



Prognostic model of invasive ductal carcinoma of the breast based on differentially expressed glycolysis-related genes

Xiaoping Li^{1,*}, Qihe Yu^{2,*}, Jishang Chen³, Hui Huang⁴, Zhuangsheng Liu⁵, Chengxing Wang¹, Yaoming He¹, Xin Zhang⁶, Weiwen Li⁷, Chao Li¹, Jinglin Zhao¹ and Wansheng Long⁵

¹Department of Gastrointestinal Surgery, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, Guangdong, China

²Department of Oncology, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, Guangdong, China

³Department of Breast Surgery, Yangjiang people's Hospital, Yangjiang, Guangdong, China

⁴Department of Breast Surgery, Jiangmen Maternity & Child Health Care Hospital, Jiangmen, Guangdong, China

⁵Department of Radiology, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, Guangdong, China

⁶Clinical Experimental Center, Jiangmen Key Laboratory of Clinical Biobanks and Translational Research, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, Guangdong, China

⁷Department of Breast and Thyroid Surgery, Affiliated Jiangmen Hospital of Sun Yat-sen University, Jiangmen, Guangdong, China

*These authors contributed equally to this work.

ABSTRACT

Background. Invasive ductal carcinoma (IDC) is a common pathological type of breast cancer that is characterized by high malignancy and rapid progression. Upregulation of glycolysis is a hallmark of tumor growth, and correlates with the progression of breast cancer. We aimed to establish a model to predict the prognosis of patients with breast IDC based on differentially expressed glycolysis-related genes (DEGRGs).

Methods. Transcriptome data and clinical data of patients with breast IDC were from The Cancer Genome Atlas (TCGA). Glycolysis-related gene sets and pathways were from the Molecular Signatures Database (MSigDB). DEGRGs were identified by comparison of tumor tissues and adjacent normal tissues. Univariate Cox regression and least absolute shrinkage and selection operator (LASSO) regression were used to screen for DEGRGs with prognostic value. A risk-scoring model based on DEGRGs related to prognosis was constructed. Receiver operating characteristic (ROC) analysis and calculation of the area under the curve (AUC) were used to evaluate the performance of the model. The model was verified in different clinical subgroups using an external dataset (GSE131769). A nomogram that included clinical indicators and risk scores was established. Gene function enrichment analysis was performed, and a protein-protein interaction network was developed.

Results. We analyzed data from 772 tumors and 88 adjacent normal tissues from the TCGA database and identified 286 glycolysis-related genes from the MSigDB. There were 185 DEGRGs. Univariate Cox regression and LASSO regression indicated that 13 of these genes were related to prognosis. A risk-scoring model based on these

Submitted 10 August 2020
Accepted 6 October 2020
Published 4 November 2020

Corresponding author
Wansheng Long, jmlws2@163.com

Academic editor
Vladimir Uversky

Additional Information and
Declarations can be found on
page 12

DOI 10.7717/peerj.10249

© Copyright
2020 Li et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

13 DEGRGs allowed classification of patients as high-risk or low-risk according to median score. The duration of overall survival (OS) was longer in the low-risk group ($P < 0.001$), and the AUC was 0.755 for 3-year OS and 0.726 for 5-year OS. The results were similar when using the GEO data set for external validation (AUC for 3-year OS: 0.731, AUC for 5-year OS: 0.728). Subgroup analysis showed there were significant differences in OS among high-risk and low-risk patients in different subgroups (T1-2, T3-4, N0, N1-3, M0, TNBC, non-TNBC; all $P < 0.01$). The C-index was 0.824, and the AUC was 0.842 for 3-year OS and 0.808 for 5-year OS from the nomogram. Functional enrichment analysis demonstrated the DEGRGs were mainly involved in regulating biological functions.

Conclusions. Our prognostic model, based on 13 DEGRGs, had excellent performance in predicting the survival of patients with IDC of the breast. These DEGRGs appear to have important biological functions in the progression of this cancer.

