and developed a nomogram to predict the survival of patients with I-III ER+ and HER2- breast cancer. At the same time, we analyzed the relationship between immune infiltration and the expression of these 4 genes or risk groups. **Results:** 4 drug resistance genes:AMIGO2, LGALS3BP, SCUBE2 and WLS were promising tools for ER+ and HER2- breast cancer risk stratification. Then, The nomogram, which combines genetic characteristics with known risk factors, produces better performance and net benefits in calibration and decision curve analysis. Similar results were validated in three separate GEO cohorts. All of these results showed that the model can be used as a prognostic classifier for clinical decision-making, individual prediction and treatment, as well as follow-up.

1 **Estrogen receptor - positive breast cancer survival Prediction**

2 **and Analysis of Resistance - related genes**

3 **Introduction**

4 Shuai.Chen[1, 2], Fengyan Yuan[1], YuLiu[2] , Chengchen Wang[2], Jiansong Wang[2], Hongye He[2]

5 [1] Hunan Nomal university school of medicine, Changsha, Hunan, China

6 [2] Department of General Surgery, Hunan Provincial people's Hospital, Changsha, Hunan, China

8 Corresponding Author:

9 Hongye He[2]

10 Department of General Surgery, Hunan Provincial people's Hospital, Changsha, Hunan, China

11 Email address: qingyunshuimo@sina.com

13 **Background**. In recent years, ER+ and HER2- breast cancer of adjuvant therapy has made great

14 progress, including chemotherapy and endocrine therapy. We found that the responsiveness of

15 breast cancer treatment was related to the prognosis of patients. However, reliable prognostic

16 signatures based on ER+ and HER2- breast cancer and drug resistance-related prognostic markers

17 have not been well confirmed, This study in amied to establish a drug resistance-related gene

18 signature for risk stratification in ER+ and HER2- breast cancer.

19 **Methods**. We used the data from The Cancer Genome Atlas(TCGA) breast cancer dataset and

20 gene expression database (Gene Expression Omnibus,GEO),constructed a risk profile based on

21 four drug resistance-related genes, and developed a nomogram to predict the survival of patients

22 with I-III ER+ and HER2- breast cancer. At the same time, we analyzed the relationship between

23 immune infiltration and the expression of these 4 genes or risk groups.

24 **Results**. 4 drug resistance genes: *AMIGO2, LGALS3BP, SCUBE2* and *WLS* were promising tools for

25 ER+ and HER2- breast cancer risk stratification. Then, The nomogram, which combines genetic

26 characteristics with known risk factors, produces better performance and net benefits in

27 calibration and decision curve analysis. Similar results were validated in three separate GEO

28   cohorts. All of these results showed that the model can be used as a prognostic classifier for

29   clinical decision-making, individual prediction and treatment, as well as follow-up.

30   **Introduction**

31   Breast cancer is a common diagnosed malignant tumor that lead to cancer-related mortality

32   in women worldwide, It can mainly classified into invasive ductal carcinoma (IDC) and invasive

33   lobular carcinoma (ILC) by histopathological classification.(Dianatinasab et al. 2019; Li et al. 2005;

34   Viale 2012) Basic status of estrogen receptor (ER), progesterone receptor (PR), and Human

35   epidermal growth factor receptor 2 (HER2) play a crucial role in molecular subtype of breast

36   cancer, previous molecular subtype of breast cancer, previous molecular studies demonstrated

37   that breast cancer can be classified into Luminal A(ER+/PR+/HER2-/low Ki-67); Luminal B

38   (ER+/PR+/HER2-/+/high Ki-67); HER2-overexpression (ER-/PR-/HER2+) and triple negative breast

39   cancers(ER-/PR-/HER2-). (Al-Thoubaity 2020; Goldhirsch et al. 2011)Different molecular subtypes

40   exhibit different responsiveness to therapies. For instance, most patients of breast cancer are

41   present as ER+ and HER2- and can be benefit from endocrine therapy.

42   Although surgery-based comprehensive therapy can significantly improve the prognosis of

43   breast cancer, the endocrine resistance may occur in approximately 33% of patients with

44   endocrine therapy.(Achinger-Kawecka et al. 2020) When endocrine-resistant appear,

45   chemotherapy is needed to control the disease progression.(Zhu et al. 2018) Nevertheless, a

46   small subset of patients are resistant to both endocrine therapy and chemotherapy, so that lead

47   to poor prognosis. In clinical practice, tamoxifen is one of the first-line endocrine therapies that

48   competes with estrogen for ER and counteracts the estrogen signal pathway, which

49   retards breast cancer growth. (Swaby et al. 2007; Ye et al. 2019)Similarly, Doxorubicin is the most

50   widely used DNA-damaging chemotherapeutic drug.(Pfitzer et al. 2019) Many studies have been

51   investigated to explore the mechanism of drug resistance. Several pathways include

52   PI3K/Akt/mTOR pathway, fibroblast growth factor receptor pathway , etc. are activated to

53   reverse the growth suppression induced by tamoxifen and doxorubicin.(Guo et al. 2016; Li et al.

54 [2020a](); [Li et al. 2019](); [Lu et al. 2019](); [Shah et al. 2015](); [Turner et al. 2010]() In addition, Several

55 studies showed that tumor infiltrating lymphocytes (TILs) play a pivotal role in chemotherapy

56 response and associate with clinical outcome in all subtypes of breast cancer([Denkert et al. 2010]();

57 [Guo et al. 2016](); [Stanton & Disis 2016](); [Turner et al. 2010]())

58   Commonly, the clinicopathological risk factors, such as TNM stage and histological grade,

59 are usually used to assess the survival risk of breast cancer. However, as a result of molecular

60 differences, clinical prognoses are mainly different even in patients with histologically similar

61 tumors, which indicates that only use these risk factors as evaluation indicators cannot accurately

62 predict the prognoses of breast cancer patients.([Tang et al. 2019]()) Therefore, it is necessary to

63 identify novel prognostic biomarkers to improve the prognosis evaluation ability of breast cancer.

64   In this study, we used the data obtained from The Cancer Genome Atlas (TCGA) breast

65 cancer dataset and Gene Expression Omnibus (GEO) database, then built a four resistance-

66 related genes based risk signature and developed a nomogram to predict the survival of patients

67 with ER+ and HER2- breast cancer in stage I-III. Finally, we analyzed the association between

68 immune infiltrates and the four genes' expression or risk groups.

69 **Materials & Methods**

70 **data download**

71   The raw counts of the RNA sequence data of the TCGA-BRCA dataset   were downloaded

72 from the TCGA website ([https://portal.gdc.cancer.gov/). The]() clinical features and the disease-

73 free survival (DFS) status of these patients were downloaded from the cBioportal database

74 ([http://www.cbioportal.org). We select]()ed 473 ER+ and HER2- breast cancer patients in stage I-III

75 for the subsequent analysis (Table 1). The R software (version 3. 6. 3) was used to standardize

76 and process data.

77   GSE21653 (100 samples of ER+ and HER2- early invasive breast cancer), GSE17705 (298

78 samples of ER+ breast cancer) and GSE11121(200 patients) datasets were acquired from GEO

79 database (https://www.ncbi.nlm.nih.gov/geo/) as the validation cohorts. GSE21653 has DFS as

80   the endpoint, GSE17705 has relapse-free survival (RFS) as the endpoint and GSE11121 has distant

81   metastases-free survival (DMFS) as the endpoint.

Screening of differentially expressed genes

83   The resistance-related differentially expressed genes (DEGs) were detected with the

84   "limma" package in R software from the GSE67916 (DEGs between MCF7 tamoxifen resistant

85   (MCF7/TamR) cell lines and MCF7 parental cell lines) and GSE24460 (DEGs between MCF7

86   doxorubicin resistant (MCF/ADR) cell lines and MCF7 parental cell lines). Defined genes with

87   adjusted p values <0. 05 and |log2foldchange| values >1. 5 as DEGs. The genes both up-regulated

88   (16 genes) and down-regulated (21 genes) in the two GEO datasets were finally screened (Fig.

89   S1b). Volcano plots and heatmaps were performed using the R packages "ggpubr", "ggthemes"

90   and "pheatmap" to exhibit the differentially expressed genes (Fig. S1a,1c).

Identification and validation of a risk signature associated with breast cancer

92   The data downloaded from TCGA datasets were used as the training cohort to analyze the

93   correlation of genes' expression with the DFS time based on univariate Cox regression analysis.

