# Prognostic analyses of genes associated with anoikis in breast cancer

Jingyu Cao [Equal first author, 1], Xinyi Ma [Equal first author, 2], Guijuan Zhang [1], Shouyi Hong [1], Ruirui Ma [1], Yanqiu Wang [1], Xianxin Yan [Corresp., 1], Min Ma [Corresp. 1]

[1] Jinan University, Guangzhou, Guangdong, China

[2] Southern Medical University, Guangzhou, China

Corresponding Authors: Xianxin Yan, Min Ma
Email address: yanxianxin@jnu.edu.cn, tmamin@jnu.edu.cn

Breast cancer (BRCA) is the most diagnosed cancer worldwide and is responsible for the highest cancer-associated mortality among women. It is evident that anoikis resistance contributes to tumour cell metastasis, and this is the primary cause of treatment failure for BRCA. However, anoikis-related gene (ARG) expression profiles and their prognostic value in BRCA remain unclear. In this study, a prognostic model of ARGs based on The Cancer Genome Atlas(TCGA) database using a least absolute shrinkage and selection operator analysis to evaluate the prognostic value of ARGs in BRCA. The risk factor graph demonstrated that the low-risk group had longer survival than the high-risk group, implying that the prognostic model had a good performance. We identified 11ARGs that exhibited differential expression between the two risk groups in TCGA and Gene Expression Omnibus databases. Through Gene Ontology and Kyoto Encyclopaedia of Genes and Genomes enrichment analyses, it was discovered that the screened ARGs were associated with tumour progression and metastasis. Furthermore, a protein-protein interaction network showed potential interactions among these ARGs. Furthermore, gene set enrichment analysis suggested that the Notch and Wnt signalling pathways were overexpressed in the high-risk group, and gene set variation analysis revealed that 38 hallmark genes differed between the two groups. Moreover, Kaplan–Meier survival curves and receiver operating characteristic curves were used to identify five ARGs (CD24, KRT15, MIA, NDRG1, TP63), and quantitative polymerase chain reaction was employed to assess the differential expression of these ARGs. Univariate and multivariate Cox regression analyses were then performed for the key ARGs, with the best prediction of 2-, 3-, and 4-year survival being obtained from the 3-year survival estimation. In conclusion, ARGs might play a crucial role in tumour progression and serve as indicators of prognosis in BRCA.

# 1 **Prognostic analyses of genes associated with anoikis in breast cancer**

3 Jingyu Cao[1*], Xinyi Ma[2*], Guijuan Zhang[3], Shouyi Hong[4], RuiruiMa[4],Yanqiu Wang[4], Xianxin

4 Yan[4✉],Min Ma[4✉]

5 [1]Department of Oncology, The First Affiliated Hospital, Jinan University, Guangzhou, China.

6 [2] The First Clinical Medical College, Southern Medical University (No: 3210090112),

7 Guangzhou, China

8 [3]School of Nursing of Jinan University, Guangzhou, China

9 [4]College of Traditional Chinese Medicine, Institute of Integrated Traditional Chinese and

10 Western Medicine, Jinan University, Guangzhou, China.

11 *Jingyu Cao, and Xinyi Ma contributed equally to this work.

12 ✉**Correspondence:**

13 Min Ma, College of Traditional Chinese Medicine of Jinan University, 601 Huangpu Avenue

14 West, Guangzhou, Guangdong 510632, People's Republic of China.

15 Email: tmamin@jnu.edu.cn.

16 Xianxin Yan, College of Traditional Chinese Medicine of Jinan University, 601 Huangpu

17 Avenue West, Guangzhou, Guangdong 510632, People's Republic of China.

18 Email: yanxianxin@jnu.edu.cn.

## 20 **Abstract**

22 Breast cancer (BRCA) is the most diagnosed cancer worldwide and is responsible for the highest

23 cancer-associated mortality among women. It is evident that anoikis resistance contributes to

24 tumour cell metastasis, and this is the primary cause of treatment failure for BRCA. However,

25 anoikis-related gene (ARG) expression profiles and their prognostic value in BRCA remain

26 unclear. In this study, a prognostic model of ARGs based on The Cancer Genome Atlas (TCGA)

27  database using a least absolute shrinkage and selection operator analysis to evaluate the

28  prognostic value of ARGs in BRCA. The risk factor graph demonstrated that the low-risk group

29  had longer survival than the high-risk group, implying that the prognostic model had a good

30  performance. We identified 11 ARGs that exhibited differential expression between the two risk

31  groups in TCGA and Gene Expression Omnibus databases. Through Gene Ontology and Kyoto

32  Encyclopaedia of Genes and Genomes enrichment analyses, it was discovered that the screened

33  ARGs were associated with tumour progression and metastasis. Furthermore, a protein-protein

34  interaction network showed potential interactions among these ARGs. Furthermore, gene set

35  enrichment analysis suggested that the Notch and Wnt signalling pathways were overexpressed

36  in the high-risk group, and gene set variation analysis revealed that 38 hallmark genes differed

37  between the two groups. Moreover, Kaplan–Meier survival curves and receiver operating

38  characteristic curves were used to identify five ARGs (CD24, KRT15, MIA, NDRG1, TP63),

39  and quantitative polymerase chain reaction was employed to assess the differential expression of

40  these ARGs. Univariate and multivariate Cox regression analyses were then performed for the

41  key ARGs, with the best prediction of 2-, 3-, and 4-year survival being obtained from the 3-year

42  survival estimation. In conclusion, ARGs might play a crucial role in tumour progression and

43  serve as indicators of prognosis in BRCA.

## Keywords
46  breast cancer; anoikis; prognosis; CD24; NDRG1

## Introduction

50  Breast cancer is the most commonly diagnosed cancer worldwide and has the highest mortality

51  rate among malignant tumours in women[1]. Despite comprehensive treatment, such as surgery,

52  cytotoxic chemotherapy, and targeted therapy, improves clinical remission rates, dissemination

53  and metastasis often lead to the failure of anticancer therapy[2]. Present reports suggest that

54  metastasis is responsible for more than 90% of breast cancer-related deaths[3]. Therefore, the

55  efficacy of treatment for breast cancer is largely dependent on its capacity to prevent metastasis.

56  Tumour cell metastasis involves multiple steps, such as basement membrane invasion,

57  intravasation and transport in the vasculature, and distant organ colonisation[4]. It has been

58   confirmed that tumour cells must overcome various stresses, including anoikis, to achieve distant

59   metastasis[5].

60

61   Anoikis is a form of apoptosis initiated when cells detach from the extracellular matrix (ECM)

62   and has been suggested to be regulated via the extrinsic death receptor pathway and intrinsic

63   mitochondrial pathway[6]. Usually, epithelial and endothelial cells maintain contact-dependent

64   growth by activating the cellular signalling pathway via the ECM; however, their apoptotic

65   program is initiated after the cell-matrix contact loss, which helps maintain tissue homeostasis by

66   eliminating shed cells[6]. Resistance to anoikis has been reported in malignancies, including

67   breast cancer, and is considered a fundamental characteristic of metastatic cancer cells.

68   Therefore, anoikis is a potential therapeutic target for inhibiting cancer metastasis[7].

69

70   Hence, anoikis has recently become the focus of intense research in the field of tumour invasion,

71   and potential molecular mechanisms of anoikis resistance have been explored in several

72   tumours[8]. A previous study indicated that 14-3-3σ induced anoikis resistance and

73   hepatocellular carcinoma cell metastasis by epidermal growth factor receptor-dependent

74   extracellular signal-regulated kinase 1/2 pathway activation[9]. In addition, anoikis-resistant

75   gastric cancer cells exhibited stronger metastatic and proangiogenic traits based on CCAAT-

76   enhancer-binding protein β-mediated platelet-derived growth factor subunit B autocrine and

77   paracrine signalling[10]. Furthermore, anoikis-resistant molecular mechanisms have been

78   explored in breast cancer[11]. Human epidermal growth factor receptor (HER) 2 and Src-

79   dependent lactate dehydrogenase A activation contributed to breast cancer cell anoikis resistance

80   and metastasis[12], and HER2-dependent B cell linker down-regulation facilitated three-

81   dimensional breast tumour growth[13]. Although research has demonstrated that several genes

82   are involved in anoikis resistance in breast cancer[11], research on the clinical relevance of

83   anoikis-related genes (ARGs) and the systematic evaluation of the implications of anoikis in

84   breast cancer is lacking.

85

86   In this research, we first constructed an ARG prognostic model based on the Cancer Genome

87   Atlas (TCGA) Breast Invasive Carcinoma (BRCA) (TCGA-BRCA) database and subsequently

88   screened 11 differentially expressed ARGs between high- and low-risk groups. Functional

89    annotation enrichment and protein-protein interaction (PPI) network analysis were performed for

90    the screened ARGs. We identified five key ARGs by assessing their potential diagnostic and

91    prognostic values and determined the messenger ribonucleic acid (RNA) (mRNA) levels of the

92    five key ARGs by quantitative polymerase chain reaction (qPCR). Univariate and multivariate

93    Cox regression analyses were employed to assess the ability of 2-, 3-, and 4-year survival

94    predictions with respect to key ARGs. Altogether, our findings demonstrate the significant roles

95    of ARGs in breast cancer progression and provide novel research perspectives for anoikis

96    resistance.

97

98

## Materials & Methods

100

**Data collection**

102    RNA sequencing data of patients with breast cancer were downloaded from TCGA database

103    (https://portal.gdc.cancer.gov/) using the R package TCGAbiolinks[14]. After eliminating

104    samples with incomplete clinical annotations, 1,083 cases of breast cancer projects, including

105    111 cases with matched adjacent tissues, were obtained (TCGA-BRCA), and the data format was

106    level 3 HTSeq-fragments per kilobase per million. Additionally, corresponding clinical

107    information was obtained from the UCSC Xena database[15]. The mRNA counts of TCGA-

108    BRCA dataset were normalised by using the limma R package[16].

109

110    The breast cancer-related datasets GSE42568[17], GSE102484[18], and GSE20685[19] were

111    downloaded from the Gene Expression Omnibus (GEO) database

112    (http://www.ncbi.nlm.nih.gov/geo/)[20] using the R package GEO query[21]. The detection

113    platform for the three datasets was GPL570 (HG-U133_Plus_2). GSE42568 included 104 breast

114    cancer samples and 17 matching normal tissues, GSE102484 comprised 683 breast cancer

115    samples, and GSE20685 comprised 327 breast cancer samples.

116

117    Eight gene sets of ARGs were retrieved from the Molecular Signatures Database (MSigDB)[22],

118    which provides the most comprehensive information on human gene sets, resulting in 862 ARGs.

119    Furthermore, 794 ARGs were retrieved from the GeneCards database[23]. By taking the

120    intersection of the two databases and omitting those genes with missing probes, we obtained 100

121    ARGs that are common genes in both databases (Table S1).

122

**Prognostic model construction**

124    The least absolute shrinkage and selection operator **(**LASSO**)** prognostic model of 100 ARGs for

125    breast cancer was constructed using the TCGA-BRCA dataset, employing 10-fold cross-

126    validation with seeds 2021 and a p-value of 0.05. We ran 1000 iterations to guard against

127    overfitting. The outcomes of the LASSO analysis were visualised. All samples from the TCGA-

128    BRCA dataset and BRCA dataset were divided into high- and low-risk groups, respectively,

129    according to the median risk score of the prognostic model, and the survival outcomes were

130    displayed in spot graphs. Moreover, the ARG expression profiles for prognostic model

131    construction were visualised using heatmaps.

132

**Differential expression genes in the high- and low-risk groups**

134    First, the batch effects of GSE42568, GSE20685, and GSE102484 were eliminated using the R

135    package sva[24], and merged BRCA datasets were retrieved from the GEO database. Second,

136    datasets from TCGA and GEO were standardised and then grouped into high- and low-risk

137    groups according to the median risk score of the LASSO prognostic model. Differential gene

138    expression analysis was performed between the two groups, and a volcano plot (ggplot2) and

139    heatmap (R package pheatmap) were created to visualise up-regulated genes (logFC >0 and p

140    <0.05) and down-regulated genes (logFC <0 and p <0.05).

141

**Gene enrichment analysis and the PPI network**

143    The clusterProfiler R package was used to perform Gene Ontology (GO)[25] and Kyoto

144    Encyclopaedia of Genes and Genomes (KEGG)[26] pathway enrichment analyses[27] with p

145    <0.05 and a false discovery rate (FDR) value (q-value) of <0.2. The Search Tool for the

146    Retrieval of Interacting Genes/Proteins (STRING) database was accessed to provide information

147    regarding known and predicted PPIs[28], From which information regarding ARGs was

148    extracted from the STRING database, and a PPI network was constructed using Cytoscape[29].