Subjects Bioinformatics, Oncology, Women's Health, Medical Genetics, Computational Science

Keywords Breast invasive ductal carcinoma, Glycolysis, Nomogram, Function enrichment analysis

INTRODUCTION

Breast invasive ductal carcinoma (IDC) is the most common malignant tumor in females worldwide (Harbeck *et al.*, 2019; Hunker, Sudhan & Arteaga, 2020). In 2018, there were more than 266,000 cases of breast IDC among females in the United States, and this cancer accounted for 30% of malignant tumors in females, far more than lung cancer (13%) (Bray *et al.*, 2018; Ahmad, 2019). The prognosis of women with breast IDC is related to the activation or silencing of various biological functions in tumor tissues and signaling pathways. There are prognostic models based on tumor immunity and autophagy, but no models have exclusively focused on IDC (Li *et al.*, 2020a; Li *et al.*, 2020b; Zhang, Zhang & Yu, 2019; Hu *et al.*, 2020) and few models examined genes that function in basic metabolism.

Glycolysis is a series of reactions that catabolize most carbohydrates. "Metabolic reprogramming" is the hallmark of tumors, and glycolysis is the main source of energy for tumor cells, even when there is insufficient oxygen (Wu *et al.*, 2020). Moreover, activation of glycolysis-related genes occurs in almost all tumor cells. For example, Dai *et al.* (2020) found that glycolysis promoted the progression of pancreatic cancer and induced gemcitabine chemotherapy resistance. Long noncoding RNAs (lncRNAs) that interact with Long Intergenic Noncoding RNA for IGF2BP2 Stability (LINRIS) activate aerobic glycolysis in tumor cells, and can affect the development and prognosis of rectal cancer (Wang *et al.*, 2019a; Wang *et al.*, 2019b).

Research on the function of glycolysis-related genes in breast tumors showed that hexokinase (HK2) had high expression in breast IDC cells, and that HK2 silencing inhibited IDC (Patra *et al.*, 2013; Cao *et al.*, 2020). Other research showed that overexpression of 6-Phosphofructo 2-kinase/fructose 2, 6-bisphosphatase 3 (PFKFB3) promoted the progression of breast IDC, and had negative associations with progression-free survival

(PFS), distant metastasis-free survival (DMFS), and overall survival (OS) in patients with breast IDC (O'Neal et al., 2016; Peng et al., 2018a; Peng et al., 2018b). Thus, glycolysis-related genes have a potentially significant impact on the progression of breast tumors and on the survival and prognosis of patients with breast tumors.

Our aim was to develop a prognostic model of breast IDC based on glycolysis-related genes and determine the potential functions of glycolysis-related genes in the progression of breast IDC.

MATERIALS AND METHODS

Data resources and preprocessing

The transcriptome data and corresponding clinical data of breast invasive ductal carcinoma were downloaded from the TCGA database (<https://www.cancer.gov/>) (Tomczak, Czerwinska & Wiznerowicz, 2015). The data set of glycolysis-related genes was obtained from the MSigDB database (<http://www.hmdb.ca>). Using $|\log_2FC| > 0.5$ and false discovery rate (FDR) < 0.05 as the cut-off value, the data was normalized with the “edgeR” package from R, and then the differential analysis was performed to obtain the differential expression of glycolysis-related genes (DEGRGs) between the tumor tissue and normal tissue.

Construction of risk-scoring model

Based on the above DEGRGs, univariate Cox regression and LASSO regression were used to screen out prognostic-related glycolysis genes. The risk score was evaluated by the coefficient of each prognostic-related glycolysis gene. The risk scoring formula was constructed as $\text{Risk scores} = \sum_1^i (\text{coefi} * \text{expri})$, where i is the number of genes used to build the model, coefi is the coefficients of the genes in the model, and expri the expression of genes in the model. Taking the median risk score as the cut-off point, patients were divided into high-risk and low-risk groups. The survival outcome of the two groups was observed by Kaplan–Meier survival analysis. Receiver operating characteristic (ROC) curve was applied to calculate the area under the curve (AUC) to evaluate the predictive ability of the risk-scoring model. Independent GEO (<https://www.ncbi.nlm.nih.gov/geo/>) data sets are used to verify the above results (Barrett et al., 2013). Univariate Cox regression and multivariate Cox regression were used to identified the independent prognostic factors among risk scores, age, tumor TNM stage, and whether triple negative breast cancer (TNBC) or not. Through clinical survival analysis, the predictive ability of the risk-scoring model in different clinical subgroups was clarified.