94   Patients without survival for less than 0 days were removed to identify prognostic genes, only

95   the genes with the p values < 0. 05 were selected as candidates (Fig. 2a). Then, the candidate

96   genes were selected for stepwise multivariate Cox regression analysis (Fig. 2b). The Akaike

97   information criterion (AIC) was used to avoid over-fitting, we chosed the best-fit predictive model

98   with the lowest AIC (AIC=359. 44). Finally, a four-gene risk signature was established and

99   calculated their risk coefficients use the formula: The riskScore = (-0. 175*expression of AMIGO2)

100 + (-0. 210* expression of LGALS3BP) + (-0. 106* expression of SCUBE2) + (-0. 293* expression of

101 WLS). Using the optimal riskScore cut-off value (-10.09) calculated by the R package "survminer",

102 we divided patients in both the training cohort and the validation cohort into high-risk and low-

103 risk group. Then, Kaplan-Meier curve and log-rank test were used to analyze the different survival

104 time between high-risk group and low-risk group. Subsequently, a time-dependent receiver

105   operating characteristic (ROC) curve was calculated to compare the sensitivity and specificity of

106   risk factors.

107 Development of the prognostic nomogram

108   To identify independent predictors of DFS, we tested clinical risk factors and the gene-based

109   riskScore. Then we used the R package "rms" to establish the prognostic nomogram. The

110   discrimination and calibration abilities of the nomogram were assessed with a concordance index

111   (C-index) and calibration plot. For further evaluation of the predictive performance, we also used

112   the time-dependent ROC curves.

113 Function and pathway enrichment analyses

114   The R package "clusterProfiler" was used to classify the 37 DEGs' GO terms or KEGG

115   pathways. Functional enrichment analyses were carried out for GO terms and KEGG pathways

116   through the significance threshold of p value<0. 05.

117 Immune infiltration analyses

118   The 22 kinds of immune cell infiltration levels of TCGA-BRCA dataset were quantified by

119   CIBERSORT algorithm with 1,000 permutations (http://cibersort.stanford.edu/).(Chen et al.

120   2018a) CIBERSORT calculates a p value to confirm the accuracy of the results, only the samples

121   with the p value <0.05 were selected for analyzing the immune infiltration levels between the

122   high- and low- risk groups. Moreover, to explore the correlation between the four genes'

123   expression and the abundance of immune cell infiltrates (B cells, CD4+ T cells, CD8+ T cells,

124   neutrophils, macrophages, and dendritic cells), the Tumor Immune Estimation Resource (TIMER,

125   https://cistrome.shinyapps.io/timer/) was utilized.

126 Statistical Analyses

127   All the statistical analyses were performed with R software (version 3. 6. 3). All the statistical

128   analyses included univariate and multivariate Cox regression analysis, ROC curve analysis, Kaplan-

129   Meier survival analysis and correlation analysis. The correlation analysis was based on the R

130   package "corrplot". Statistical differences were compared by Wilcoxon test between two groups

131   and Kruskal-Wallis H test for multigroup comparison. P<0. 05 was considered statistically

132   significant.

Results

Screening DFS-related resistance-related DEGs

135   The flow chart was shown in Fig 1, To screen the DFS-related resistance-related DEGs, two

136   datasets was downloaded from GEO database. After "limma" filtering between MCF7

137   doxorubicin or tamoxifen resistant cell lines and MCF parental cell lines (Fig. S1a), the intersection

138   of up-regulated and down-regulated genes was taken respectively (Fig. S1b). A total of 37

139   resistance-related DEGs were obtained. The expression profile of the 37 DEGs in GSE24460 and

140   GSE67916 were shown in Fig. S1c respectively.

141   Next, we investigated whether genes could be used as prognostic candidates in ER positive

142   (ER+) and HER2 negative (HER2-) breast cancer patients in stage I-III. The univariate Cox

143   regression analysis found that eight genes were associated with DFS time in ER+ and HER2- breast

144   cancer, all of these genes were protective factors (Fig. 2a). The eight candidates were further

145   analyzed by multivariate Cox regression analysis and examined four genes (*AMIGO2, LGALS3BP,*

146   *SCUBE2* and *WLS*) as the prognostic candidates (Table 2, Fig. 2b). In addition, we verified the

147   prognosis of these four genes in the Kaplan-Meier Plotter website

148   (https://kmplot.com/analysis/). Consistent with our results, *WLS* and *SCUBE2* were protective

149   factors, high expression of the two genes had longer RFS time (Fig. S2c). The results of other two

150   genes were not statistically significant. A risk score formula (the formula shown in methods)

151   derived from the expression levels of the four prognostic candidates was constructed by the

152   multivariate Cox regression analysis.

Identification and validation of a four-gene based risk signature

154   All enrolled patients were divided into high- and low-risk groups with its optimal cut-off

155   value among the training cohort by using the riskScore. The survival status plot and heatmap for

156   the expression of the four genes were showed in Fig. 2c-d. The prognostic capacity of the four-

157   gene based riskScore was estimated by calculating the AUC of a time-dependent ROC curve (Fig.

158   3a). The AUCs of the four-gene risk signature were 0.686 and 0.714 for 3- and 5-year DFS

159   respectively, suggesting that the risk signature had certain sensitivity and specificity. Kaplan-

160   Meier curve showed that patients in the high-risk group had a lower DFS time than those in the

161   low-risk group (Fig. 3a).

162   We also analyzed the correlation of the four genes in ER+ and HER2- breast cancer patients

163   in TCGA-BRCA dataset. The correlation heatmap showed that the maximum R values was 0.14,

164   indicating that these four genes were independent of each other (Fig. 2f).

165   To validate the prognostic prediction performance of riskScore, the same as training cohort,

166   patients in the validation cohorts were classified into high- and low-risk groups using the same

167   formula obtained from the training cohort. In all validation cohorts, patients with a high riskScore

168   had a shorter survival time and a high riskScore was associated with worse prognosis (Fig. 3b-d).

169   Additionally, the time-dependent ROC analyses for the survival prediction of the risk signature

170   showed AUCs of 0. 624 at 3 years, 0. 613 at 5 years in GSE21653 and 0. 618 at 3 years, 0. 616 at

171   5 years in GSE17705, demonstrating that this risk signature was capable of predicting survival in

172   ER+ and HER2- or ER+ and HER2 equivocal patients (Fig. 3b-c). Moreover, we also used the time-

173   dependent ROC analysis to explore the prognostic prediction performance if the subtype of

174   breast cancer was unknown. The AUCs were 0.730 at 3 years and 0.669 at 5 years in

175   GSE11121(Fig. 3d), suggesting that our risk signature also has good predictive ability when breast

176   cancer subtype is unknown.

177   **Stratified analyses with clinical factors**

178   As shown in Fig 4, the four-gene based risk signature could serve as a promising candidate

179   for predicting the survival of breast cancer in different subtypes stratified by the clinical factors,

180   including N0-1 type (p=0. 0014), N2-3 (p=0. 026), T1-2 (p=0. 0006) T3-4(p=0. 04), stage I-II (p=0.

181   0022), stage III (p= 0. 0078), PR positive (PR+, p<0. 0001) patients, respectively. But the PR

182   negative (PR-) subtype was not statistically significant.

183   As shown in Fig 5, the riskScore value was higher expression in PR- subtype than in PR+

184   subtype, *SCUBE2* and *WLS* were higher expression in PR+ subtype than in PR-subtype, the

185   expression of *SCUBE2* was negatively correlated with T and Stage, these results were consistent

186   with the clinical prognostic significance.

187   **Functional enrichment analyses**

188   To explain the function of the resistance-related DEGs, we performed enrichment analyses

189   of the GO terms and KEGG pathways (Fig. S2a-b).

190   GO analyses showed that these resistance-related DEGs were enriched in several biological

191   processes (BP), including Antigen processing and presentation of peptide antigen via MHC class

192   I, Antigen processing and presentation of exogenous peptide antigen via MHC class I and

193   Epithelial structure maintenance (Fig. S2a).

194   KEGG analysis showed that the resistance-related DEGs were related to Endocytosis, Cellular

195   senescence, Antigen processing and presentation and Adherens junction (Fig. S2b).

196   **Revealing the correlation between the risk signature and tumor-infiltrating immune cells**

197   Because TILs play a key role in chemotherapy response and our enrichment analyses results

198   showed that the DEGs associated with antigen presentation, first, we evaluated the abundance

199   of the 22 kinds of immune cells in each TCGA-BRCA patient (Fig. 6a). Next, we utilized the

200   Wilcoxon test to compare the difference immune cell infiltration levels between the high- and

201   low groups and found that B cells memory (p=0.042) and T cells follicular helper (p=0.016) were

202   high infiltrated in the high-risk group, and B cells naive (p=0.047) and mast cells resting (p=0.042)

203   were low infiltrated in the high-risk group (Fig. 6b). Unfortunately, due to the limitation of sample

204   size, the relationship between the immune cell infiltration levels and survival is unknown.