149    Additionally, GeneMANIA was used for gene prediction of the selected ARGs, and the results

150    were visualised[30].

151

**Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)**

153   GSEA is an analytical method that determines whether predefined gene sets demonstrate

154   statistically significant and consistent differences between two biological states[31]. In the

155   current study, TCGA-BRCA genes were divided into phenotype-related high- and low-risk

156   groups based on the risk score median of the LASSO prognostic model, and GSEA was

157   performed using the clusterProfiler R package. The procedure was repeated 1,000 times for each

158   analysis, and c2.cp.v7.2.symbols were obtained from MSigDB to serve as the reference gene

159   collection. Statistical significance was set at FDR <0.25 and p <0.05.

160

161   We downloaded "h.all.v7.4.symbols.gmt" from the MSigDB database to perform GSVA based

162   on gene expression levels in TCGA-BRCA dataset and analyse variations of functional

163   enrichment between the high- and low-risk groups. Statistical significance was set at p <0.05.

164

**Analysis of the prognostic value and receiver operating characteristic (ROC) curve for**
**ARGs**

167   Kaplan–Meier curves were used to estimate the overall survival of patients with breast cancer

168   based on the ARG expression derived from TCGA-BRCA dataset. The ROC curve was used to

169   assess the relationship between ARGs and tumourigenesis, and genes with an area under the

170   curve (AUC) of the ROC curve >0.6 were considered key genes.

171

**Correlation analysis between clinical characteristics and prognosis**

173   First, we performed univariate Cox regression to evaluate the prognostic value of clinical

174   characteristics and key ARGs in patients, and a multivariate Cox regression model was

175   constructed from factors with a p-value of <0.1 in univariate Cox regression. Subsequently,

176   nomograms were used to predict the 2-, 3-, and 4-year survival probabilities were predicted

177   individually based on a multivariate Cox regression model. The calibration curves were

178   evaluated graphically assessed by plotting the nomogram-predicted probabilities against the

179   observed occurrences, and the 45° line represented the best predictive values. The nomograms

180   and calibration curves were produced by the R package rms. Decision curve analysis (DCA) was

181 performed using the R package ggDCA[32] to evaluate the predictive effect of the nomograms

182 for survival probability.

183

184 **Cell cultures and real-time qPCR**

185 The human normal breast epithelial cell line Michigan Cancer Foundation (MCF)-10A and

186 breast cancer cell line HCC1954, MDA-MB-231, and MCF-7 were procured from the American

187 Type Culture Collection. MDA-MB 231 is oestrogen (ER)/progesterone (PR)/HER2 negative,

188 HCC1954 is ER/PR negative and HER2 positive, and MCF-7 is ER/PR positive and HER2

189 negative. MCF-10A cells were cultured in Dulbecco's Modified Eagle Medium (DMEM)/F12

190 medium (GIBCO, USA) with 5% horse serum (HyClone, USA) and other supplements. MDA-

191 MB-231and MCF-7 cells were cultured in DMEM (GIBCO, USA) containing 10% foetal bovine

192 serum (GIBCO, USA), and HCC1954 cells were cultured in Rosewell Park Memorial Institute

193 1640 medium (GIBCO, USA) containing 10% foetal bovine serum (GIBCO, USA). All cell lines

194 were incubated at 37°C under 5% carbon dioxide according to the recommended protocols.

195

196 Total RNA was extracted from cells which had been cultured in dishes using TRIzol (Life

197 Technologies, USA). Reverse transcription was then performed using an Evo M-MLV RT Mix

198 Tracking Kit with genomic deoxyribonucleic acid (DNA) Clean (Accurate Biology, Changsha,

199 China), and qPCR was performed using an HS SYBR Green Premix Pro Taq qPCR Tracking Kit

200 (Accurate Biology, Changsha, China). All steps were performed according to the manufacturer's

201 instructions. The primer sequences were as follows:

| Genes | Forward primer | Reverse primer |
|---|---|---|
| CD24 | CCCCAAATCCAACTAATGC | GGACTTCCAGACGCCATT |
| Keratin 15 (KRT15) | GACGGAGATCACAGACCTGAG | CTCCAGCCGTGTCTTTATGTC |
| Melanoma-inhibiting activity (MIA) | CCTCCGTGTCCACTAAAT | TTCTTGCCGTTCATACTC |
| N-Myc downstream regulated 1 (NDRG1) | ACACCTACCGCCAGCACA | GCCACAGTCCGCCATCTT |
| Tumour protein p63 (TP63) | CCTTACATCCAGCGTTTC | TTTGTCGCACCATCTTCT |

202    The ΔΔCt method was used to analyse qPCR data. Normalised ΔCt values were obtained by

203    subtracting the Ct values of the housekeeping reference gene (β-actin) from the target gene Ct

204    values. ΔΔCt values were calculated by subtracting the testing group target gene ΔCt values from

205    the average ΔCt of the control group. Finally, relative mRNA expression was obtained by $2^{-\Delta\Delta Ct}$.

206

207    **Immunohistochemical (IHC) analysis**

208    IHC information on key genes in breast cancer and normal breast tissues was obtained from the

209    Human Protein Atlas (HPA) database[33].

210

211    **Statistical analysis**

212    Clinical data handling in the study was performed using the statistical software R (version 4.1.2).

213    Continuous variables were summarised using the mean ± standard deviation. The Wilcoxon-rank

214    sum test was used to compare the two groups, and Student's t-test was used to compare the

215    normally distributed data of the two groups. The Kruskal–Wallis test was used for comparing

216    more than two groups, and categorical variables were compared using the chi-square or Fisher's

217    exact tests. Survival analyses and univariate and multivariate Cox regressions were performed

218    using the survival package in R, and LASSO analyses were performed using the R package

219    glmnet. Furthermore, survivalROC in the R package was used for creating the ROC curve. The

220    significance of survival difference was tested with the log-rank test. In addition, two-sided

221    Spearman's rank correlation coefficients were employed. One-way analysis of variance was used

222    for comparing multiple samples of experimental data. Differences with p-values <0.05 were

223    considered significant.

224

225    # Results

226

227    **LASSO prognostic models were established and 11 ARGs with possible prognostic values**

228    **were screened**

229    A flow chart of the study is provided in the Figure 1. BRCA datasets of GEO were obtained from

230    GSE42568, GSE20685, and GSE102484 after eliminating batch effects. Box plots and principal

231    component analysis plots displayed the datasets before and after eliminating batch effects (Figs.

232    S1 A–D). The results revealed that the batch effect had been effectively eliminated. We

233    performed LASSO analysis of ARGs based on the TCGA-BRCA dataset (Fig. 2A) to evaluate

234 the prognostic value of 100 ARGs from the intersection of MSigDB and the GeneCards database

235 (Table S1), and the variable trajectories graph was used to visualise the results (Fig. 2B). The

236 risk factor graph demonstrated that those in low-risk groups had longer survival compared with

237 those in high-risk groups (Fig. 2C) and 12 ARGs (BIRC3, CD24, FGFR1, IVL, KRT15, L1CAM,

238 MIA, NDRG1, NOS2, PTPN3, SERPINA1, and TP63) of potential prognostic value were

239 screened (Fig. 2C) (Table 1). Next, we explored the 12 ARG profiles between two risk groups in

240 TCGA-BRCA and BRCA datasets, respectively, and visualised the results with pheatmaps (Figs.

241 2F, G). The box plot further showed a comparison of each ARG between the two risk groups,

242 and there were 11 significant differentially expressed ARGs (BIRC3, CD24, IVL, KRT15,

243 L1CAM, MIA, NDRG1, NOS2, PTPN3, SERPINA1, and TP63) with the same expression trends

244 in both datasets (Figs. 2H, I).

245

246 Subsequently, we found 878 differentially expressed genes between the two risk groups based on

247 TCGA-BRCA dataset following thresholds of |logFC| >0.5 and P.adj <0.05. Of these, 244 genes

248 were up-regulated, and 634 were down-regulated in the high-risk group, and the volcano plot

249 displayed the results (Fig. 2D). In addition, 196 genes were obtained from the BRCA dataset

250 under the same thresholds comprising 57 up-regulated genes and 139 down-regulated genes (Fig.

251 2E).

252

253 **Results of the gene enrichment analysis and PPI network based on screened ARGs**

254 We performed GO and KEGG enrichment analysis, including cellular components, biological

255 processes (BPs), molecular functions (MFs), and signalling pathways, with the criteria of

256 pc<0.05 and FDR (q-value) <0.25 (Table 2) to elucidate the biological functions of the selected

257 11 ARGs. GO analysis demonstrated that 11 ARGs are predominantly enriched in BP-related

258 processes (such as keratinocyte differentiation, response to hypoxia, and negative proteolysis

259 regulation) and MF-related processes (such as endopeptidase inhibitor activity, peptidase

260 inhibitor activity, and flavin mononucleotide [FMN] binding) (Fig. S2A). KEGG pathway

261 analysis showed that the 11 ARGs were primarily associated with small-cell lung cancer and

262 toxoplasmosis (Fig. S2A). Fig. S2B illustrates the potential links between these items.

263 Furthermore, the chord diagram illustrated each ARG's distribution in the GO items and KEGG

264 pathways (Fig. S2C). CD24, NOS2, and NDRG1 were associated with response to hypoxia.

265    KRT15, TP63, and IVL were mainly enriched in the process of keratinocyte differentiation.

266    BIRC3 and NOS2 were closely associated with small-cell lung cancer pathways. SERPINA1 and

267    BIRC3 were associated with peptidase inhibitor activity.

268

269    PPI networks were constructed using the STRING database with an interaction threshold of 0.15

270    to ascertain the interactions between the 11ARGs in breast cancer. Visualisations were

271    performed using the Cytoscape software (Fig. S2D). The results indicated potential interaction

272    between ARGs. Furthermore, genes with similar functions to ARGs were predicted, and a

273    protein network was constructed using GeneMANIA, with ephrin type-B receptor 2, RAN

274    binding protein 9, and activated leukocyte cell adhesion molecule being closely associated with

275    the 11 ARGs (Fig. S2E).

276

277    **Signalling pathways and hallmark gene sets associated with risk groupings**

278    We subsequently performed GSEA and GSVA between high- and low-risk groups based on

279    TCGA-BRCA dataset. to explore potential relationships between ARGs and gene sets related to

280    breast cancer. The results of GSEA indicated several significantly enriched BPs (Table 3),

281    including pre-Notch expression and processing (Fig. 3B), signalling by Notch (Fig. 3C), T cell

282    factor-dependent signalling in response to Wnt (Fig. 3D), and signalling by Wnt (Fig. 3E), as

283    illustrated in Figure 3A. Other BPs, including transcription of androgen receptor-regulated genes

284    kallikrein-related peptidase (KLK) 2 and KLK3, DNA damage telomere stress-induced

285    senescence, and processes associated with cell cycle regulation, are presented in Table 3. We

286    performed GSVA on all genes in TCGA-BRCA to further explore the differentially expressed

287    hallmark gene sets between the two risk groups. Table 4 reveals that 38 hallmarks exhibited

288    differences, which were visualised using heat maps (Fig. 3F), and the gene sets with p-values

289    <0.001 were compared between two risk groups and visualised using box plot diagrams (Fig.

290    3G). The results implied that gene sets of Myc targets, E2F targets, hypoxia, p53, tumour

291    necrosis factor (TNF)-α/nuclear factor kappa B (NF-κB), interleukin (IL)-6/Janus kinase

292    (JAK)/signal transducer and activator of transcription 3 (STAT3), and mammalian target of

293    rapamycin complex 1 (mTORC1) signalling pathway had marked differences in their enrichment

294    between the two groups.

295

**Analysis of the prognostic significance and ROC curve of the ARGs**

Kaplan–Meier survival curves were plotted for 11 ARGs based on TCGA-BRCA dataset (Figs. 4A–I) to further explore the roles of ARGs in the prognosis of patients with breast cancer. We found that nine ARGs impacted the prognosis of patients with breast cancer (p <0.05). Specifically, BIRC3 (hazard ratio [HR] = 0.67, p = 0.015), KRT15 (HR = 0.61, p = 0.002), MIA (HR = 0.68, p = 0.021), SERPINA1 (HR = 0.58, p = 0.001), and TP63 (HR = 0.64, P = 0.007) were positively correlated with overall survival, while CD24 (HR = 1.69, p = 0.002), IVL (HR = 1.71, p = 0.001), NDRG1 (HR = 1.41, p = 0.036), and NOS2 (HR = 1.46, p = 0.022) were associated with a poor prognosis. We also performed ROC curves to evaluate the correlation between ARGs and tumourigenesis using TCGA-BRCA dataset. As a result, the AUC of the ROC curve was calculated, and five key ARGs (CD24, KRT15, MIA, NDRG1, and TP63) with AUC >0.6 were demonstrated (Figs. 4 J–N).