External validation of the risk scoring model

The TCGA results were validated using the GEO (<https://www.ncbi.nlm.nih.gov/geo/>) dataset (GSE131769). For this validation, the outcomes of the two groups were compared using Kaplan–Meier survival analysis. ROC curves were used to calculate AUCs and evaluate the predictive performance of the risk-scoring model.

Construction of the nomogram

A nomogram was constructed based on the results of the multivariate Cox regression, with clinical information such as age, TMN stage, TNBC status, and DEGRG risk scores, to predict 3-year and 5-year OS. The predictive ability of the nomogram was evaluated by calculating the C-index and the calibration chart, clinical decision curve analysis, and an ROC curve.

Function enrichment analysis

We analyzed the genes that were differentially expressed in the high-risk and low-risk groups using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) to identify pathway enrichment (*Kanehisa et al., 2019*). This analysis allowed identification of the functions of the differentially expressed genes. We then examined whether the differentially expressed genes were involved in the development of breast cancer.

Construction of interactive network diagram

To determine the relationship of the DEGRGs model with prognosis, a network between genes was developed. The prognosis-related genes were imported into Search Tool for the Retrieval of Interacting Genes Proteins (STRING) to construct an interaction network (*Szklarczyk et al., 2019*).

RESULTS

DEGRGs in breast IDC

We obtained gene expression data and clinical data of females with IDC of the breast (772 tumor tissues, 88 adjacent normal tissues) from the TCGA database and the glycolysis gene set (286 genes) from the GSEA website. Based on standard cut-off values for fold-change in gene expression ($|\log_2(\text{FC})| > 0.5$) and false discovery rate ($\text{FDR} < 0.05$), the IDC tissues had 185 DEGRGs, with 67 down-regulated genes and 118 up-regulated genes (*Figs. 1A, 1B, Table S1*).

Relationship of DEGRGs with prognosis and risk-scoring model

We then identified patients with follow-up times greater than 30 days, and performed univariate Cox regression and LASSO regression to screen for DEGRGs that were related to prognosis (*Figs. 2A, 2B, 2C*). This analysis indicated that 13 DEGRGs were closely related to prognosis.

We constructed a risk-scoring model based on multivariate Cox regression and divided patients into high-risk and low-risk groups based on median risk score. Kaplan–Meier survival analysis showed that patients with high-risk had significantly reduced duration of OS ($P = 9.795 \times 10^{-8}$, *Fig. 3A*). ROC analysis indicated the AUC was 0.755 for 3-year OS and 0.726 for 5-year OS (*Fig. 3B*). The risk curve and scatterplot (*Figs. 3C, 3D*) show the risk scores and survival status of all patients, and indicate that the risk coefficient and mortality rate were greater in the high-risk group. We plotted a “heat map” to visualize the expression of the 13 DEGRGs in the high-risk and low-risk groups (*Fig. 3E*). Taken together, these results confirm that 13 DEGRGs were significant prognostic indicators for patients with IDC of the breast.

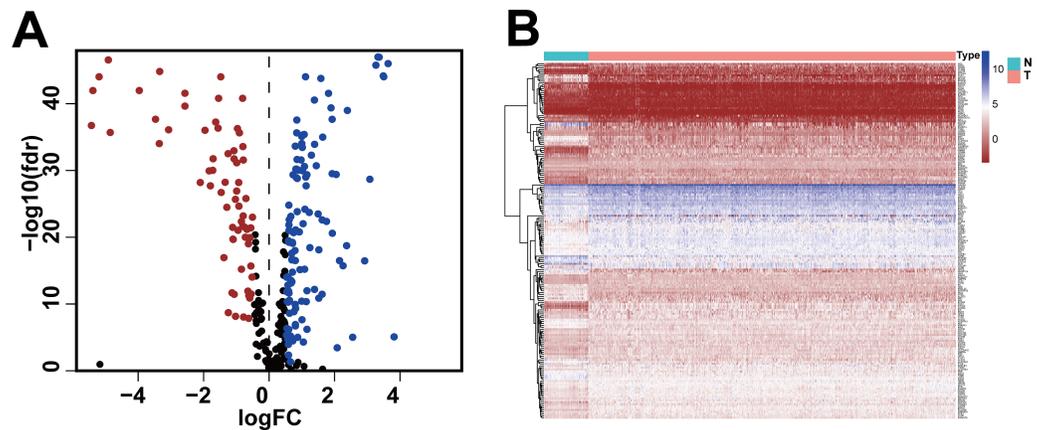


Figure 1 Differentially expressed glycolysis-related genes between breast invasive ductal carcinoma and normal tissues. (A) The volcano graph showed that compared with normal tissues, 67 DEGRGs were down-regulated and 118 DEGRGs were up-regulated in breast invasive ductal carcinoma. ($P < 0.05$) (B) The heat map showed the expression of 185 DEGRGs in both tumor tissues and normal tissues.