205   Whereafter, we assessed the correlation between the four resistance-related DEGs and six

206   infiltrating immune cells by using the TIMER database. The results revealed that the genes had

207   good correlation with immune cells. (Fig. S3)

208   **Building a prognostic nomogram combining riskScore with clinical risk factors**

209   Nomograms are widely used in predicting cancer patients' prognoses, principally because

210   they have the ability to decrease the statistical prediction models into a single numerical

211   assessment, which is tailored to the profile of an individual patient.(Iasonos et al. 2008; Long et

212   al. 2018)

213   In this article, we combined the four-gene based riskScore with clinical risk factors to

214   construct a nomogram to assess the probability of 3- and 5- year DFS for patients with ER+ and

215   HER2- early stage invasive breast cancer in TCGA-BRCA dataset. After univariate and multivariate

216   Cox regression analyses filtering, the riskScore, T and Stage remained in the final Cox model for

217   DFS (Fig. 7a-b). A nomogram was established based on the final Cox model for DFS to evaluate of

218   the 3- or 5-year DFS (Fig. 7c). The clinical stage and riskScore of the patient can be matched with

219   this Nomogram to obtain the total points and estimate the survival rate of the patient. The

220   nomogram achieved a C-index of 0. 740 (95% CI, 0. 653 - 0. 826), and the calibration plots

221   demonstrated that the nomogram has good predictive power (Fig. 7d). Time-dependent ROC

222   analyses also indicated that the nomogram was accurate in prediction of DFS compared to use

223   riskScore or clinical risk factors alone (Fig. 7e). In conclusion, all the results suggest the clinical

224   practical use of the nomogram for prediction of DFS.

225   **Discussion**

226   Breast cancer is highly heterogeneous in molecular and cellular, the morbidity of it increases

227   every year. In recent years, molecular biomarkers of breast cancer become more and more

228   important. Understanding the biological functions of breast cancer may offer new methods that

229   can be used to predict and treat the disease. However, there are still few specific biomarkers with

230   significant prognostic to predict survival of breast cancer. Therefore, in order to improve the

231   prognosis of breast cancer, molecular screening of biomarkers is urgently needed.

232   In this study, we developed and validated a novel risk signature based on four resistance-

233   related genes to predict patients with stage I-III ER+ and HER2- breast cancer. This article showed

234   that our risk signature could effectively distinguish the breast cancer patients between low and

235   high risk of survival groups. In addition, the four-gene based model was demonstrate to be

236   independent from other clinicopathological features. Then, a comprehensive nomogram was

237   constructed to predict the individualized DFS, which performed better than clinical factors and

238   riskScore to classify the patients' survival risk. Finally, we explored the correlation between the

239   four resistance-related DEGs and immune infiltrates.

240   Previous investigation proved that protein-based signature could predict tamoxifen

241   treatment outcome in recurrent breast cancer and Long Noncoding RNA based signature could

242   predict recurrence among ER+ breast cancer patients treated with tamoxifen. (De Marchi et al.

243   2016; Wang et al. 2018)In reality, the expression profile-based prognostic biomarkers signature

244   has been already investigated in several types of cancer, including glioblastoma multiforme,

245   colon cancer, hepatocellular carcinoma, gastric cancer and non-small cell lung cancer. (Long et

246   al. 2018; Qiu et al. 2020; Wang et al. 2020; Zuo et al. 2019)But there is no research on the

247   application of predictive models of resistance-related genes in ER+ and HER2- breast cancer.

248   There are four genes involved in building the risk signature in this study, they are *AMIGO2,*

249   *LGALS3BP, SCUBE2* and *WLS*. All of these genes are protective markers in our study, there are no

250   previous studies demonstrated their role in drug resistance. These genes paly different roles in

251   different diseases.

252   Adhesion molecule with IgG-like domain 2 (*AMIGO2*) is a multifunctional amphoterin-

253   induced genes.(Kuja-Panula et al. 2003) *AMIGO2* controls cell survival and angiogenesis in

254   endothelial cells via Akt activation, knockout of *AMIGO2* can reduce cell survival and induce

255   apoptosis.(Park et al. 2015) *AMIGO2* also can promote liver metastasis by regulating the tumor

256   cell adhesion to liver endothelial cells and it function as a pro-proliferation factor in human

257   melanoma in vitro and in vivo. (Fontanals-Cirera et al. 2017; Kanda et al. 2017)

258   Lectin galactoside-binding soluble 3 binding protein (*LGALS3BP*) is a large oligomeric

259   glycoprotein, is overexpressed in many cancers and always associated with cancer growth and

260   migration.(Woo et al. 2017) *LGALS3BP* can induce vascular endothelial growth factor in human

261 MDA-MB-231 cells and promotes angiogenesis, it also act as a E-selectin ligand that associated

262 with metastasis and poor prognosis in ER- breast cancer. (Piccolo et al. 2013; Traini et al. 2014)But

263 in colon cancer, different from previous studies, *LGALS3BP* suppresses tumor proliferation in

264 colon cancer cells, high expression of *LGALS3BP* have a good prognosis. (Piccolo et al. 2015)There

265 are no research to explore the function of *LGALS3BP* in ER+ and HER2- invasive breast cancer.

266   The findings of Signal Peptide-CUB-EGF Domain-Containing Protein 2(*SCUBE2*) are

267 controversial. Jia-Hong Chen et. al proved that upregulated *SCUBE2* expression in breast cancer

268 stem cells can promote triple negative breast cancer aggression.(Chen et al. 2018b) But Lin, Y.

269 Cet. al demonstrated that *SCUBE2* inhibited MDA-MB-231 cell proliferation. (Lin et al. 2014)In

270 MCF-7 breast cancer cell line, overexpression of *SCUBE2* protein suppresses proliferation, this

271 result is consistent with our study. (Cheng et al. 2009)*SCUBE2* also performed as protective factor

272 in glioma, non-small cell lung cancer and colorectal cancer and inhibited the proliferation and

273 migration. (Guo et al. 2017; Song et al. 2015; Yang et al. 2018)

274   The Wnt secretion protein Wntless (*WLS*) has been proved to be involved in the

275 development of several human cancers. *WLS* mostly plays the role of oncogene in most tumors.

276 In triple negative breast cancer, glioma and colon cancer, *WLS* promotes the cell proliferation.

277 (Augustin et al. 2012; Voloshanenko et al. 2013)In lung adenocarcinoma cell lines, WLS is

278 upregulated by small nucleolar RNA host gene 17 (*SNHG17*), the up-regulation of WLS could

279 reverse the inhibition of proliferation and promote apoptosis effects of *SNHG17* down-

280 regulation. (Li et al. 2020b)But in melanoma cell, *WLS* inhibits proliferation through the β-catenin

281 signaling pathway and reduces spontaneous metastasis. (Yang et al. 2012)

282   Tumor-infiltrating immune cells in the TME play a crucial role in antitumor and

283 chemotherapeutic effects.(Denkert et al. 2010; Stanton & Disis 2016) Therefore, we used

284 CIBERSORT algorithm and TIMER database to detect the relationship between risk signature and

285 immune cells. The results proved that B cells memory and T cells follicular helper (Tfh) were

286 positive correlated infiltrated in the high-risk group, but B cells naive and mast cells resting were

287   negative correlated infiltrated in the high-risk group. Tfh cells are crucial for maintenance of

288   humoral memory that assist B cells in producing germinal centers. ([Crotty 2019](#))Mast cells may

289   play protumoral or antitumoral roles in different tumor types, for instance, it could inhibit

290   prostate adenocarcinoma development but promote the tumorigenesis of highly malignant

291   neuroendocrine cancers.([Pittoni et al. 2011](#)) But these cells' function in tumor microenvironment

292   are undiscovered.

293   This study also has some Limitations. This is a retrospective research and is susceptible to

294   inherent biases. First, the study has potential selection bias and limited sample size. Then the

295   riskScore and nomogram cannot guide treatment strategies of breast cancer, so that further

296   verification and prospective trials are needed. Moreover, all of the data were obtained from the

297   database, the population ethnicities in the TCGA database are mainly composed of White people

298   and Black people, therefore the findings to other ethnicities still need proof, and some other

299   clinicopathological features including histological Grade, are not contained in the nomogram.

300   Because of the lack of clinical data in database, the nomogram has not been verified, so further

301   researchs are needed to verify the nomogram. Finally, the precise mechanism of resistance-

302   related genes involved in the risk signature need further in vitro and in vivo validation

303   experiment.