**Correlation analysis between the clinical characteristics and prognosis**

Clinical information of patients from TCGA-BRCA dataset was extracted to assess the LASSO prognostic model (Table 5). We initially performed univariate Cox regression to evaluate the prognostic value of crucial ARGs (CD24, KRT15, MIA, NDRG1, and TP63) in patients and then formulated a multivariate Cox regression model comprising of those factors with a p-value <0.1 in the univariate Cox regression (Table 6). Forest plots revealed that two genes (CD24 and NDRG1) were associated with increased risk with HRs >1, while three other genes (KRT15, MIA, and TP63) were protective genes with HRs <1 (Fig. 5A). Subsequently, we constructed nomograms for ARGs to evaluate the predictive power of the Cox regression models (Fig. 5B). Furthermore, we performed calibration curves for 2-, 3-, and 4- year survival predictive power of Cox regression model nomograms, respectively (Figs. 5C–E). The results indicated that the blue line corresponding to the 3-year survival was closest to the ideal diagonal 45° line (grey line), implying that the Cox regression models yielded the most reliable prediction for the 3-year survival. Finally, DCA was used to assess the clinical applicability of the LASSO-Cox prognostic model for the 2-, 3-, and 4-year survival (Figs. 5F–H), and the blue line in the graphs represents the model's predictive power. The range on the x-axis of the blue line was higher than that of the red line (all positive) and grey line (all negative) for the 3-year survival, indicating the best clinical applicability in terms of the 3-year survival.

327

328   **mRNA levels of ARGs in MCF-10A, MDA-MB 231, HCC1954, and MCF-7 cell lines.**

329   Gene mRNA levels of key ARGs in normal mammary and breast cancer cell lines were detected

330   using qPCR. We ascertained that CD24 expression was up-regulated in breast cancer cells

331   compared with normal breast epithelial cells (Fig. 6A). Moreover, KRT15, MIA, and TP63

332   levels significantly decreased in breast cancer cells (Figs. 6B, C, E). However, mRNA levels of

333   NDRG1 varied across the three breast cancer lines, being up-regulated in MDA-MB 231 cells

334   and down-regulated in HCC1954 and MCF-7 cells (Fig. 6D).

335

336   **CD24 and NDRG1 protein levels in breast cancer tissues**

337   The expression levels of CD24 and NDRG1 that had an HR of >1 in the multivariable Cox

338   regression model were detected using IHC based on the HPA database. It was observed, in

339   comparison to normal breast tissues, there was a significant increase in the CD24 and NDRG1

340   expression in breast cancer (Figs. S3 A-D).

341

342   # Discussion

343   Breast cancer remains a serious threat to women due to its high incidence and mortality rates[1],

344   with distant metastasis being the primary cause of breast cancer-associated mortality

345   worldwide[34]. However, treatments for metastasis remain elusive. Therefore, it is necessary to

346   study the underlying molecular mechanisms of metastasis, while at the same time, there is an

347   urgent need to explore novel clinical diagnostic and prognostic biomarkers and therapeutic

348   targets for breast cancer. Anoikis resistance has been established as a hallmark of advanced

349   malignancy due to its ability to promote tumour cell invasion and metastasis[7]. Recently,

350   researchers have been focusing more on ARGs due to their potential prognostic and therapeutic

351   value[35]. However, a comprehensive understanding of ARGs in relation to breast cancer is

352   lacking.

353

354   In this study, we performed a LASSO analysis based on TCGA-BRCA dataset through ARGs

355   and divided all samples from TCGA-BRCA and BRCA datasets into high- and low-risk groups

356   based on the median risk score of the prognostic model. Then we screened 11 significant

357   differentially expressed ARGs between two risk groups with the same expression patterns in

358 both datasets, suggesting their potential prognostic value in breast cancer samples. Thus, to
359 further investigate the potential function and associated molecular mechanisms of the screened
360 ARGs, we performed function and pathway enrichment analysis. The results revealed that ARGs
361 were primarily enriched in response to hypoxia, proteolysis regulation, peptidase inhibitor
362 activity, FMN binding, and small-cell lung cancer. To the best of our knowledge, resistance to a
363 hypoxic environment, which is one of the hallmarks of tumour cells, could enhance the anoikis
364 resistance of tumour cells and promote cell survival and distant metastasis[36,37]. As predicted,
365 our results remain consistent with those previous studies that found ARGs to be enriched in
366 hypoxia-related functional processes. Proteolysis is a fundamental metabolic process essential
367 for life and abnormal regulation of proteolysis and peptidase activity are associated with cancer.
368 For example, tumour cell metastasis can be facilitated by the degradation of the extracellular
369 matrix by secreted proteinases[38]. FMN, which is a cofactor of enzymes, has been the focus of
370 photodynamic therapy for tumours[39]. Therefore, our data enhance our understanding of ARGs
371 in breast cancer and guides further research.

372

373 It is now evident that activation of the Notch and WNT signalling pathway was associated with
374 anoikis resistance in various malignancies, including breast cancer[40-42]. GSEA results
375 revealed that the Notch and Wnt signalling pathways were enriched in the high-risk group, which
376 implies that the poor prognosis of patients with breast cancer in the model might be associated
377 with abnormal pathway activation. Notably, the results further revealed potential regulatory
378 relationships between ARGs and the Notch and Wnt signalling pathways, which need to be
379 further explored in the future. Growing evidence has indicated that the androgen receptor
380 modulates breast cancer progression[43]. However, the relationship between anoikis and KLK
381 expression remains unclear. GSEA revealed that ARGs might be involved in the process of
382 transcription of androgen receptor-regulated genes KLK2 and KLK3 in breast cancer, which
383 might offer clues for a more in-depth investigation of ARGs and anoikis resistance in breast
384 cancer. In addition, GSVA results revealed 30 gene sets that exhibited significant differences
385 between the two risk groups, comprising Myc targets, E2F targets, hypoxia, p53, TNFα/NF-κB,
386 IL-6/JAK/STAT3, and mTORC1 signalling pathways, implying a potential correlation between
387 these pathways and the role of ARGs in the invasion and metastasis of breast cancer. Previous
388 studies have demonstrated that p53 induced anoikis and decreased breast cancer metastasis[44],

389    and mTORC1 has been reported to be associated with anoikis in breast cancer[45]. This further

390    supported our study. However, the relationship between anoikis and Myc and E2F in breast

391    cancer is yet to be identified. Thus, we are the first to uncover this association in breast cancer.

392

393    Subsequently, to elucidate the prognostic value of ARGs and their correlation with breast cancer

394    progression, we plotted survival and ROC curves individually. As a result, five AGRs (CD24,

395    KRT15, MIA, NDRG1, and TP63), which met the criteria were identified as key genes and were

396    used for univariate and multivariable Cox regression analysis combined with clinical

397    characteristics from TCGA samples. Furthermore, calibration curves demonstrated a good

398    prognostic performance of the Cox regression model, with the most accurate prediction being

399    made at the 3-year survival. DCA indicated the best clinical applicability in the 3-year survival.

400    Precise prognosis prediction is essential for treating patients with breast cancer. This new

401    information from our study might provide more information for developing prognostic tools.

402    Therefore, the five ARGs might play significant roles in breast cancer progression and

403    aggressiveness.

404

405    We confirmed the strong expression of CD24 in breast cancer through qPCR and IHC, which is

406    consistent    with    previous    reports[46].    CD24    is    a    mucin-like    glycosylated

407    glycophosphatidylinositol-anchored molecule, which could help resist anoikis in various

408    tumours[47]. It has been suggested recently that CD24 is an immune checkpoint promoting

409    immune evasion through its ability to bind to sialic acid-binding immunoglobulin-like lectin

410    10[46]. Furthermore, it has been confirmed that CD24 might be a potential therapeutic target,

411    according to preclinical studies[48]. Nevertheless, there are some conflicting reports on the role

412    of CD24 in cancer[48]. Our study provided clear evidence of significant CD24 up-regulation at

413    the protein and mRNA levels in breast cancer, and it has been suggested to be a high-risk factor

414    resulting in poor prognosis in patients with breast cancer.

415

416    KRT15, a type I keratin, has been shown to be highly expressed in oesophageal carcinoma,

417    colorectal cancer, and gastric cancer[49], whereas it is down-regulated in breast and prostate

418    cancers[50]. Nevertheless, its precise functional role and associated molecular mechanisms in

419    breast cancer remain unclear. Our data suggest that KRT15 might be a tumour suppressor gene

420    and could be associated with longer overall survival of patients with breast cancer. Furthermore,

421    enrichment analysis proposed further avenues of exploration into the role of KRT15 in anoikis.

422

423    MIA is secreted by malignant melanoma cells and is associated with cell detachment from the

424    ECM and the promotion of melanoma cell invasion and metastasis[51]. To date, the relationship

425    between MIA and breast cancer has not been reported. Our study is the first to reveal that MIA

426    acts in a protective manner in patients with breast cancer. We also used qPCR to confirm that the

427    MIA mRNA levels were lower in breast cancer cell lines compared with normal breast cells,

428    which gave insights into breast cancer progression and metastasis.

429

430    Currently, the exact role of NDRG1in breast cancer remains controversial[52]. Herein, survival

431    analysis demonstrated that NDRG1 might be an oncogene associated with a poor prognosis.

432    Furthermore, IHC analysis of the HPA database demonstrated that NDRG1 protein expression

433    was higher in breast cancer tissues. However, NDRG1 mRNA levels were increased in MDA-

434    MB 231 cells and decreased in HCC 1954 cells and MCF-7 cells compared with normal

435    mammary cells, indicating that NDRG1 might play distinct roles in different breast cancer

436    subtypes. Consequently, we shall further examine the protein and mRNA levels of NDRG1 in

437    additional clinical samples. Simultaneously, a study of the underlying molecular mechanisms of

438    NDRG1 is required to determine its role in breast cancer.

439

440    Protein p63 encoded by the TP63 gene belongs to the P53 transcription factor family, and

441    various isoforms are generated due to alternative splicing[53]. Two major subtypes of p63

442    (TAp63 and ΔNp63) have contrasting roles in breast cancer. TAp63 acts as a tumour suppressor,

443    while ΔNp63 might promote tumourigenesis and progression in breast cancer[53]. Our study

444    demonstrated that decreased TP63 expression is a negative and independent prognostic factor in

445    breast cancer, and the TAp63 mRNA level was down-regulated in MDA-MB 231, HCC1954,

446    and MCF-7 cells in laboratory tests, consistent with previous reports. In conclusion, five key

447    ARGs were closely associated with breast cancer progression.

448

449    In this study, we initially examined the prognostic values and expression profiles of ARGs in

450    patients with breast cancer. However, our study has some limitations. We did not perform a

451 prognostic analysis of breast cancer subtypes, and the laboratory data did not comprise the key

452 ARG protein levels in clinical samples and cell lines. We leave these questions to future studies

453 and further investigate the molecular mechanisms of key ARGs in breast cancer.

454

## Conclusions

456

457 This study primarily aimed to determine the critical role of ARGs in breast cancer progression

458 and search for valuable diagnostic and prognostic markers. The findings revealed that the ARG-

459 based predictive model could accurately predict the survival of patients with breast cancer. We

460 also investigated the possible BPs of ARGs and elucidated expression features and potential

461 prognostic value for each ARG. Our data emphasise the significant roles of ARGs in breast

462 cancer progression and provide novel research ideas for anoikis resistance.

463

## Acknowledgements

465 We are grateful to The Fang Zheng laboratory of Jinan University for instrument support and

466 assistance.

467

## Funding

469 This work was supported by the National Natural Science Foundation of China (nos. 82074430,

470 81803979, and 81673979), Guangdong Basic and Applied Basic Research Foundation, China

471 (no. 2022A1515011674), the Natural Science Foundation of Guangdong Province, China (nos.