Full-size DOI: 10.7717/peerj.10249/fig-1

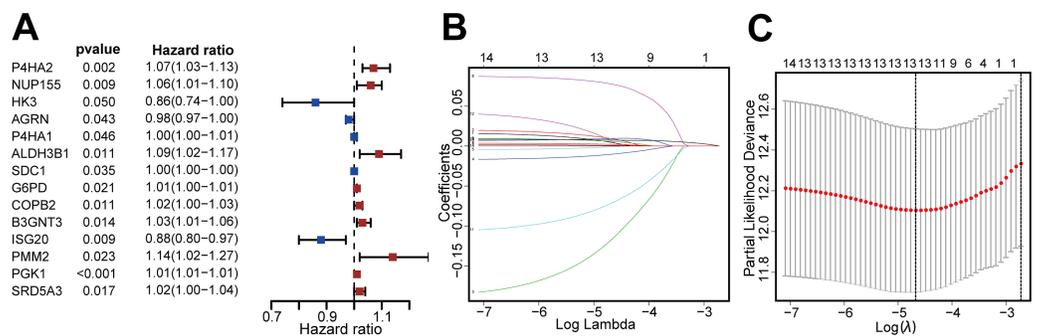


Figure 2 Screening DEGRGs with prognostic value. (A) Based on univariate Cox regression, the forest map showed that there were 14 DEGRGs with prognostic significance ($P < 0.05$). (B, C) LASSO regression further screened out 13 DEGRGs that were closely related to the prognosis.

Full-size DOI: 10.7717/peerj.10249/fig-2

We performed univariate and multivariate Cox regression to evaluate the effect of risk score, age, triple-negative breast cancer (TNBC), and TNM stage on patient prognosis (Figs. 4A, 4B). The results confirmed that the risk score was an independent prognostic factor for patients with IDC of the breast (adjusted hazard ratio: 2.71, 95% CI [1.87–3.94]).

We also performed survival analysis of different subgroups based on TNM status (Fig. 5). This analysis indicated that the risk-scoring model had good predictive value in the T1-2 subgroup, T3-4 subgroup, N0 subgroup, N1-3 subgroup, M0 subgroup, TNBC subgroup, and non-TNBC subgroup (all $P < 0.001$), but not in the M1 subgroup ($P = 0.857$).

External validation of the risk scoring model

We verified the model using the GEO dataset (GSE131769). Kaplan–Meier survival curves showed that patients with high-risk had a significantly shorter duration of OS ($P = 3.245$