304   **Conclusions**

305   In this study, We used the data from TCGA-BRCA dataset and GEO database, constructed a risk

306   signature based on four drug resistance-related genes and developed a nomogram to predict

307   the survival of patients with stage I-III in ER+ and HER2- breast cancer. The results were

308   validated in three GEO cohorts. All the results indicated that this risk signature might serve as

309   novel molecular biomarkers for predicting breast cancer survival and provide guidance for

310   therapeutic strategies.

311

References

313 Achinger-Kawecka J, Valdes-Mora F, Luu PL, Giles KA, Caldon CE, Qu W, Nair S, Soto S, Locke WJ,
314   Yeo-Teh NS, Gould CM, Du Q, Smith GC, Ramos IR, Fernandez KF, Hoon DS, Gee JMW,
315   Stirzaker C, and Clark SJ. 2020. Epigenetic reprogramming at estrogen-receptor binding
316   sites alters 3D chromatin landscape in endocrine-resistant breast cancer. *Nat Commun*
317   11:320. 10.1038/s41467-019-14098-x
318 Al-Thoubaity FK. 2020. Molecular classification of breast cancer: A retrospective cohort study.
319   *Ann Med Surg (Lond)* 49:44-48. 10.1016/j.amsu.2019.11.021
320 Augustin I, Goidts V, Bongers A, Kerr G, Vollert G, Radlwimmer B, Hartmann C, Herold-Mende C,
321   Reifenberger G, von Deimling A, and Boutros M. 2012. The Wnt secretion protein
322   Evi/Gpr177 promotes glioma tumourigenesis. *EMBO Mol Med* 4:38-51.
323   10.1002/emmm.201100186
324 Chen B, Khodadoust MS, Liu CL, Newman AM, and Alizadeh AA. 2018a. Profiling Tumor Infiltrating
325   Immune Cells with CIBERSORT. *Methods Mol Biol* 1711:243-259. 10.1007/978-1-4939-
326   7493-1_12
327 Chen JH, Kuo KT, Bamodu OA, Lin YC, Yang RB, Yeh CT, and Chao TY. 2018b. Upregulated SCUBE2
328   expression in breast cancer stem cells enhances triple negative breast cancer aggression
329   through modulation of notch signaling and epithelial-to-mesenchymal transition. *Exp Cell*
330   *Res* 370:444-453. 10.1016/j.yexcr.2018.07.008
331 Cheng CJ, Lin YC, Tsai MT, Chen CS, Hsieh MC, Chen CL, and Yang RB. 2009. SCUBE2 suppresses
332   breast tumor cell proliferation and confers a favorable prognosis in invasive breast cancer.
333   *Cancer Res* 69:3634-3641. 10.1158/0008-5472.CAN-08-3615
334 Crotty S. 2019. T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. *Immunity*
335   50:1132-1148. 10.1016/j.immuni.2019.04.011
336 De Marchi T, Liu NQ, Stingl C, Timmermans MA, Smid M, Look MP, Tjoa M, Braakman RB, Opdam
337   M, Linn SC, Sweep FC, Span PN, Kliffen M, Luider TM, Foekens JA, Martens JW, and Umar
338   A. 2016. 4-protein signature predicting tamoxifen treatment outcome in recurrent breast
339   cancer. *Mol Oncol* 10:24-39. 10.1016/j.molonc.2015.07.004
340 Denkert C, Loibl S, Noske A, Roller M, Muller BM, Komor M, Budczies J, Darb-Esfahani S,
341   Kronenwett R, Hanusch C, von Torne C, Weichert W, Engels K, Solbach C, Schrader I, Dietel
342   M, and von Minckwitz G. 2010. Tumor-associated lymphocytes as an independent
343   predictor of response to neoadjuvant chemotherapy in breast cancer. *J Clin Oncol* 28:105-
344   113. 10.1200/JCO.2009.23.7370
345 Dianatinasab M, Fararouei M, Daneshi N, Rezaian S, Mohammadianpanah M, Chaman R, and
346   Ghiasvand R. 2019. Heterogeneity in risk factors for ductal and lobular breast carcinomas:
347   A case-control study. *Int J Cancer* 145:2917-2925. 10.1002/ijc.32182

348 Fontanals-Cirera B, Hasson D, Vardabasso C, Di Micco R, Agrawal P, Chowdhury A, Gantz M, de
349   Pablos-Aragoneses A, Morgenstern A, Wu P, Filipescu D, Valle-Garcia D, Darvishian F, Roe
350   JS, Davies MA, Vakoc CR, Hernando E, and Bernstein E. 2017. Harnessing BET Inhibitor
351   Sensitivity Reveals as a Melanoma Survival Gene. *Mol Cell* 68:731-744 e739.
352   10.1016/j.molcel.2017.11.004

353 Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ, and Panel m. 2011.
354   Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St.
355   Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer
356   2011. *Ann Oncol* 22:1736-1747. 10.1093/annonc/mdr304

357 Guo E, Liu H, and Liu X. 2017. Overexpression of SCUBE2 Inhibits Proliferation, Migration, and
358   Invasion in Glioma Cells. *Oncol Res* 25:437-444. 10.3727/096504016X14747335734344

359 Guo Y, Ding Y, Zhang T, and An H. 2016. Sinapine reverses multi-drug resistance in MCF-7/dox
360   cancer cells by downregulating FGFR4/FRS2alpha-ERK1/2 pathway-mediated NF-kappaB
361   activation. *Phytomedicine* 23:267-273. 10.1016/j.phymed.2015.12.017

362 Iasonos A, Schrag D, Raj GV, and Panageas KS. 2008. How to build and interpret a nomogram for
363   cancer prognosis. *J Clin Oncol* 26:1364-1370. 10.1200/JCO.2007.12.9791

364 Kanda Y, Osaki M, Onuma K, Sonoda A, Kobayashi M, Hamada J, Nicolson GL, Ochiya T, and Okada
365   F. 2017. Amigo2-upregulation in Tumour Cells Facilitates Their Attachment to Liver
366   Endothelial Cells Resulting in Liver Metastases. *Sci Rep* 7:43567. 10.1038/srep43567

367 Kuja-Panula J, Kiiltomaki M, Yamashiro T, Rouhiainen A, and Rauvala H. 2003. AMIGO, a
368   transmembrane protein implicated in axon tract development, defines a novel protein
369   family with leucine-rich repeats. *J Cell Biol* 160:963-973. 10.1083/jcb.200209074

370 Li CI, Uribe DJ, and Daling JR. 2005. Clinical characteristics of different histologic types of breast
371   cancer. *Br J Cancer* 93:1046-1052. 10.1038/sj.bjc.6602787

372 Li D, Ji H, Niu X, Yin L, Wang Y, Gu Y, Wang J, Zhou X, Zhang H, and Zhang Q. 2020a. Tumor-
373   associated macrophages secrete CC-chemokine ligand 2 and induce tamoxifen resistance
374   by activating PI3K/Akt/mTOR in breast cancer %J Cancer Sci. *Cancer Sci* 111:47-58.
375   10.1111/cas.14230 %/ (c) 2019 The Authors. Cancer Science published by John Wiley &
376   Sons Australia, Ltd on behalf of Japanese Cancer Association.

377 Li W, Zheng Y, Mao B, Wang F, Zhong Y, and Cheng D. 2020b. SNHG17 upregulates WLS expression
378   to accelerate lung adenocarcinoma progression by sponging miR-485-5p. *Biochem*
379   *Biophys Res Commun* 533:1435-1441. 10.1016/j.bbrc.2020.09.130

380 Li Z, Qian J, Li J, and Zhu C. 2019. Knockdown of lncRNA-HOTAIR downregulates the drug-
381   resistance of breast cancer cells to doxorubicin via the PI3K/AKT/mTOR signaling
382   pathway. *Exp Ther Med* 18:435-442. 10.3892/etm.2019.7629

383 Lin YC, Lee YC, Li LH, Cheng CJ, and Yang RB. 2014. Tumor suppressor SCUBE2 inhibits breast-
384   cancer cell migration and invasion through the reversal of epithelial-mesenchymal
385   transition. *J Cell Sci* 127:85-100. 10.1242/jcs.132779

Long J, Zhang L, Wan X, Lin J, Bai Y, Xu W, Xiong J, and Zhao H. 2018. A four-gene-based prognostic model predicts overall survival in patients with hepatocellular carcinoma. *J Cell Mol Med* 22:5928-5938. 10.1111/jcmm.13863

Lu PW, Li L, Wang F, and Gu YT. 2019. Inhibitory role of large intergenic noncoding RNA-ROR on tamoxifen resistance in the endocrine therapy of breast cancer by regulating the PI3K/Akt/mTOR signaling pathway. *J Cell Physiol* 234:1904-1912. 10.1002/jcp.27066