472 2018A030313393 and 2016A030313114), Science and Technology Program of Guangzhou,

473 China (nos. 201803010051, 201707010245, and 201704020117) and the Fourth Batch of TCM

474 Clinical Outstanding Talent Program of China (no. 444258).

475

## Legends

477

478 **Figure 1. Flowchart of the study.**

479 TCGA-BRCA and BRCA datasets were obtained from TCGA and GEO databases, respectively.

480  LASSO analysis was constructed using 100 ARGs based on TCGA-BRCA dataset, and we
481  screened 12 ARGs with possible prognostic values. Subsequently, all samples from TCGA-
482  BRCA and BRCA datasets were divided into high- and low-risk groups based on the prognostic
483  model's median risk score. Each ARG was compared between the two risk groups, and 11
484  significant differentially expressed ARGs with the same expression trends in both datasets were
485  identified. Next, GO and KEGG enrichment analyses and PPI network construction were
486  performed using the 11 ARGs. We further performed GSEA and GSVA in TCGA-BRCA dataset
487  and obtained five key ARGs by plotting KM survival curves and ROC curves (AUC >0.6) for
488  each ARG based on TCGA-BRCA dataset. Finally, univariate and multivariate Cox regression
489  analyses were performed for key ARGs, and qPCR was used to detect gene expression qPCR in
490  different cell lines. Results of ARG protein expression with an HR of >1 were downloaded from
491  the HPA database.
492  TCGA: The Cancer Genome Atlas; BRCA: breast invasive carcinoma; GEO: Gene Expression
493  Omnibus; LASSO: least absolute shrinkage and selection operator; ARGs: anoikis-related genes;
494  GO: Gene Ontology; KEGG: Kyoto Encyclopaedia of Genes and Genomes; PPI: protein-protein
495  interaction; GSEA: gene set enrichment analysis; GSVA: gene set variation analysis; KM:
496  Kaplan–Meier curve; ROC: receiver operating characteristic curve; AUC: area under the curve;
497  qPCR: quantitative polymerase chain reaction; HR: hazard ratio; HPA: human protein atlas.
498
499  **Figure 2. Prognostic model establishment and DEG screening between the high- and low-**
500  **risk groups in breast cancer.**
501  (A) A LASSO prognostic model constructed using 100 ARGs. (B-C) Variable trajectories graph
502  and risk factor graph of the prognostic model. (D-E) Volcanic maps of the DEGs between the
503  two risk groups based on TCGA-BRCA and BRCA datasets, respectively. (F-G) Heat maps of
504  ARG expression between the two risk groups in TCGA-BRCA and BRCA datasets, respectively.
505  (H-I) Box plots comparing each ARG between the two risk groups in TCGA-BRCA and BRCA
506  datasets. ns, p ≥0.05, *, p <0.05, **, p <0.01, ***, p <0.001, and p <0.05 were considered
507  statistically significant. DEGs: differentially expressed genes; LASSO: least absolute shrinkage
508  and selection operator; TCGA: The Cancer Genome Atlas; BRCA: breast invasive carcinoma;
509  ARGs: anoikis-related genes.
510  **Figure 3. GSEA and GSVA**

511  (A) Four biological processes of GSEA for TCGA-BRCA dataset. Enrichment analysis of the

512  pre-Notch expression and processing (B), signalling by Notch (C), TCF dependent signalling in

513  response to Wnt (D), signalling by Wnt (E) in TCGA-BRCA dataset. (F) Heat maps of GSVA

514  for TCGA-BRCA dataset. (G) Box plots of significantly differently enriched hallmark genes

515  between the two risk groups. ***, $p < 0.001$ and $p < 0.05$ were considered statistically significant.

516  GSEA: gene set enrichment analysis; TCGA: The Cancer Genome Atlas; BRCA: breast invasive

517  carcinoma; TCF: T cell factor; GSVA: gene set variation analysis.

518

519  **Figure 4. KM survival curves and ROC curves**

520  KM curves of BIRC3 (A), CD24 (B), IVL (C), KRT15 (D), MIA (E), NDRG1 (F), NOS2 (G),

521  SERPINA1 (H), and TP63 (I). ROC curves of CD24 (J), KRT15 (K), MIA (L), NDRG1 (M),

522  and TP63 (N). KM: Kaplan–Meier; BIRC3: baculoviral IAP repeat-containing protein 3; CD24:

523  cluster of differentiation 24; IVL: involucrin; MIA: melanoma-inhibiting activity; KRT15,

524  keratin 15; NDRG1: N-Myc downstream regulated 1; NOS2: nitric oxide synthase 2; SERPINA1:

525  serpin family A member 1; TP63: tumour protein p63; ROC: receiver operating characteristic.

526

527  **Figure 5. Analysis of clinical correlation with the prognostic models.**

528  (A-B) Forest plots (A) and nomograms (B) of the multivariate Cox regression model for key

529  ARGs. (C-E) Calibration curves of the 2- (C), 3- (D), and 4-year (E) survival for nomograms of

530  multivariate Cox regression. Black lines at the top correspond to the distributions of predicted

531  survival probabilities of samples. (F-H) 2- (F), 3- (G), and 4-year (H) DCA of multivariate Cox

532  regression. ARGs: anoikis-related genes; DCA: decision curve analysis.

533

534  **Figure 6. mRNA levels of ARGs in MCF-10A, MDA-MB 231, HCC1954, and MCF-7 cell**

535  **lines.**

536  mRNA levels of CD24 (A), KRT15 (B), MIA (C), NDRG1 (D), and TP63 (E) determined by

537  qPCR. The mean ± SD is shown. * $p < 0.05$.

538  mRNA: messenger ribonucleic acid; ARGs: anoikis-related genes; MCF: Michigan Cancer

539  Foundation; CD24: cluster of differentiation 24; KRT15: keratin 15; MIA: melanoma-inhibiting

540  activity; NDRG1: N-Myc downstream regulated 1; TP63: tumour protein p63; qPCR:

541  quantitative polymerase chain reaction; SD: standard deviation.

542

**Figure S1. Box plots and PCA plots displayed the BRCA dataset before and after the**

**elimination of batch effects.** (A-B) Box plots displayed the BRCA dataset before (A) and after

the elimination of batch effects (B). (C-D) PCA plots displayed the BRCA dataset before (C) and

after the elimination of batch effects (D).

BRCA: breast invasive carcinoma; PCA: principal component analysis.

548

**Figure S2. Gene enrichment analysis and PPI network**

(A-B) The histogram and network diagrams indicate GO and KEGG analyses of 11 ARGs. Red

and blue dots represent genes and pathways, respectively. (C) Chordal graph of GO and KEGG

analyses of ARGs combined with logFC. (D) PPI network of ARGs constructed using Cytoscape

based on the STRING database. (E) PPI network of ARGs and related genes based on

GeneMANIA.

PPI: protein-protein interaction; GO: Gene Ontology; KEGG: Kyoto Encyclopaedia of Genes

and Genomes; ARGs: anoikis-related genes; STRING: Search Tool for the Retrieval of

Interacting Genes/Proteins.

558

**Figure S3. IHC results of CD24 and NDRG1 from the HPA database**

(A-B) CD24 protein levels in breast cancer tissue (A) and normal breast tissue (B). (C-D)

NDRG1 protein levels in breast cancer tissue (C) and normal breast tissue (D).

IHC: immunohistochemical; CD24: cluster of differentiation 24; NDRG1: N-Myc downstream

regulated 1; BRCA: breast invasive carcinoma; HPA: human protein atlas.

564

# References

566

[1] H. Sung, J. Ferlay, R.L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA: a cancer journal for clinicians 71 (2021) 209-249. 10.3322/caac.21660.

[2] S. Loibl, P. Poortmans, M. Morrow, C. Denkert, G. Curigliano, Breast cancer, Lancet (London, England) 397 (2021) 1750-1769. 10.1016/s0140-6736(20)32381-3.

[3] G.P. Gupta, J. Massagué, Cancer metastasis: building a framework, Cell 127 (2006) 679-695. 10.1016/j.cell.2006.11.001.

[4] S. Valastyan, R.A. Weinberg, Tumor metastasis: molecular insights and evolving paradigms, Cell 147 (2011) 275-292. 10.1016/j.cell.2011.09.024.

[5] D. Fanfone, Z. Wu, J. Mammi, K. Berthenet, D. Neves, K. Weber, A. Halaburkova, F. Virard,

578 F. Bunel, C. Jamard, H. Hernandez-Vargas, S.W.G. Tait, A. Hennino, G. Ichim, Confined
579 migration promotes cancer metastasis through resistance to anoikis and increased invasiveness,
580 Elife 11 (2022). 10.7554/eLife.73150.

581 [6] A.P. Gilmore, Anoikis, Cell death and differentiation 12 Suppl 2 (2005) 1473-1477.
582 10.1038/sj.cdd.4401723.

583 [7] F.O. Adeshakin, A.O. Adeshakin, L.O. Afolabi, D. Yan, G. Zhang, X. Wan, Mechanisms for
584 Modulating Anoikis Resistance in Cancer and the Relevance of Metabolic Reprogramming,
585 Front Oncol 11 (2021) 626577. 10.3389/fonc.2021.626577.

586 [8] J. Wang, Z. Luo, L. Lin, X. Sui, L. Yu, C. Xu, R. Zhang, Z. Zhao, Q. Zhu, B. An, Q. Wang, B.
587 Chen, E.L. Leung, Q. Wu, Anoikis-Associated Lung Cancer Metastasis: Mechanisms and
588 Therapies, Cancers 14 (2022). 10.3390/cancers14194791.

589 [9] J. Song, Y. Liu, F. Liu, L. Zhang, G. Li, C. Yuan, C. Yu, X. Lu, Q. Liu, X. Chen, H. Liang, Z.
590 Ding, B. Zhang, The 14-3-3σ protein promotes HCC anoikis resistance by inhibiting EGFR
591 degradation and thereby activating the EGFR-dependent ERK1/2 signaling pathway,
592 Theranostics 11 (2021) 996-1015. 10.7150/thno.51646.

593 [10] S. Du, Z. Yang, X. Lu, S. Yousuf, M. Zhao, W. Li, J. Miao, X. Wang, H. Yu, X. Zhu, H. Chen,
594 L. Shi, E. Xu, X. Xia, W. Guan, Anoikis resistant gastric cancer cells promote angiogenesis and
595 peritoneal metastasis through C/EBPβ-mediated PDGFB autocrine and paracrine signaling,
596 Oncogene 40 (2021) 5764-5779. 10.1038/s41388-021-01988-y.

597 [11] A. Tajbakhsh, M. Rivandi, S. Abedini, A. Pasdar, A. Sahebkar, Regulators and mechanisms
598 of anoikis in triple-negative breast cancer (TNBC): A review, Critical reviews in
599 oncology/hematology 140 (2019) 17-27. 10.1016/j.critrevonc.2019.05.009.

600 [12] L. Jin, J. Chun, C. Pan, G.N. Alesi, D. Li, K.R. Magliocca, Y. Kang, Z.G. Chen, D.M. Shin,
601 F.R. Khuri, J. Fan, S. Kang, Phosphorylation-mediated activation of LDHA promotes cancer cell
602 invasion and tumour metastasis, Oncogene 36 (2017) 3797-3806. 10.1038/onc.2017.6.

603 [13] X. Liu, S. Chipurupalli, P. Jiang, M. Tavasoli, B.H. Yoo, M. McPhee, S. Mazinani, G.
604 Francia, R.S. Kerbel, K.V. Rosen, ErbB2/Her2-dependent downregulation of a cell death-
605 promoting protein BLNK in breast cancer cells is required for 3D breast tumor growth, Cell
606 Death Dis 13 (2022) 687. 10.1038/s41419-022-05117-9.

607 [14] A. Colaprico, T.C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T.S. Sabedot, T.M.
608 Malta, S.M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks:
609 an R/Bioconductor package for integrative analysis of TCGA data, Nucleic acids research 44
610 (2016) e71. 10.1093/nar/gkv1507.

611 [15] M.J. Goldman, B. Craft, M. Hastie, K. Repečka, F. McDade, A. Kamath, A. Banerjee, Y. Luo,
612 D. Rogers, A.N. Brooks, J. Zhu, D. Haussler, Visualizing and interpreting cancer genomics data
613 via the Xena platform, Nat Biotechnol 38 (2020) 675-678. 10.1038/s41587-020-0546-8.

614 [16] M.E. Ritchie, B. Phipson, D. Wu, Y. Hu, C.W. Law, W. Shi, G.K. Smyth, limma powers
615 differential expression analyses for RNA-sequencing and microarray studies, Nucleic acids
616 research 43 (2015) e47. 10.1093/nar/gkv007.

617 [17] C. Clarke, S.F. Madden, P. Doolan, S.T. Aherne, H. Joyce, L. O'Driscoll, W.M. Gallagher,
618 B.T. Hennessy, M. Moriarty, J. Crown, S. Kennedy, M. Clynes, Correlating transcriptional
619 networks to breast cancer survival: a large-scale coexpression analysis, Carcinogenesis 34
620 (2013) 2300-2308. 10.1093/carcin/bgt208.

621 [18] S.H. Cheng, T.T. Huang, Y.H. Cheng, T.B.K. Tan, C.F. Horng, Y.A. Wang, N.S. Brian, L.S.
622 Shih, B.L. Yu, Validation of the 18-gene classifier as a prognostic biomarker of distant
623 metastasis in breast cancer, PloS one 12 (2017) e0184372. 10.1371/journal.pone.0184372.

624 [19] K.J. Kao, K.M. Chang, H.C. Hsu, A.T. Huang, Correlation of microarray-based breast
625 cancer molecular subtypes and clinical outcomes: implications for treatment optimization, BMC
626 Cancer 11 (2011) 143. 10.1186/1471-2407-11-143.