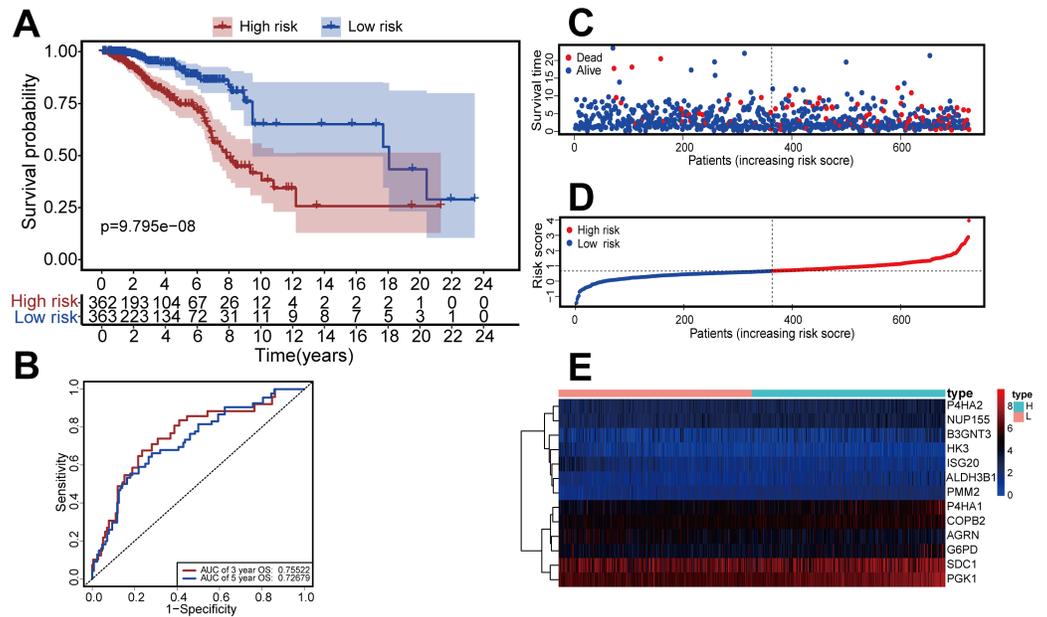


Figure 3 Construction of the risk-scoring model for patients with breast invasive ductal carcinoma based on DGRG. (A) Kaplan–Meier survival analysis showed that patients in the high-risk group had a shorter OS than that of the low-risk group ($P = 9.795 \times 10^{-8}$). (B) The ROC curve showed that the AUC of the 3-year OS and 5-year OS were 0.755 and 0.726, respectively. (C–E) The overview of survival time for each patient, the distributions of risk scores for each patient and heatmaps of expression profiles for 13-DEGRGs between the low-risk group and the high-risk group.

Full-size [DOI: 10.7717/peerj.10249/fig-3](https://doi.org/10.7717/peerj.10249/fig-3)

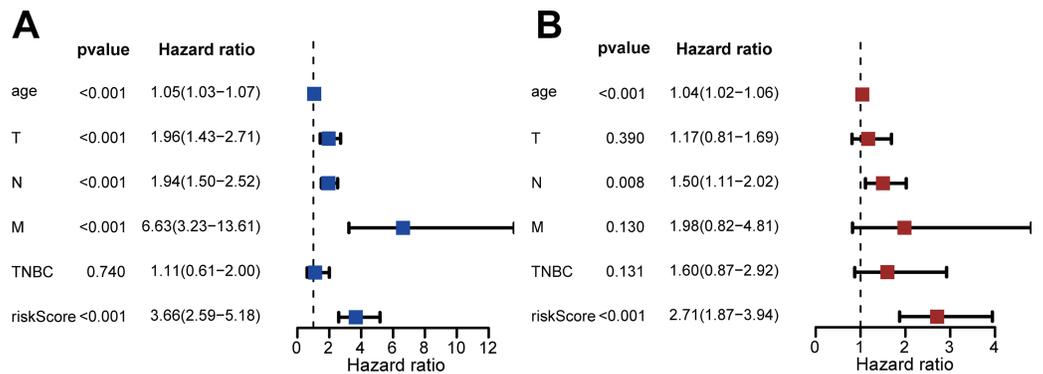


Figure 4 Assessment of risk scores and prognostic value of clinical data. (A) Univariate Cox analysis showed that risk scores and clinical variables including age, TMN stage, and whether it was TNBC or not were significantly related to OS. (B) Multivariate Cox analysis manifested that the 13-DEGRGs signature was an independent prognostic indicator for patients with breast invasive ductal carcinoma.

Full-size [DOI: 10.7717/peerj.10249/fig-4](https://doi.org/10.7717/peerj.10249/fig-4)