Park H, Lee S, Shrestha P, Kim J, Park JA, Ko Y, Ban YH, Park DY, Ha SJ, Koh GY, Hong VS, Mochizuki N, Kim YM, Lee W, and Kwon YG. 2015. AMIGO2, a novel membrane anchor of PDK1, controls cell survival and angiogenesis via Akt activation. *J Cell Biol* 211:619-637. 10.1083/jcb.201503113

Pfitzer L, Moser C, Gegenfurtner F, Arner A, Foerster F, Atzberger C, Zisis T, Kubisch-Dohmen R, Busse J, Smith R, Timinszky G, Kalinina OV, Muller R, Wagner E, Vollmar AM, and Zahler S. 2019. Targeting actin inhibits repair of doxorubicin-induced DNA damage: a novel therapeutic approach for combination therapy. *Cell Death Dis* 10:302. 10.1038/s41419-019-1546-9

Piccolo E, Tinari N, D'Addario D, Rossi C, Iacobelli V, La Sorda R, Lattanzio R, D'Egidio M, Di Risio A, Piantelli M, Natali PG, and Iacobelli S. 2015. Prognostic relevance of LGALS3BP in human colorectal carcinoma. *J Transl Med* 13:248. 10.1186/s12967-015-0606-x

Piccolo E, Tinari N, Semeraro D, Traini S, Fichera I, Cumashi A, La Sorda R, Spinella F, Bagnato A, Lattanzio R, D'Egidio M, Di Risio A, Stampolidis P, Piantelli M, Natoli C, Ullrich A, and Iacobelli S. 2013. LGALS3BP, lectin galactoside-binding soluble 3 binding protein, induces vascular endothelial growth factor in human breast cancer cells and promotes angiogenesis. *J Mol Med (Berl)* 91:83-94. 10.1007/s00109-012-0936-6

Pittoni P, Tripodo C, Piconese S, Mauri G, Parenza M, Rigoni A, Sangaletti S, and Colombo MP. 2011. Mast Cell Targeting Hampers Prostate Adenocarcinoma Development but Promotes the Occurrence of Highly Malignant Neuroendocrine Cancers. *Cancer Research* 71:5987-5997. 10.1158/0008-5472.Can-11-1637

Qiu J, Sun M, Wang Y, and Chen B. 2020. Identification and validation of an individualized autophagy-clinical prognostic index in gastric cancer patients. *Cancer Cell Int* 20:178. 10.1186/s12935-020-01267-y

Shah KN, Wilson EA, Malla R, Elford HL, and Faridi JS. 2015. Targeting Ribonucleotide Reductase M2 and NF-kappaB Activation with Didox to Circumvent Tamoxifen Resistance in Breast Cancer. *Mol Cancer Ther* 14:2411-2421. 10.1158/1535-7163.MCT-14-0689

Song Q, Li C, Feng X, Yu A, Tang H, Peng Z, and Wang X. 2015. Decreased expression of SCUBE2 is associated with progression and prognosis in colorectal cancer. *Oncol Rep* 33:1956-1964. 10.3892/or.2015.3790

Stanton SE, and Disis ML. 2016. Clinical significance of tumor-infiltrating lymphocytes in breast cancer. *J Immunother Cancer* 4:59. 10.1186/s40425-016-0165-6

424 Swaby RF, Sharma CG, and Jordan VC. 2007. SERMs for the treatment and prevention of breast
425   cancer. *Rev Endocr Metab Disord* 8:229-239. 10.1007/s11154-007-9034-4
426 Tang J, Ren J, Cui Q, Zhang D, Kong D, Liao X, Lu M, Gong Y, and Wu G. 2019. A prognostic 10-
427   lncRNA expression signature for predicting the risk of tumour recurrence in breast cancer
428   patients. *J Cell Mol Med* 23:6775-6784. 10.1111/jcmm.14556
429 Traini S, Piccolo E, Tinari N, Rossi C, La Sorda R, Spinella F, Bagnato A, Lattanzio R, D'Egidio M, Di
430   Risio A, Tomao F, Grassadonia A, Piantelli M, Natoli C, and Iacobelli S. 2014. Inhibition of
431   tumor growth and angiogenesis by SP-2, an anti-lectin, galactoside-binding soluble 3
432   binding protein (LGALS3BP) antibody. *Mol Cancer Ther* 13:916-925. 10.1158/1535-
433   7163.MCT-12-1117
434 Turner N, Pearson A, Sharpe R, Lambros M, Geyer F, Lopez-Garcia MA, Natrajan R, Marchio C,
435   Iorns E, Mackay A, Gillett C, Grigoriadis A, Tutt A, Reis-Filho JS, and Ashworth A. 2010.
436   FGFR1 amplification drives endocrine therapy resistance and is a therapeutic target in
437   breast cancer. *Cancer Res* 70:2085-2094. 10.1158/0008-5472.CAN-09-3746
438 Viale G. 2012. The current state of breast cancer classification. *Ann Oncol* 23 Suppl 10:x207-210.
439   10.1093/annonc/mds326
440 Voloshanenko O, Erdmann G, Dubash TD, Augustin I, Metzig M, Moffa G, Hundsrucker C, Kerr G,
441   Sandmann T, Anchang B, Demir K, Boehm C, Leible S, Ball CR, Glimm H, Spang R, and
442   Boutros M. 2013. Wnt secretion is required to maintain high levels of Wnt activity in colon
443   cancer cells. *Nat Commun* 4:2610. 10.1038/ncomms3610
444 Wang K, Li J, Xiong Y-F, Zeng Z, Zhang X, and Li H-Y. 2018. A Potential Prognostic Long Noncoding
445   RNA Signature to Predict Recurrence among ER-positive Breast Cancer Patients Treated
446   with Tamoxifen. *Scientific Reports* 8:3179. 10.1038/s41598-018-21581-w
447 Wang Y, Zhao W, Xiao Z, Guan G, Liu X, and Zhuang M. 2020. A risk signature with four autophagy-
448   related genes for predicting survival of glioblastoma multiforme. *J Cell Mol Med* 24:3807-
449   3821. 10.1111/jcmm.14938
450 Woo JK, Jang JE, Kang JH, Seong JK, Yoon YS, Kim HC, Lee SJ, and Oh SH. 2017. Lectin, Galactoside-
451   Binding Soluble 3 Binding Protein Promotes 17-N-Allylamino-17-demethoxygeldanamycin
452   Resistance through PI3K/Akt Pathway in Lung Cancer Cell Line. *Mol Cancer Ther* 16:1355-
453   1365. 10.1158/1535-7163.MCT-16-0574
454 Yang B, Miao S, and Li Y. 2018. SCUBE2 inhibits the proliferation, migration and invasion of human
455   non-small cell lung cancer cells through regulation of the sonic hedgehog signaling
456   pathway. *Gene* 672:143-149. 10.1016/j.gene.2018.06.012
457 Yang PT, Anastas JN, Toroni RA, Shinohara MM, Goodson JM, Bosserhoff AK, Chien AJ, and Moon
458   RT. 2012. WLS inhibits melanoma cell proliferation through the beta-catenin signalling
459   pathway and induces spontaneous metastasis. *EMBO Mol Med* 4:1294-1307.
460   10.1002/emmm.201201486
461 Ye L, Lin C, Wang X, Li Q, Li Y, Wang M, Zhao Z, Wu X, Shi D, Xiao Y, Ren L, Jian Y, Yang M, Ou R,
462   Deng G, Ouyang Y, Chen X, Li J, and Song L. 2019. Epigenetic silencing of SALL2 confers

463   tamoxifen resistance in breast cancer. *EMBO Mol Med* 11:e10638.
464   10.15252/emmm.201910638

465   Zhu Y, Liu Y, Zhang C, Chu J, Wu Y, Li Y, Liu J, Li Q, Li S, Shi Q, Jin L, Zhao J, Yin D, Efroni S, Su F, Yao
466   H, Song E, and Liu Q. 2018. Tamoxifen-resistant breast cancer cells are resistant to DNA-
467   damaging chemotherapy because of upregulated BARD1 and BRCA1. *Nat Commun*
468   9:1595. 10.1038/s41467-018-03951-0

469   Zuo S, Wei M, Zhang H, Chen A, Wu J, Wei J, and Dong J. 2019. A robust six-gene prognostic
470   signature for prediction of both disease-free and overall survival in non-small cell lung
471   cancer. *Journal of Translational Medicine* 17:152. 10.1186/s12967-019-1899-y

472

Figure 1

Figure1

The workflow of identification of breast cancer survival-related Four- Resistance - related genes signature.