627 [20] T. Barrett, S.E. Wilhite, P. Ledoux, C. Evangelista, I.F. Kim, M. Tomashevsky, K.A. Marshall,
628 K.H. Phillippy, P.M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C.L. Robertson, N.

629 Serova, S. Davis, A. Soboleva, NCBI GEO: archive for functional genomics data sets--update,
630 Nucleic acids research 41 (2013) D991-995. 10.1093/nar/gks1193.
631 [21] S. Davis, P.S. Meltzer, GEOquery: a bridge between the Gene Expression Omnibus (GEO)
632 and BioConductor, Bioinformatics 23 (2007) 1846-1847. 10.1093/bioinformatics/btm254.
633 [22] A. Liberzon, C. Birger, H. Thorvaldsdóttir, M. Ghandi, J.P. Mesirov, P. Tamayo, The
634 Molecular Signatures Database (MSigDB) hallmark gene set collection, Cell Syst 1 (2015) 417-
635 425. 10.1016/j.cels.2015.12.004.
636 [23] G. Stelzer, N. Rosen, I. Plaschkes, S. Zimmerman, M. Twik, S. Fishilevich, T.I. Stein, R.
637 Nudel, I. Lieder, Y. Mazor, S. Kaplan, D. Dahary, D. Warshawsky, Y. Guan-Golan, A. Kohn, N.
638 Rappaport, M. Safran, D. Lancet, The GeneCards Suite: From Gene Data Mining to Disease
639 Genome Sequence Analyses, Curr Protoc Bioinformatics 54 (2016) 1.30.31-31.30.33.
640 10.1002/cpbi.5.
641 [24] J.T. Leek, W.E. Johnson, H.S. Parker, A.E. Jaffe, J.D. Storey, The sva package for
642 removing batch effects and other unwanted variation in high-throughput experiments,
643 Bioinformatics 28 (2012) 882-883. 10.1093/bioinformatics/bts034.
644 [25] Gene Ontology Consortium: going forward, Nucleic acids research 43 (2015) D1049-1056.
645 10.1093/nar/gku1179.
646 [26] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic acids
647 research 28 (2000) 27-30. 10.1093/nar/28.1.27.
648 [27] G. Yu, L.G. Wang, Y. Han, Q.Y. He, clusterProfiler: an R package for comparing biological
649 themes among gene clusters, Omics : a journal of integrative biology 16 (2012) 284-287.
650 10.1089/omi.2011.0118.
651 [28] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic,
652 N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C.V. Mering, STRING v11: protein-protein
653 association networks with increased coverage, supporting functional discovery in genome-wide
654 experimental datasets, Nucleic acids research 47 (2019) D607-d613. 10.1093/nar/gky1131.
655 [29] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B.
656 Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of
657 biomolecular interaction networks, Genome Res 13 (2003) 2498-2504. 10.1101/gr.1239303.
658 [30] M. Franz, H. Rodriguez, C. Lopes, K. Zuberi, J. Montojo, G.D. Bader, Q. Morris,
659 GeneMANIA update 2018, Nucleic acids research 46 (2018) W60-w64. 10.1093/nar/gky311.
660 [31] A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A.
661 Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis:
662 a knowledge-based approach for interpreting genome-wide expression profiles, Proc Natl Acad
663 Sci U S A 102 (2005) 15545-15550. 10.1073/pnas.0506580102.
664 [32] T. Tataranni, C. Piccoli, Dichloroacetate (DCA) and Cancer: An Overview towards Clinical
665 Applications, Oxid Med Cell Longev 2019 (2019) 8201079. 10.1155/2019/8201079.
666 [33] K. Colwill, S. Gräslund, A roadmap to generate renewable protein binders to the human
667 proteome, Nat Methods 8 (2011) 551-558. 10.1038/nmeth.1607.
668 [34] Y. Liang, H. Zhang, X. Song, Q. Yang, Metastatic heterogeneity of breast cancer: Molecular
669 mechanism and potential therapeutic targets, Seminars in cancer biology 60 (2020) 14-27.
670 10.1016/j.semcancer.2019.08.012.
671 [35] J. Liu, G. Lu, C. Liang, Y. Tian, Z. Jiang, Roles of anoikis in colorectal cancer therapy and
672 the assessment of anoikis-regulatory molecules as therapeutic targets, Pathol Res Pract 241
673 (2022) 154256. 10.1016/j.prp.2022.154256.
674 [36] R. Wang, I. Godet, Y. Yang, S. Salman, H. Lu, Y. Lyu, Q. Zuo, Y. Wang, Y. Zhu, C. Chen, J.
675 He, D.M. Gilkes, G.L. Semenza, Hypoxia-inducible factor-dependent ADAM12 expression
676 mediates breast cancer invasion and metastasis, Proc Natl Acad Sci U S A 118 (2021).
677 10.1073/pnas.2020490118.
678 [37] H.M. Wobma, M.A. Tamargo, S. Goeta, L.M. Brown, R. Duran-Struuck, G. Vunjak-
679 Novakovic, The influence of hypoxia and IFN-γ on the proteome and metabolome of therapeutic

680  mesenchymal stem cells, Biomaterials 167 (2018) 226-234. 10.1016/j.biomaterials.2018.03.027.

681  [38] K.O. Osuala, K. Ji, R.R. Mattingly, B.F. Sloane, Breast Cancer: Proteolysis and Migration,
682  Adv Exp Med Biol 1152 (2019) 401-411. 10.1007/978-3-030-20301-6_21.

683  [39] M.Y. Yang, C.J. Chang, L.Y. Chen, Blue light induced reactive oxygen species from flavin
684  mononucleotide and flavin adenine dinucleotide on lethality of HeLa cells, J Photochem
685  Photobiol B 173 (2017) 325-332. 10.1016/j.jphotobiol.2017.06.014.

686  [40] M. Tan, M. Asad, V. Heong, M.K. Wong, T.Z. Tan, J. Ye, K.T. Kuay, J.P. Thiery, C. Scott,
687  R.Y. Huang, The FZD7-TWIST1 axis is responsible for anoikis resistance and tumorigenesis in
688  ovarian carcinoma, Mol Oncol 13 (2019) 757-780. 10.1002/1878-0261.12425.

689  [41] O.J. Kwon, J.M. Valdez, L. Zhang, B. Zhang, X. Wei, Q. Su, M.M. Ittmann, C.J. Creighton, L.
690  Xin, Increased Notch signalling inhibits anoikis and stimulates proliferation of prostate luminal
691  epithelial cells, Nature communications 5 (2014) 4416. 10.1038/ncomms5416.

692  [42] K.G. Leong, K. Niessen, I. Kulic, A. Raouf, C. Eaves, I. Pollet, A. Karsan, Jagged1-
693  mediated Notch activation induces epithelial-to-mesenchymal transition through Slug-induced
694  repression of E-cadherin, The Journal of experimental medicine 204 (2007) 2935-2948.
695  10.1084/jem.20071082.

696  [43] A. Magklara, T.J. Brown, E.P. Diamandis, Characterization of androgen receptor and
697  nuclear receptor co-regulator expression in human breast cancer cell lines exhibiting differential
698  regulation of kallikreins 2 and 3, International journal of cancer 100 (2002) 507-514.
699  10.1002/ijc.10520.

700  [44] H. Cheng, P. Liu, Z.C. Wang, L. Zou, S. Santiago, V. Garbitt, O.V. Gjoerup, J.D. Iglehart, A.
701  Miron, A.L. Richardson, W.C. Hahn, J.J. Zhao, SIK1 couples LKB1 to p53-dependent anoikis
702  and suppresses metastasis, Science signaling 2 (2009) ra35. 10.1126/scisignal.2000369.

703  [45] T.L. Ng, G. Leprivier, M.D. Robertson, C. Chow, M.J. Martin, K.R. Laderoute, E. Davicioni,
704  T.J. Triche, P.H. Sorensen, The AMPK stress response pathway mediates anoikis resistance
705  through inhibition of mTOR and suppression of protein synthesis, Cell death and differentiation
706  19 (2012) 501-510. 10.1038/cdd.2011.119.

707  [46] A.A. Barkal, R.E. Brewer, M. Markovic, M. Kowarsky, S.A. Barkal, B.W. Zaro, V. Krishnan, J.
708  Hatakeyama, O. Dorigo, L.J. Barkal, I.L. Weissman, CD24 signalling through macrophage
709  Siglec-10 is a target for cancer immunotherapy, Nature 572 (2019) 392-396. 10.1038/s41586-
710  019-1456-0.

711  [47] Y.H. Li, X. Sun, H.B. Wang, Role of CD24 in anoikis resistance of ovarian cancer cells, J
712  Huazhong Univ Sci Technolog Med Sci 35 (2015) 390-396. 10.1007/s11596-015-1443-0.

713  [48] Y.H. Ni, X. Zhao, W. Wang, CD24, A Review of its Role in Tumor Diagnosis, Progression
714  and Therapy, Curr Gene Ther 20 (2020) 109-126. 10.2174/1566523220666200623170738.

715  [49] W. Chen, C. Miao, KRT15 promotes colorectal cancer cell migration and invasion through
716  β-catenin/MMP-7 signaling pathway, Med Oncol 39 (2022) 68. 10.1007/s12032-021-01619-2.

717  [50] L.Y. Chong, P.Y. Cheok, W.J. Tan, A.A. Thike, G. Allen, M.K. Ang, A.S. Ooi, P. Tan, B.T.
718  Teh, P.H. Tan, Keratin 15, transcobalamin I and homeobox gene Hox-B13 expression in breast
719  phyllodes tumors: novel markers in biological classification, Breast cancer research and
720  treatment 132 (2012) 143-151. 10.1007/s10549-011-1555-6.

721  [51] J. Schmidt, A.K. Bosserhoff, Processing of MIA protein during melanoma cell migration,
722  International journal of cancer 125 (2009) 1587-1594. 10.1002/ijc.24508.

723  [52] V. Joshi, S.R. Lakhani, A.E. McCart Reed, NDRG1 in Cancer: A Suppressor, Promoter, or
724  Both?, Cancers 14 (2022). 10.3390/cancers14235739.

725  [53] V. Gatti, L. Bongiorno-Borbone, C. Fierro, M. Annicchiarico-Petruzzelli, G. Melino, A.
726  Peschiaroli, p63 at the Crossroads between Stemness and Metastasis in Breast Cancer,
727  International journal of molecular sciences 20 (2019). 10.3390/ijms20112683.

728

List of description and expression difference of anoikis-related genes of LASSO model.

1 **Table 1:List of description and expression difference of anoikis-related genes of LASSO**

| Gene Symbol | Description | logFC | P.Value | adj.P |
|---|---|---|---|---|
| BIRC3 | Baculoviral IAP Repeat Containing 3 | -0.551334371 | 1.89883E-10 | 4.32641E-09 |
| CD24 | CD24 Molecule | 0.406051661 | 0.000275766 | 0.001024153 |
| FGFR1 | Fibroblast Growth Factor Receptor 1 | 0.19873266 | 0.008331891 | 0.01969228 |
| IVL | Involucrin | 0.606150008 | 0.000322142 | 0.001169971 |
| KRT15 | Keratin 15 | -1.257950408 | 3.6222E-15 | 3.52105E-13 |
| L1CAM | L1 Cell Adhesion Molecule | 0.319550324 | 0.019858044 | 0.041563644 |
| MIA | MIA SH3 Domain Containing | -1.265363973 | 5.8105E-14 | 4.00434E-12 |
| NDRG1 | N-Myc Downstream Regulated 1 | 0.371969381 | 2.8489E-06 | 1.86671E-05 |
| NOS2 | Nitric Oxide Synthase 2 | 0.482118761 | 4.97052E-12 | 1.8252E-10 |
| PTPN3 | Protein Tyrosine Phosphatase Non-Receptor Type 3 | 0.034313831 | 0.482891244 | 0.581437763 |
| SERPINA1 | Serpin Family A Member 1 | -0.677585034 | 4.08221E-09 | 6.07581E-08 |
| TP63 | Tumor Protein P63 | -0.865989129 | 2.1598E-10 | 4.82553E-09 |

2 **model.**

5

GO and KEGG enrichment analysis results of anoikis-related genes.