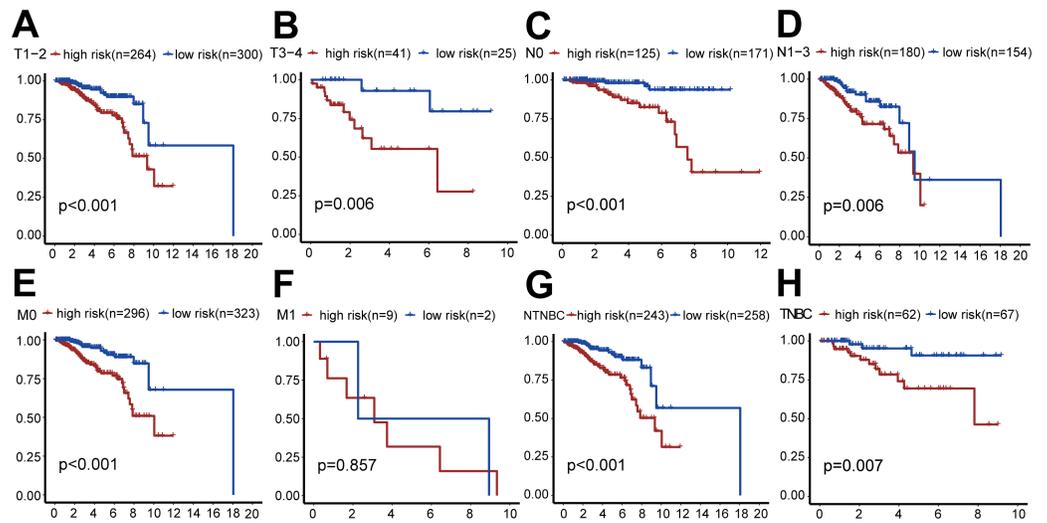


Figure 5 Clinical survival analysis of subgroups' DGRG risk scores for breast invasive ductal carcinoma. (A) In the T1–T2 subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P < 0.001$). (B) In the T3–T4 subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P = 0.006$). (C) In the N0 subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P < 0.001$). (D) In the N0–N3 subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P = 0.006$). (E) In the M0 subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P < 0.001$). (F) In the M1 subgroup, there was no significant difference in OS between patients in the high-risk group and the low-risk group, because the number of cases in the M1 subgroup is relatively small ($P = 0.857$). (G) In the TNBC subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P < 0.001$). (H) In the NTNBC subgroup, the OS of the high-risk group was lower than that of the low-risk group ($P = 0.007$). The above results suggested that the DEGRGs risk-scoring model had a good predictive ability.

Full-size DOI: [10.7717/peerj.10249/fig-5](https://doi.org/10.7717/peerj.10249/fig-5)

$\times 10^{-3}$, Fig. 6A). ROC analysis indicated that the AUC was 0.731 for 3-year OS and was 0.728 for 5-year OS (Fig. 6B). These results confirm the validity of our risk scoring model.

Construction of the prediction model

Based on the results of the multivariate Cox regression, we developed a nomogram based on age, TMN stage, risk score, and TNBC status to predict 3-year and 5-year OS (Fig. 7A). We then used the C-index, clinical decision curve, calibration chart, and ROC curve to evaluate the predictive performance of the nomogram (Figs. 7B–7F). The results indicated that the prognostic model had good prediction accuracy, with a C-index of 0.824, an AUC for 3-year OS of 0.842, and an AUC for 5-year OS of 0.808. These results verified the predictive ability of the nomogram.

Gene function enrichment analysis

To explore the potential mechanisms of prognosis-related DEGRGs in breast invasive ductal carcinoma, we performed KEGG enrichment analysis and GO enrichment analysis. KEGG pathway enrichment analysis showed that the gene set was enriched in the cell cycle, DNA replication, glycolysis and gluconeogenesis, RNA degradation, arachidonic acid metabolism, cytokine-cytokine receptor interactions, cytosolic DNA sensing, and ribosome function (Figs. 8A–8H). GO enrichment analysis showed that the gene set was enriched in