Screening DEGs

GSE24460
GSE67916

|logFC|>1.5 & adj.p<0.05
limma differential expression analysis

37 DEGs

GO enrichment analyses

KEGG enrichment analyses

select the ER+ and HER2-
patients in the TCGA database

Univariate Cox regression analysis

8 genes significantly related to DFS(p<0.05)

Stepwise multivariate Cox regression analysis

four-gene-based riskScore

Kaplan-Meier curve

ROC analysis

Building a predictive nomogram

Validation of prognostic model in
GSE11121,GSE21653 and GSE17705

Kaplan-Meier curve

ROC analysis

Figure 2

Figure2

Determination and analysis of the eight genes signature in the training cohort.

(A,B)Univariate Cox (A) and multivariate Cox (B) regression analysis results. (C) The four-gene signature riskScore distribution plot. (D) The vital status plot of patients. (E) The heatmap of the expression profiles of the four genes. (F) The correlation of four genes in TCGA.

A

	pvalue	Hazard ratio
AMIGO2	0.008	0.718(0.563-0.916)
AREG	0.040	0.878(0.775-0.994)
CXCL12	0.045	0.757(0.577-0.993)
GEM	0.048	0.754(0.570-0.998)
LGALS3BP	0.034	0.740(0.560-0.977)
SCUBE2	0.004	0.830(0.731-0.943)
TFF1	0.041	0.911(0.833-0.996)
WLS	0.003	0.695(0.544-0.886)

B

	pvalue	Hazard ratio
AMIGO2	0.139	0.839(0.665-1.059)
LGALS3BP	0.098	0.811(0.633-1.040)
SCUBE2	0.094	0.900(0.795-1.018)
WLS	0.029	0.746(0.573-0.971)

C

D

E

F
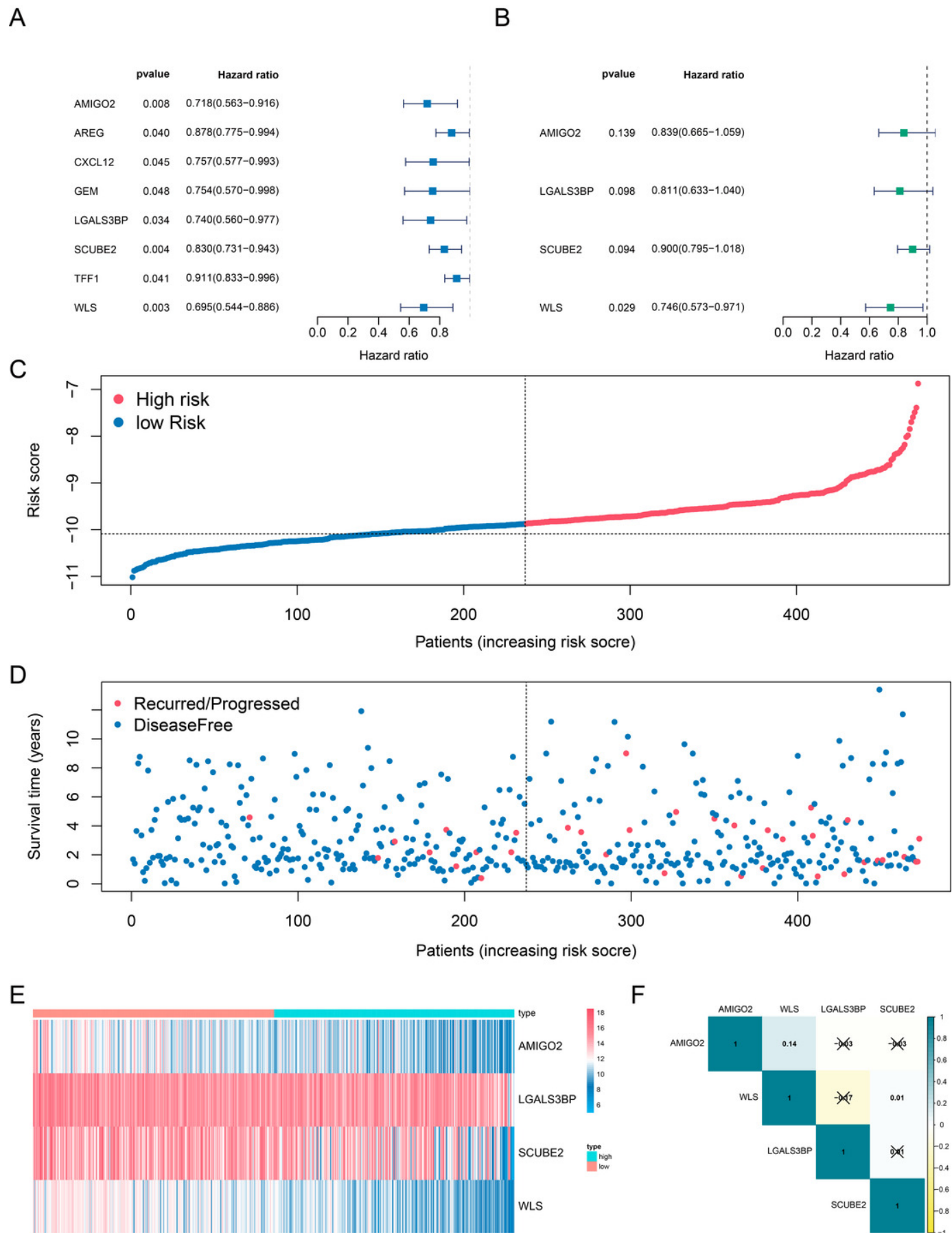
Figure 3

Figure3

Kaplan–Meier survival curves of DFS between high-risk and low-risk patients and ROC curves in (A) TCGA-BRCA , (B) GSE21653 , (C) GSE17705 and (D) GSE11121.
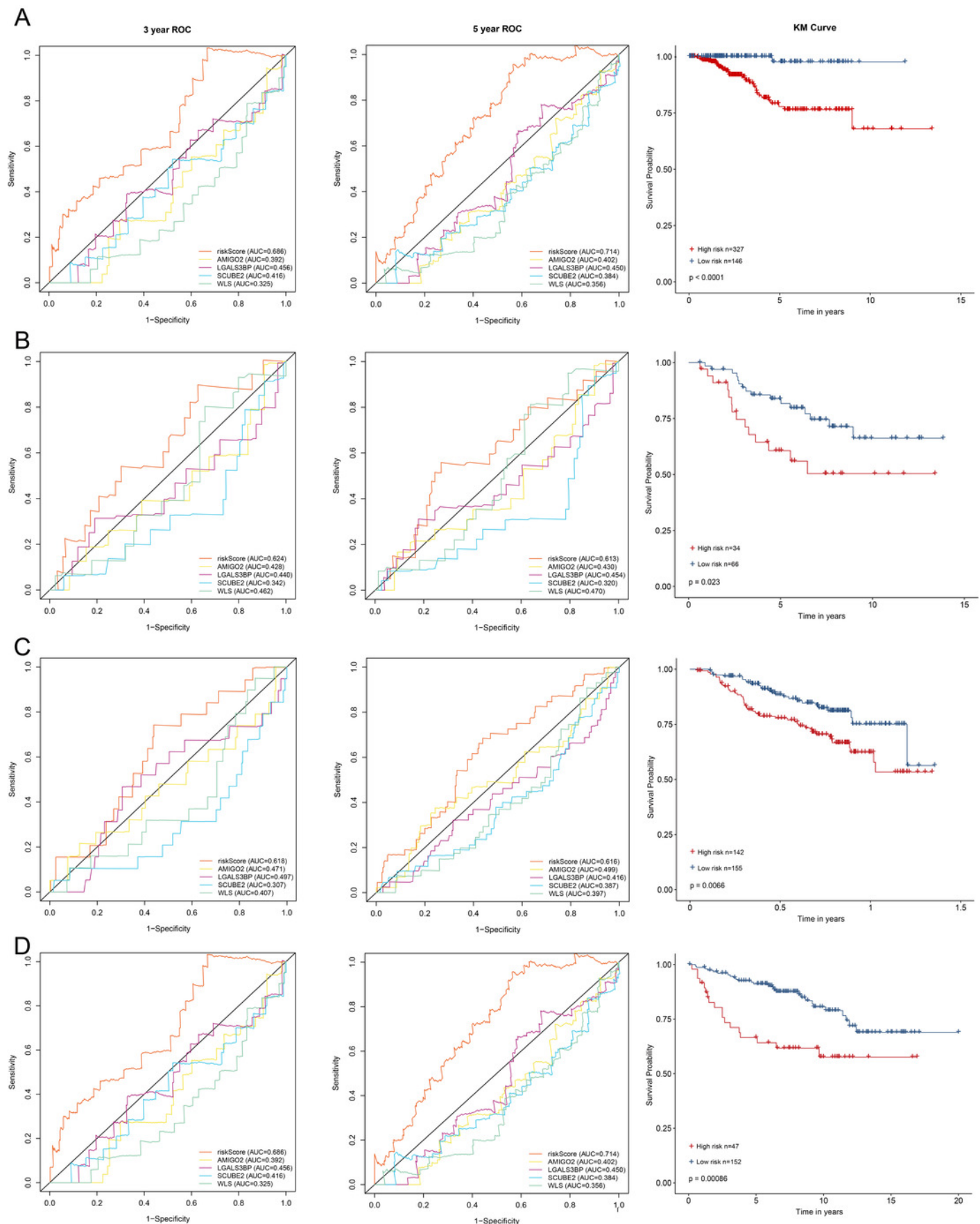
Figure 4