1    **Table 2. GO and KEGG enrichment analysis results of anoikis-related genes.**

| Ontology | ID | Description | GeneRatio | BgRatio | pvalue | p.adjust | qvalue |
|---|---|---|---|---|---|---|---|
| BP | GO:0030216 | keratinocyte differentiation | 3/11 | 305/18670 | 6.46e-04 | 0.096 | 0.057 |
| BP | GO:0001666 | response to hypoxia | 3/11 | 359/18670 | 0.001 | 0.096 | 0.057 |
| BP | GO:0045861 | negative regulation of proteolysis | 3/11 | 363/18670 | 0.001 | 0.096 | 0.057 |
| MF | GO:0004866 | endopeptidase inhibitor activity | 2/11 | 175/17697 | 0.005 | 0.097 | 0.059 |
| MF | GO:0030414 | peptidase inhibitor activity | 2/11 | 182/17697 | 0.005 | 0.097 | 0.059 |
| MF | GO:0010181 | FMN binding | 1/11 | 16/17697 | 0.010 | 0.099 | 0.061 |
| KEGG | hsa05222 | Small cell lung cancer | 2/7 | 92/8076 | 0.003 | 0.071 | 0.064 |
| KEGG | hsa05145 | Toxoplasmosis | 2/7 | 112/8076 | 0.004 | 0.071 | 0.064 |

2    **GO, Gene Ontology. BP, biological process. CC, cellular component. MF, molecular function. KEGG,**

3    **Kyoto Encyclopedia of Genes and Genomes.**

GSEA analysis of dataset TCGA-BRCA.

1  **Table 3. GSEA analysis of dataset TCGA-BRCA.**

| Description | setSize | enrichmentScore | NES | pvalue | p.adjust |
|---|---|---|---|---|---|
| condensation of prophase chromosomes | 72 | 0.648142653 | 2.609133444 | 0.000384468 | 0.007527598 |
| DNA methylation | 63 | 0.64924649 | 2.550957964 | 0.000368053 | 0.007527598 |
| deposition of new cenpa containing nucleosomes at the centromere | 73 | 0.62272815 | 2.511036721 | 0.000387747 | 0.007527598 |
| meiotic recombination | 85 | 0.597551004 | 2.469681726 | 0.000414594 | 0.007527598 |
| SIRT1 negatively regulates rrna expression | 66 | 0.615267088 | 2.433992405 | 0.000374532 | 0.007527598 |
| PRC2 methylates histones and DNA | 71 | 0.599962935 | 2.413657315 | 0.000381825 | 0.007527598 |
| activated PKN1 stimulates transcription of AR androgen receptor regulated genes KLK2 and KLK3 | 65 | 0.598544909 | 2.36548707 | 0.000370508 | 0.007527598 |
| recognition and association of DNA glycosylase with site containing an | 56 | 0.603186125 | 2.318484147 | 0.000351 | 0.007527598 |

| | | | | | |
|---|---|---|---|---|---|
| affected purine | | | | | |
| DNA damage | | | | | |
| telomere stress | 80 | 0.563392337 | 2.306378423 | 0.000403551 | 0.007527598 |
| induced senescence | | | | | |
| mitotic prophase | 141 | 0.504722456 | 2.267243796 | 0.000542005 | 0.008140373 |
| nonhomologous end | 69 | 0.566496673 | 2.266201035 | 0.000376223 | 0.007527598 |
| joining NHEJ | | | | | |
| pre NOTCH | | | | | |
| expression and | 107 | 0.467738962 | 2.020954761 | 0.000458295 | 0.007527598 |
| processing | | | | | |
| signaling by NOTCH | 234 | 0.287501505 | 1.378143085 | 0.003220612 | 0.03396672 |
| TCF dependent | | | | | |
| signaling in response | 232 | 0.286479654 | 1.373337984 | 0.006405124 | 0.058235094 |
| to WNT | | | | | |
| signaling by WNT | 329 | 0.262444276 | 1.302165188 | 0.006944444 | 0.060142944 |

2    GSEA, Gene Set Enrichment Analysis. TCGA, the cancer genome atlas. BRCA, breast invasive carcinoma.

GSVA analysis of dataset TCGA-BRCA.

1 **Table 4. GSVA analysis of dataset TCGA-BRCA.**

| Ontology | logFC | AveExpr | t | P.Value | adj.P |
|---|---|---|---|---|---|
| HALLMARK_P53_PATHWAY | 0.119329283 | -0.031254621 | 8.399741679 | 1.39E-16 | 6.95E-15 |
| HALLMARK_IL6_JAK_STAT3_SIGNALING | 0.160554553 | -0.019901999 | 7.747293348 | 2.15E-14 | 5.38E-13 |
| HALLMARK_TNFA_SIGNALING_VIA_NFKB | 0.149491843 | -0.018780161 | 7.593211543 | 6.72E-14 | 1.12E-12 |
| HALLMARK_ALLOGRAFT_REJECTION | 0.165689567 | -0.01910031 | 7.335435252 | 4.33E-13 | 5.41E-12 |
| HALLMARK_G2M_CHECKPOINT | -0.160835941 | -0.030063227 | -7.23791024 | 8.62E-13 | 8.62E-12 |
| HALLMARK_KRAS_SIGNALING_DN | 0.077950216 | 0.027929351 | 7.179824705 | 1.30E-12 | 1.08E-11 |
| HALLMARK_INTERFERON_GAMMA_RESPONSE | 0.162439186 | -0.023700528 | 6.976246774 | 5.27E-12 | 3.76E-11 |
| HALLMARK_APOPTOSIS | 0.105082386 | -0.016017823 | 6.93050969 | 7.19E-12 | 4.49E-11 |
| HALLMARK_INFLAMMATORY_RESPONSE | 0.1292374 | -0.01320064 | 6.588259013 | 6.93E-11 | 3.85E-10 |
| HALLMARK_E2F_TARGETS | -0.160184208 | -0.027553265 | -6.55175209 | 8.78E-11 | 4.39E-10 |
| HALLMARK_MITOTIC_SPINDLE | -0.107442266 | -0.026580477 | -6.479749206 | 1.39E-10 | 6.33E-10 |

| | | | | | |
|---|---|---|---|---|---|
| HALLMARK_IL2_STAT5_SIGNALING | 0.10427956 | -0.023708254 | 6.38643486 | 2.52E-10 | 1.05E-09 |
| HALLMARK_INTERFERON_ALPHA_RESPONSE | 0.162750622 | -0.025587788 | 6.349727433 | 3.17E-10 | 1.13E-09 |
| HALLMARK_SPERMATOGENESIS | -0.079775897 | 0.024668073 | -6.349696603 | 3.17E-10 | 1.13E-09 |
| HALLMARK_COAGULATION | 0.102292767 | 0.017160336 | 6.17528166 | 9.33E-10 | 3.11E-09 |
| HALLMARK_ESTROGEN_RESPONSE_LATE | 0.09490817 | -0.013436064 | 6.131517793 | 1.22E-09 | 3.81E-09 |
| HALLMARK_COMPLEMENT | 0.103046183 | -0.015778926 | 6.011855981 | 2.50E-09 | 7.37E-09 |
| HALLMARK_XENOBIOTIC_METABOLISM | 0.075977134 | 0.000803129 | 5.709659661 | 1.46E-08 | 4.06E-08 |
| HALLMARK_MYC_TARGETS_V1 | -0.110050823 | -0.039521363 | -5.536657383 | 3.87E-08 | 9.83E-08 |
| HALLMARK_ESTROGEN_RESPONSE_EARLY | 0.098760419 | 0.003536364 | 5.533710477 | 3.93E-08 | 9.83E-08 |
| HALLMARK_KRAS_SIGNALING_UP | 0.093897872 | -0.006226696 | 5.43040401 | 6.94E-08 | 1.65E-07 |
| HALLMARK_WNT_BETA_CATENIN_SIGNALING | 0.081765544 | -0.029949865 | 4.613361613 | 4.44E-06 | 1.01E-05 |
| HALLMARK_MTORC1_SIGNALING | -0.08067193 | -0.030607917 | -4.531864271 | 6.50E-06 | 1.41E-05 |
| HALLMARK_MYC_TARGETS_V2 | -0.094920667 | -0.035952966 | -3.947820502 | 8.40E-05 | 0.000174914 |

| | | | | |
|---|---|---|---|---|
| HALLMARK_BILE_ACID_METABOLISM | 0.053809556 | 0.000674898 | 3.933254506 | 8.91E-05 | 0.000178277 |
| HALLMARK_PROTEIN_SECRETION | -0.068629916 | -0.00796288 | -3.623357806 | 0.000304306 | 0.000585204 |
| HALLMARK_HYPOXIA | 0.053946213 | -0.019883061 | 3.427859931 | 0.000631247 | 0.001139569 |
| HALLMARK_UNFOLDED_PROTEIN_RESPONSE | -0.053072207 | -0.042682246 | -3.424870074 | 0.000638159 | 0.001139569 |
| HALLMARK_APICAL_SURFACE | 0.05269076 | 0.000821367 | 3.364386226 | 0.00079404 | 0.001369035 |
| HALLMARK_FATTY_ACID_METABOLISM | 0.045000694 | -0.013810827 | 3.099805318 | 0.001986498 | 0.003302312 |
| HALLMARK_APICAL_JUNCTION | 0.049957108 | -0.006908797 | 3.090764305 | 0.002047433 | 0.003302312 |
| HALLMARK_ADIPOGENESIS | 0.041776924 | -0.029142429 | 2.799430327 | 0.005210284 | 0.008141069 |
| HALLMARK_MYOGENESIS | 0.044253317 | -0.00702594 | 2.695291931 | 0.007141694 | 0.010820749 |
| HALLMARK_GLYCOLYSIS | -0.038843193 | -0.042366566 | -2.663039667 | 0.007858759 | 0.011556999 |
| HALLMARK_NOTCH_SIGNALING | 0.047821813 | -0.018464182 | 2.570911437 | 0.010275841 | 0.014679772 |
| HALLMARK_PANCREAS_BETA_CELLS | 0.036594431 | 0.073863399 | 2.521151193 | 0.011839923 | 0.016444337 |
| HALLMARK_HEME_METABOLISM | 0.028007544 | -0.020449587 | 2.287809617 | 0.022340814 | 0.030190289 |
| HALLMARK_REACTIVE | 0.036597558 | -0.039232763 | 2.049238101 | 0.04067 | 0.05352 |

| _OXYGEN_SPECIES_PA | 9537 | 5707 |
| THWAY | | |

2   GSVA, Gene Set Variation Analysis. TCGA, the cancer genome atlas. BRCA, breast invasive carcinoma.

Patient Characteristics of BRCA patients in the TCGA datasets.

1 **Table 5. Patient Characteristics of BRCA patients in the TCGA datasets.**

| Characteristic | levels | Overall |
| --- | --- | --- |
| n | | 1083 |
| T stage, n (%) | T1 | 277 (25.6%) |
| | T2 | 629 (58.2%) |
| | T3 | 139 (12.9%) |
| | T4 | 35 (3.2%) |
| N stage, n (%) | N0 | 514 (48.3%) |
| | N1 | 358 (33.6%) |
| | N2 | 116 (10.9%) |
| | N3 | 76 (7.1%) |
| M stage, n (%) | M0 | 902 (97.8%) |
| | M1 | 20 (2.2%) |
| Pathologic stage, n (%) | Stage I | 181 (17.1%) |
| | Stage II | 619 (58.4%) |
| | Stage III | 242 (22.8%) |
| | Stage IV | 18 (1.7%) |
| Age, n (%) | <=60 | 601 (55.5%) |
| | >60 | 482 (44.5%) |

| | | |
|---|---|---|
| OS event, n (%) | Alive | 931 (86%) |
| | Dead | 152 (14%) |
| DSS event, n (%) | Alive | 978 (92%) |
| | Dead | 85 (8%) |
| PFI event, n (%) | Alive | 936 (86.4%) |
| | Dead | 147 (13.6%) |
| Age, median (IQR) | | 58 (48.5, 67) |

Cox regression to identify hub genes and clinical features associated with OS.