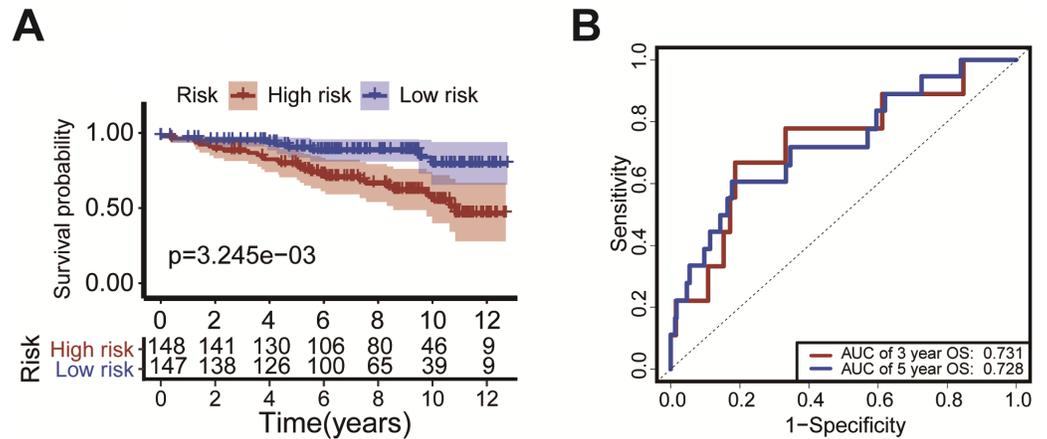


Figure 6 External validation of the risk scoring model The model was verified in GEO datasets (GSE131769). Kaplan-Meier survival curves showed that the OS of patients with high-risk was significantly shorter than in the low-risk group ($P = P = 3.245 \times 10^{-3}$, Fig. 6A). The ROC curve was drawn to calculate the AUC of the 3-year OS and 5-year OS as 0.731 and 0.728, respectively (Fig. 6B).

Full-size [DOI: 10.7717/peerj.10249/fig-6](https://doi.org/10.7717/peerj.10249/fig-6)

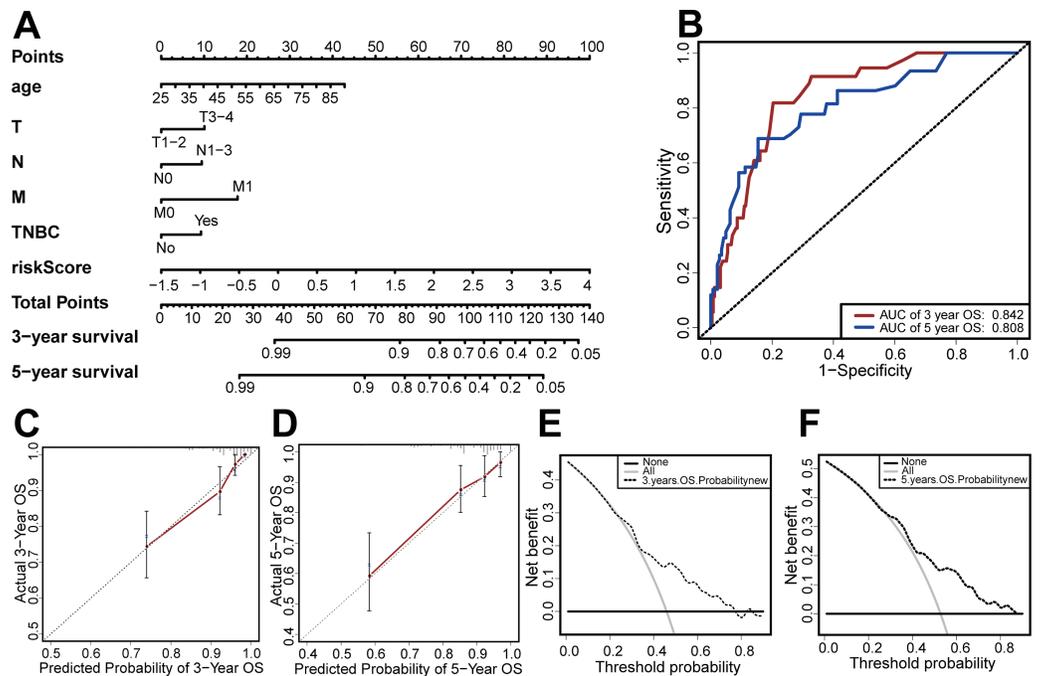


Figure 7 Establishment and evaluation of the nomogram. (A) The nomograms for predicting the patients' OS. (B) ROC curve analysis showed that AUC of 3-year and 5-year OS were 0.842 and 0.808, respectively. (C) The calibration curve for the 3-year OS of the nomogram. (D) The calibration curve for the 5-year OS of the nomogram. (E) The clinical decision curve for the 3-year OS of the nomogram. (F) The clinical decision curve for the 5-year OS of the nomogram.