Figure4

Performed the risk stratified subgroup analysis according to clinical factors.
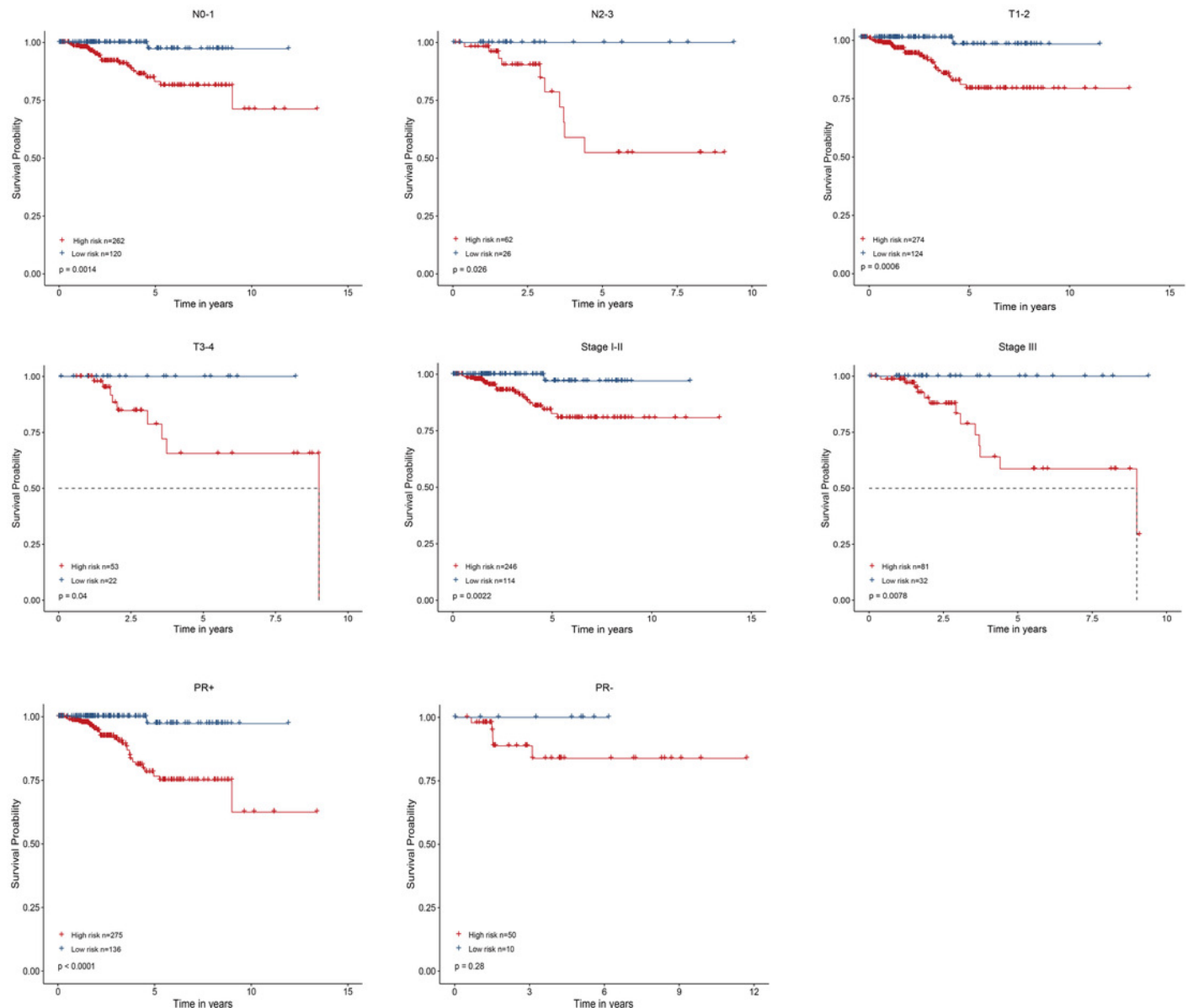
Figure 5

Figure5

The expression of the four genes in different clinicopathological risk factors.
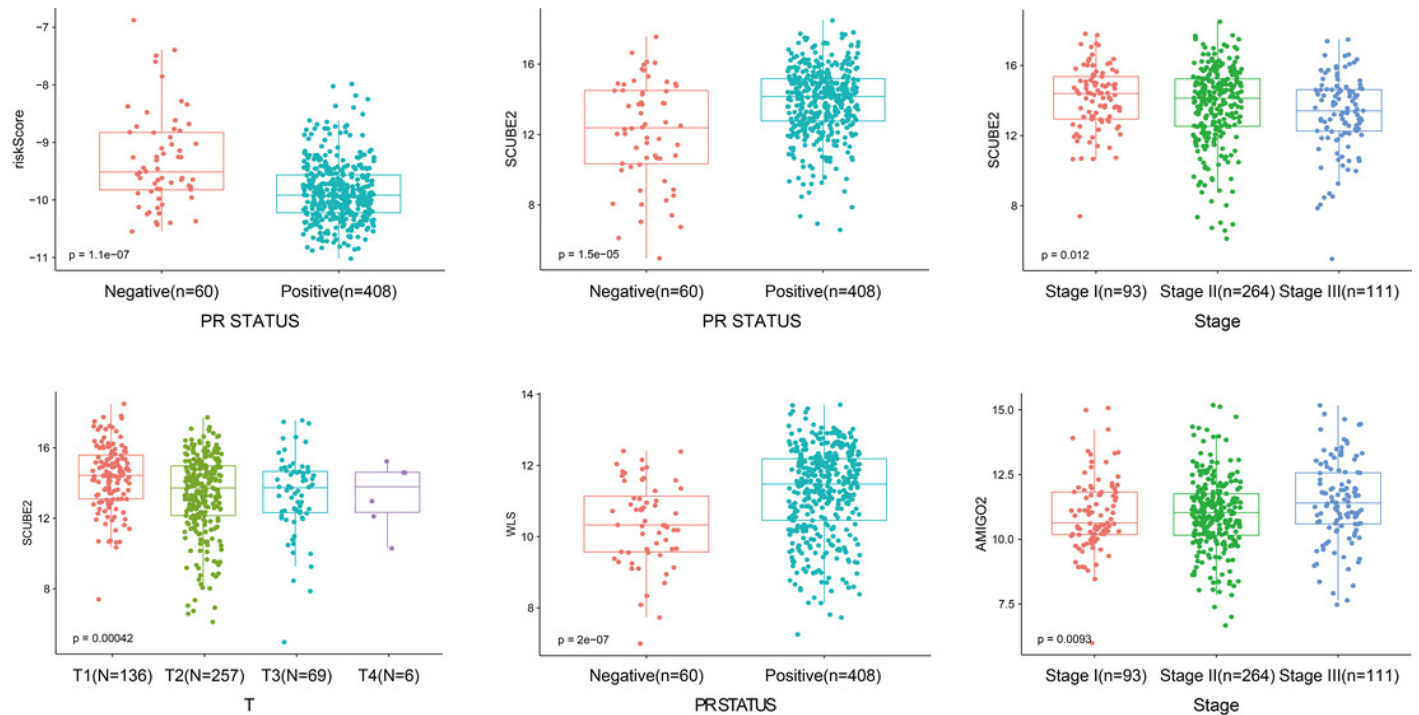
Figure 6