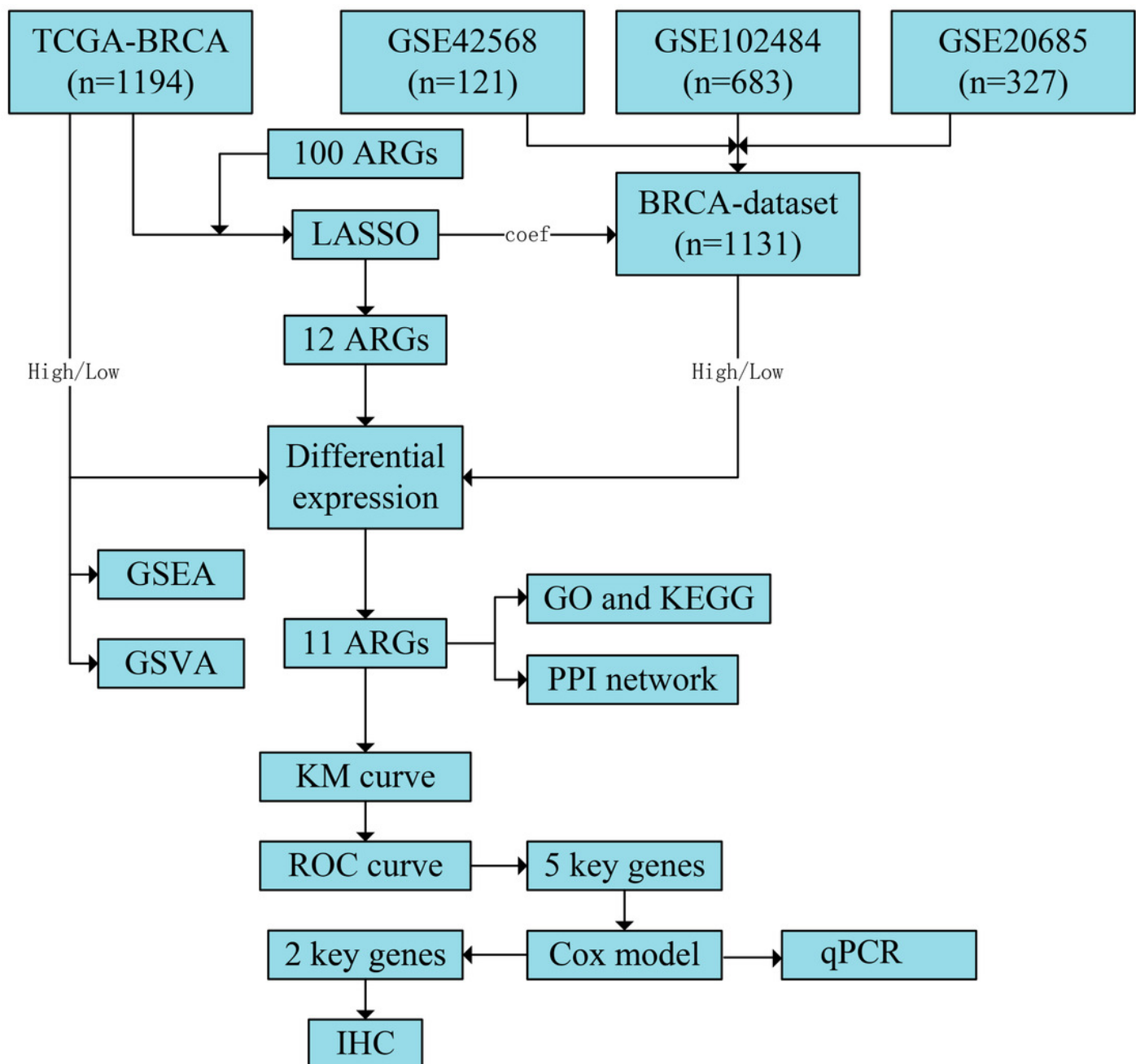
1 **Table 6. Cox regression to identify hub genes and clinical features associated with OS.**

| Characteristics | Total(N) | Univariate analysis | | Multivariate analysis | |
|---|---|---|---|---|---|
| | | Hazard ratio (95% CI) | P value | Hazard ratio (95% CI) | P value |
| CD24 | 1082 | | | | |
| Low | 540 | Reference | | | |
| High | 542 | 1.694 (1.217-2.359) | **0.002** | 1.583 (1.132-2.213) | **0.007** |
| KRT15 | 1082 | | | | |
| Low | 540 | Reference | | | |
| High | 542 | 0.606 (0.438-0.837) | **0.002** | 0.673 (0.468-0.969) | **0.033** |
| MIA | 1082 | | | | |
| Low | 540 | Reference | | | |
| High | 542 | 0.685 (0.497-0.944) | **0.021** | 0.836 (0.568-1.231) | 0.365 |
| NDRG1 | 1082 | | | | |
| Low | 540 | Reference | | | |
| High | 542 | 1.410 (1.023-1.944) | **0.036** | 1.409 (1.010-1.966) | **0.043** |
| TP63 | 1082 | | | | |
| Low | 540 | Reference | | | |
| High | 542 | 0.640 (0.464-0.883) | **0.007** | 0.798 (0.558-1.142) | 0.218 |

2 OS, overall survival.

# Figure 1

Flowchart of the study.

TCGA-BRCA and BRCA datasets were obtained from TCGA and GEO databases, respectively. LASSO analysis was constructed using 100 ARGs based on TCGA-BRCA dataset, and we screened 12 ARGs with possible prognostic values. Subsequently, all samples from TCGA-BRCA and BRCA datasets were divided into high- and low-risk groups based on the prognostic model's median risk score. Each ARG was compared between the two risk groups, and 11 significant differentially expressed ARGs with the same expression trends in both datasets were identified. Next, GO and KEGG enrichment analyses and PPI network construction were performed using the 11 ARGs. We further performed GSEA and GSVA in TCGA-BRCA dataset and obtained five key ARGs by plotting KM survival curves and ROC curves (AUC >0.6) for each ARG based on TCGA-BRCA dataset. Finally, univariate and multivariate Cox regression analyses were performed for key ARGs, and qPCR was used to detect gene expression qPCR in different cell lines. Results of ARG protein expression with an HR of >1 were downloaded from the HPA database. TCGA: The Cancer Genome Atlas; BRCA: breast invasive carcinoma; GEO: Gene Expression Omnibus; LASSO: least absolute shrinkage and selection operator; ARGs: anoikis-related genes; GO: Gene Ontology; KEGG: Kyoto Encyclopaedia of Genes and Genomes; PPI: protein-protein interaction; GSEA: gene set enrichment analysis; GSVA: gene set variation analysis; KM: Kaplan–Meier curve; ROC: receiver operating characteristic curve; AUC: area under the curve; qPCR: quantitative polymerase chain reaction; HR: hazard ratio; HPA: human protein atlas.

TCGA-BRCA (n=1194)

GSE42568 (n=121)

GSE102484 (n=683)

GSE20685 (n=327)

100 ARGs

LASSO —coef→ BRCA-dataset (n=1131)