Figure 6

Immune infiltration analyses results.(A)The relative infiltration level of immune cells in pat ients of TCGA-BRCA dataset , the 22 immune cells were annotated by various colors under the legend. (B) T he difference immune cell infiltration levels between the high- and low groups.
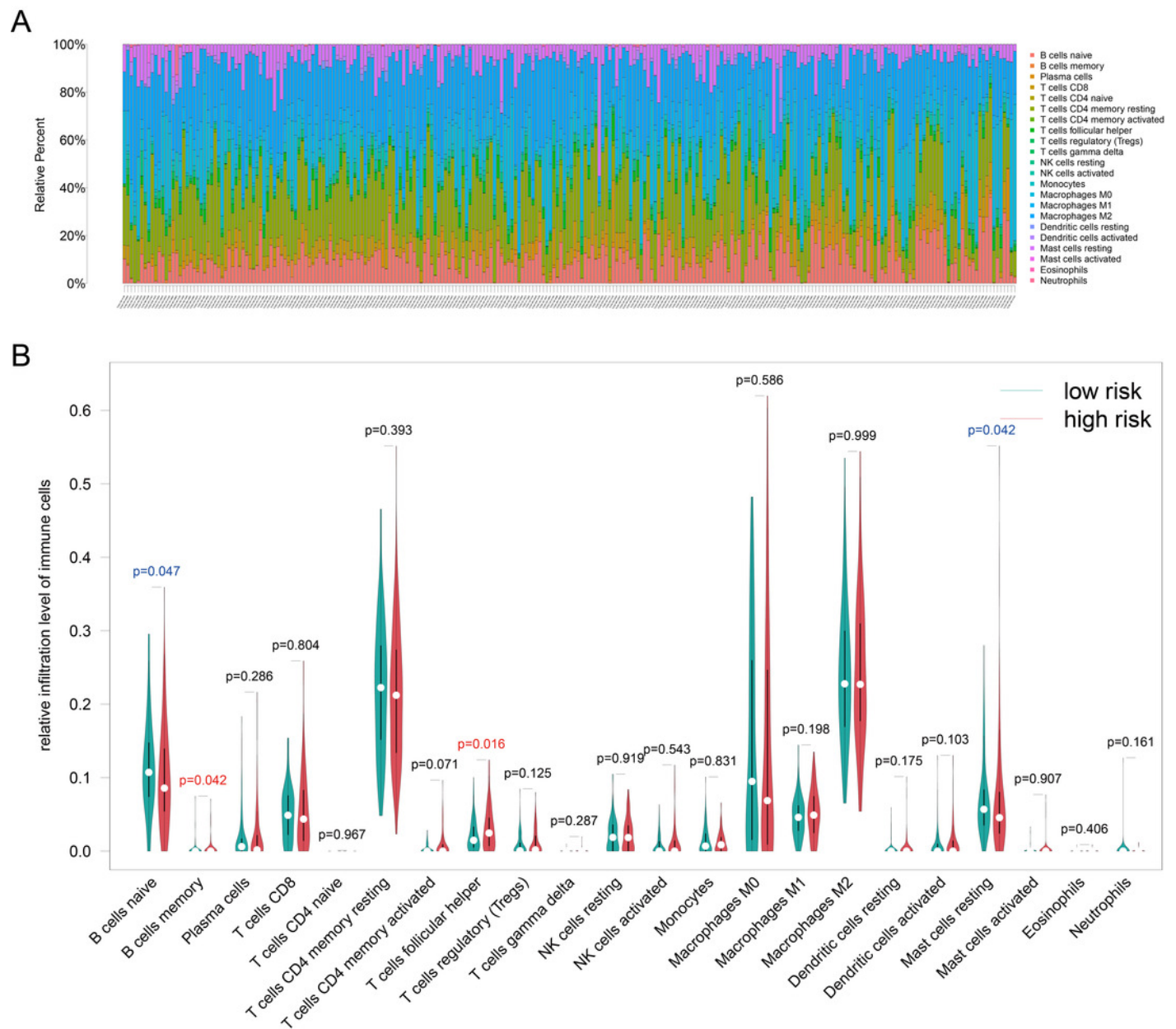
Figure 7

Figure7

Building a prognostic nomogram combining riskScore with clinical risk factors.
(A,B) Univariate Cox (A) and multivariate Cox (B) association of the risk signa ture and
clinicopathological characteristics with DFS in TCGA-BRCA . (C) Nomogram predicting 3- and
5-year DFS for patients with ER+ and HER2- breast cancer patients in stage I-III .(D) T he
calibration plots of the nomogram in prediction of the 3-year and 5-year DFS in the training
cohort . (E)Time-dependent ROC analyses evaluate the accuracy of the nomogram and
riskScore or clinical risk factors.
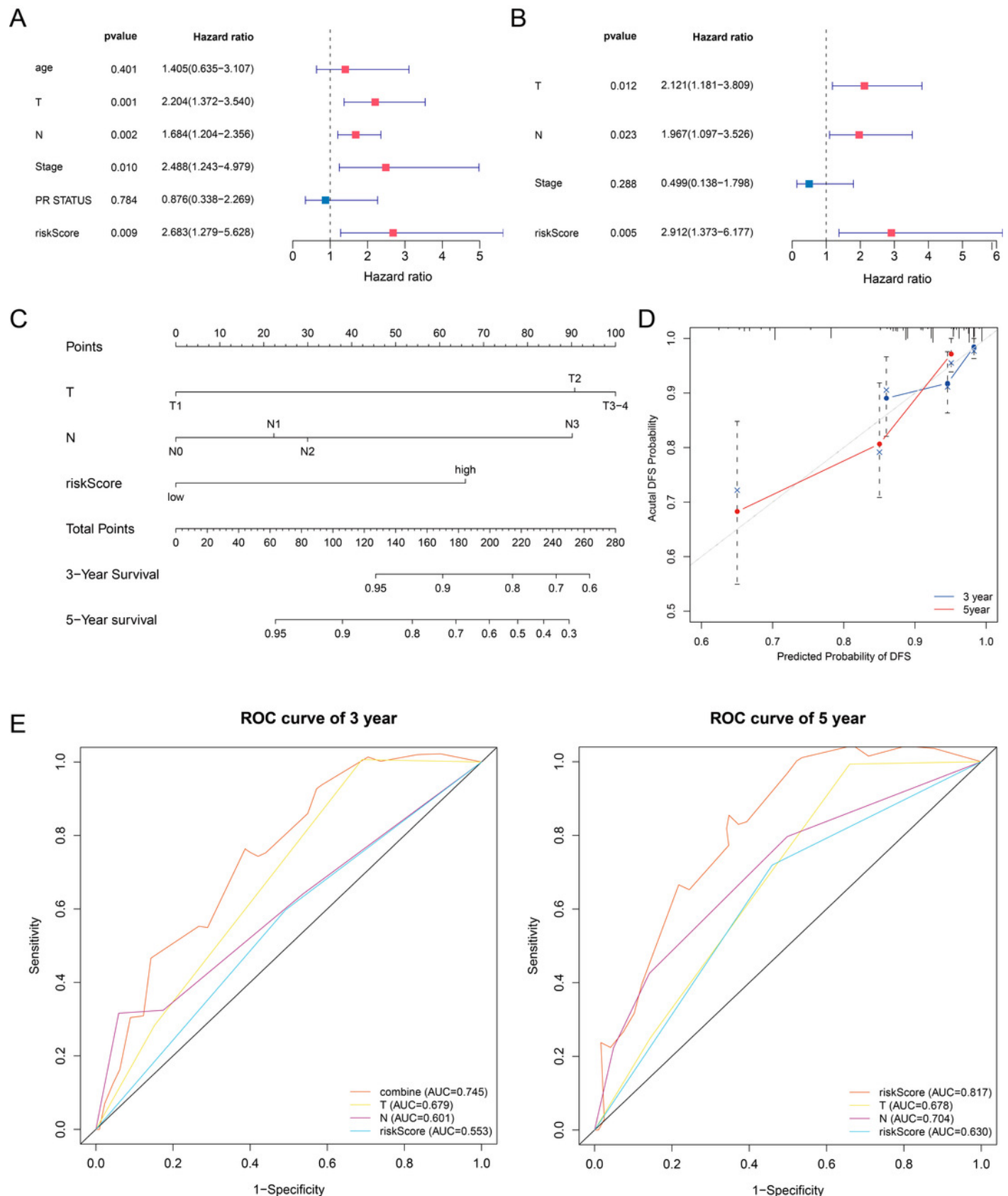
Table 1

Baseline data sheet for patients in TCGA(n=473)

	NO.	%
Age		
≥50 year	348	73.57%
<50 year	125	26.43%
T		
T1	139	29.39%
T2	259	54.76%
T3	69	14.59%
T4	6	1.27%
N		
N0	214	45.24%
N1	168	35.52%
N2	53	11.21%
N3	35	7.40%
NX	3	0.63%
Pathologic Stage		
Stage I	95	20.08%
Stage II	265	56.03%
Stage III	113	23.89%
PR Status		
positive	411	86.89%
negative	60	12.68%
unknown	2	0.42%

—

Table 2

Four resistance-related genes associated with DFS

Gene	Coef	HR	Group
AMIGO2	-0.17541	0.839117	Down-regulated
LGALS3BP	-0.20954	0.810961	Up-regulated
SCUBE2	-0.10588	0.899535	Down-regulated
WLS	-0.29345	0.745685	Up-regulated