12 ARGs

High/Low

High/Low

Differential expression

GSEA

GSVA

11 ARGs

GO and KEGG

PPI network

KM curve

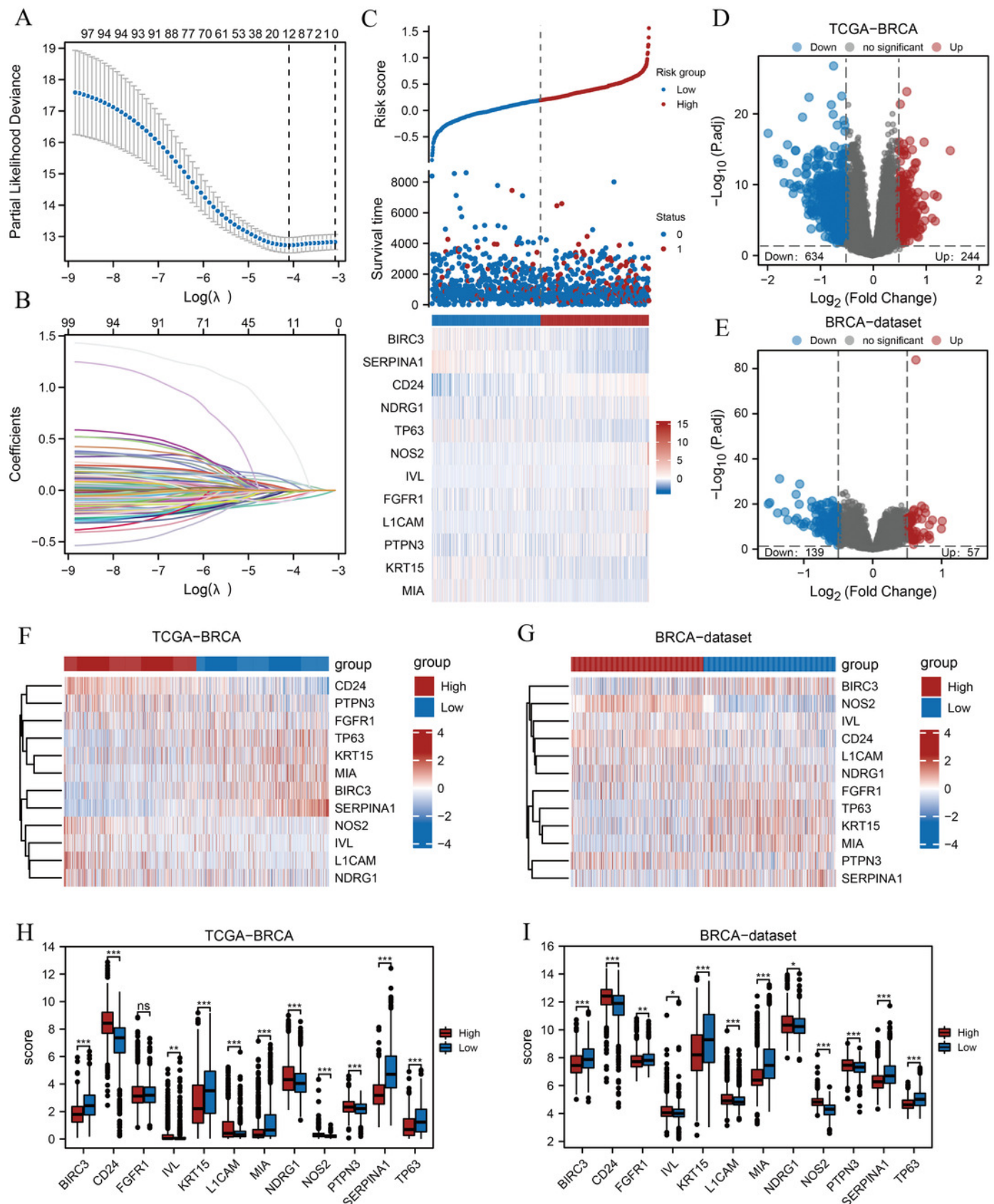
ROC curve → 5 key genes

2 key genes ← Cox model → qPCR

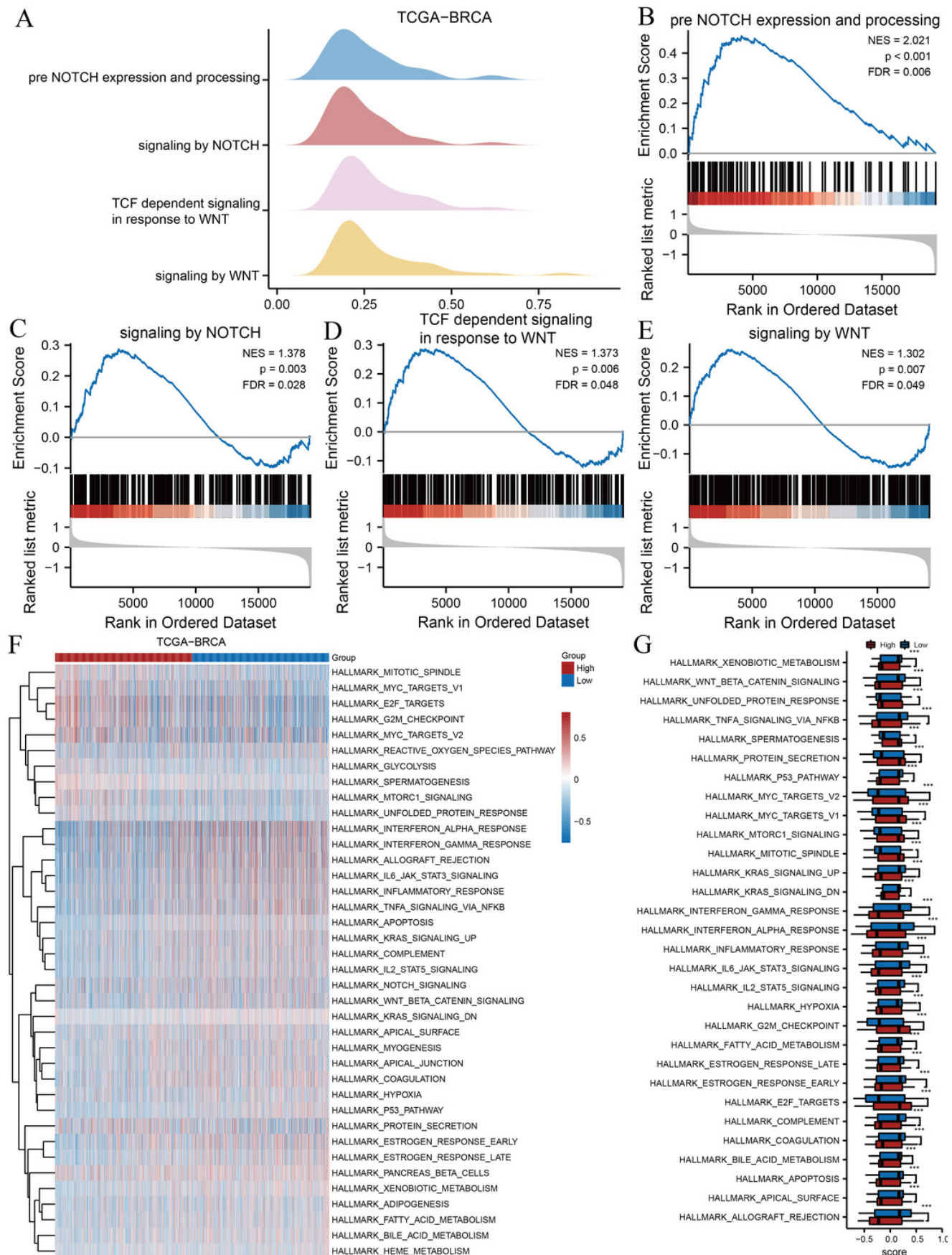
IHC

# Figure 2

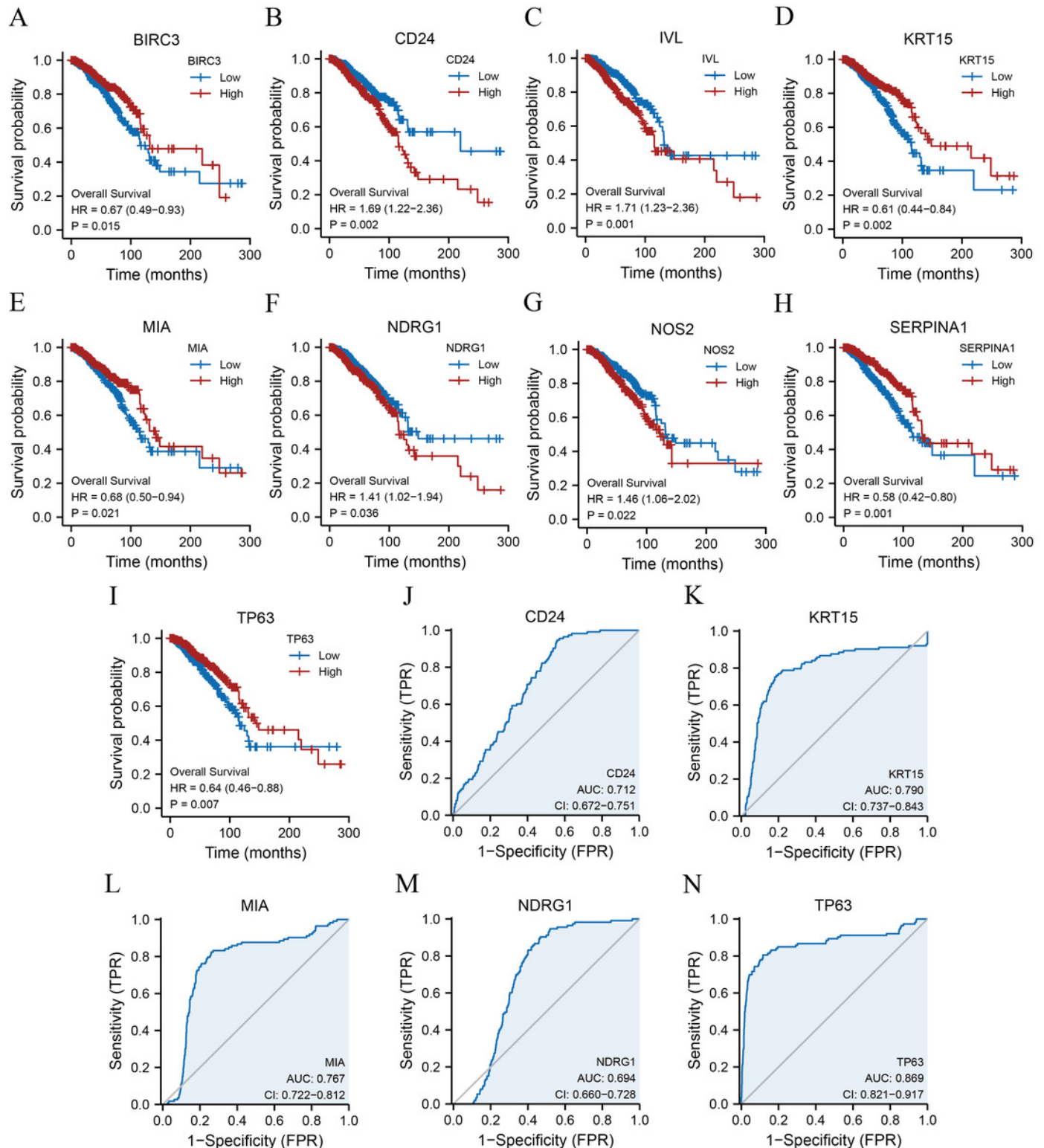
Prognostic model establishment and DEG screening between the high- and low-risk groups in breast cancer.

(A) A LASSO prognostic model constructed using 100 ARGs. (B-C) Variable trajectories graph and risk factor graph of the prognostic model. (D-E) Volcanic maps of the DEGs between the two risk groups based on TCGA-BRCA and BRCA datasets, respectively. (F-G) Heat maps of ARG expression between the two risk groups in TCGA-BRCA and BRCA datasets, respectively. (H-I) Box plots comparing each ARG between the two risk groups in TCGA-BRCA and BRCA datasets. ns, p ≥0.05, *, p <0.05, **, p <0.01, ***, p <0.001, and p <0.05 were considered statistically significant. DEGs: differentially expressed genes; LASSO: least absolute shrinkage and selection operator; TCGA: The Cancer Genome Atlas; BRCA: breast invasive carcinoma; ARGs: anoikis-related genes.

# Figure 3

GSEA and GSVA

(A) Four biological processes of GSEA for TCGA-BRCA dataset. Enrichment analysis of the pre-Notch expression and processing (B), signalling by Notch (C), TCF dependent signalling in response to Wnt (D), signalling by Wnt (E) in TCGA-BRCA dataset. (F) Heat maps of GSVA for TCGA-BRCA dataset. (G) Box plots of significantly differently enriched hallmark genes between the two risk groups. ***, p <0.001 and p <0.05 were considered statistically significant. GSEA: gene set enrichment analysis; TCGA: The Cancer Genome Atlas; BRCA: breast invasive carcinoma; TCF: T cell factor; GSVA: gene set variation analysis.

# Figure 4

KM survival curves and ROC curves

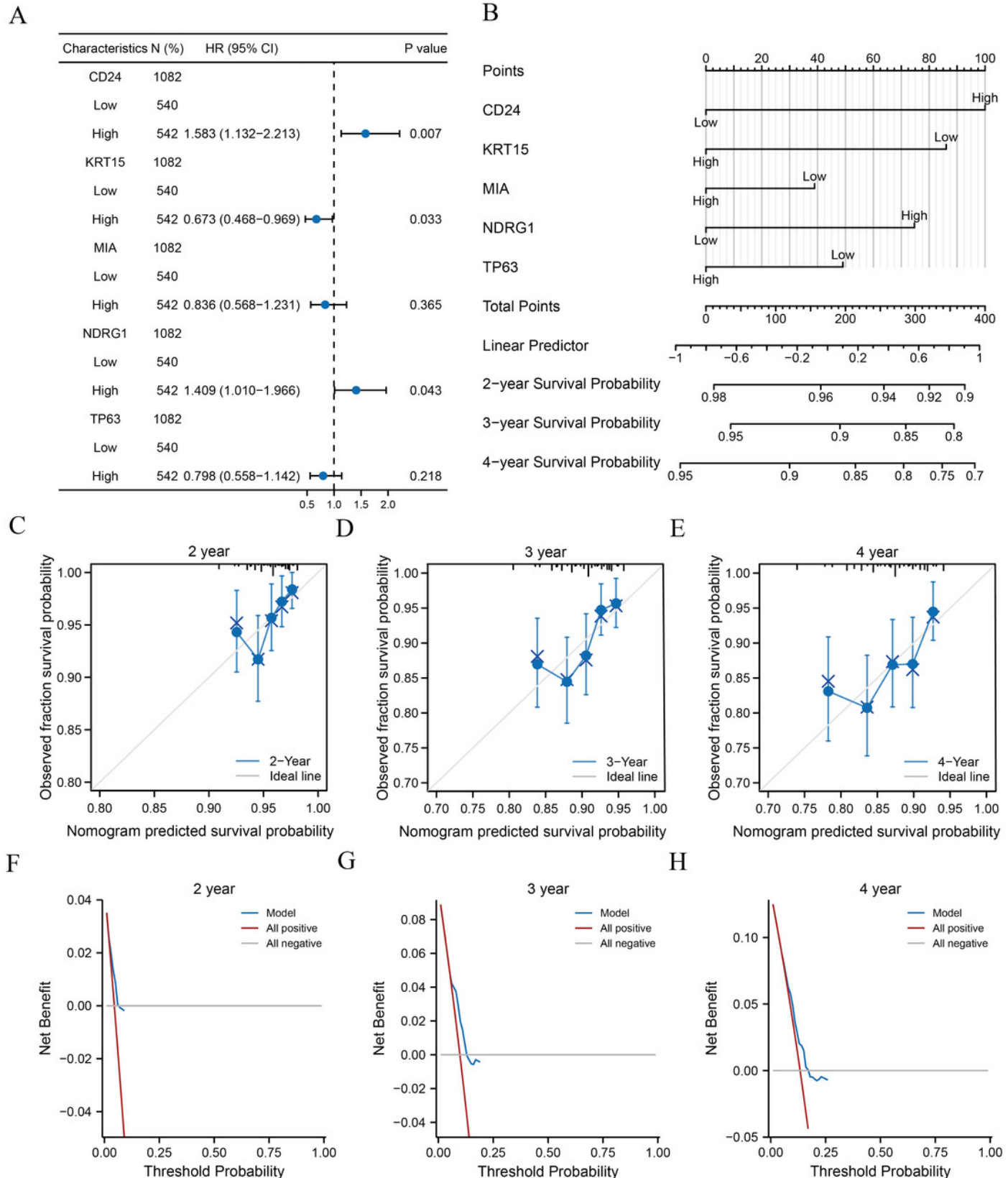
KM curves of BIRC3 (A), CD24 (B), IVL (C), KRT15 (D), MIA (E), NDRG1 (F), NOS2 (G), SERPINA1 (H), and TP63 (I). ROC curves of CD24 (J), KRT15 (K), MIA (L), NDRG1 (M), and TP63 (N). KM: Kaplan–Meier; BIRC3: baculoviral IAP repeat-containing protein 3; CD24: cluster of differentiation 24; IVL: involucrin; MIA: melanoma-inhibiting activity; KRT15, keratin 15; NDRG1: N-Myc downstream regulated 1; NOS2: nitric oxide synthase 2; SERPINA1: serpin family A member 1; TP63: tumour protein p63; ROC: receiver operating characteristic.

# Figure 5

Analysis of clinical correlation with the prognostic models.

(A-B) Forest plots (A) and nomograms (B) of the multivariate Cox regression model for key ARGs. (C-E) Calibration curves of the 2- (C), 3- (D), and 4-year (E) survival for nomograms of multivariate Cox regression. Black lines at the top correspond to the distributions of predicted survival probabilities of samples. (F-H) 2- (F), 3- (G), and 4-year (H) DCA of multivariate Cox regression. ARGs: anoikis-related genes; DCA: decision curve analysis.

A

| Characteristics | N (%) | HR (95% CI) | P value |
|---|---|---|---|
| CD24 | 1082 | | |
| Low | 540 | | |
| High | 542 | 1.583 (1.132–2.213) | 0.007 |
| KRT15 | 1082 | | |
| Low | 540 | | |
| High | 542 | 0.673 (0.468–0.969) | 0.033 |
| MIA | 1082 | | |
| Low | 540 | | |
| High | 542 | 0.836 (0.568–1.231) | 0.365 |
| NDRG1 | 1082 | | |
| Low | 540 | | |
| High | 542 | 1.409 (1.010–1.966) | 0.043 |
| TP63 | 1082 | | |
| Low | 540 | | |
| High | 542 | 0.798 (0.558–1.142) | 0.218 |

B

# Figure 6

mRNA levels of ARGs in MCF-10A, MDA-MB 231, HCC1954, and MCF-7 cell lines.

mRNA levels of CD24 (A), KRT15 (B), MIA (C), NDRG1 (D), and TP63 (E) determined by qPCR. The mean ± SD is shown. * p <0.05. mRNA: messenger ribonucleic acid; ARGs: anoikis-related genes; MCF: Michigan Cancer Foundation; CD24: cluster of differentiation 24; KRT15: keratin 15; MIA: melanoma-inhibiting activity; NDRG1: N-Myc downstream regulated 1; TP63: tumour protein p63; qPCR: quantitative polymerase chain reaction; SD: standard deviation.