Full-size [DOI: 10.7717/peerj.10249/fig-7](https://doi.org/10.7717/peerj.10249/fig-7)

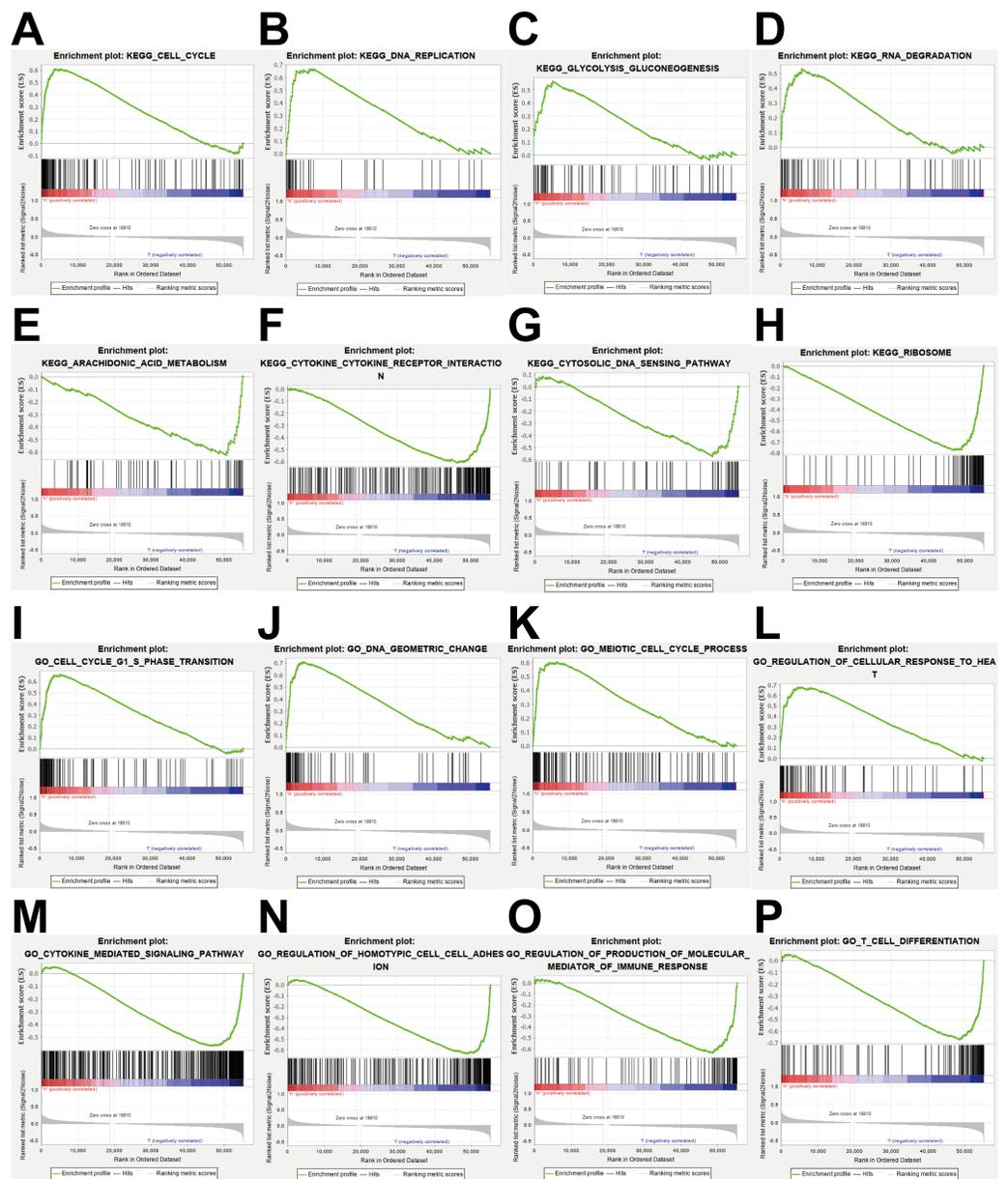


Figure 8 Enrichment analysis. (A–H) KEGG pathway analysis showed that these genes were involved in cell cycle, DNA replication, glycolysis gluconeogenesis, RNA degradation, arachidonic acid metabolism, cytokine cytokine receptor interaction, cytosolic DNA sensing pathway and ribosome. (I–P) GO enrichment analysis showed that the genes were enriched in the cell cycle G1-S phase transition, DNA geometric change, meiotic cell cycle process, regulation of cellular response to heat, cytokine mediated signaling pathway, regulation of homotypic cell cell adhesion, regulation of production of molecular mediator of immune response, T cell differentiation.

Full-size DOI: [10.7717/peerj.10249/fig-8](https://doi.org/10.7717/peerj.10249/fig-8)

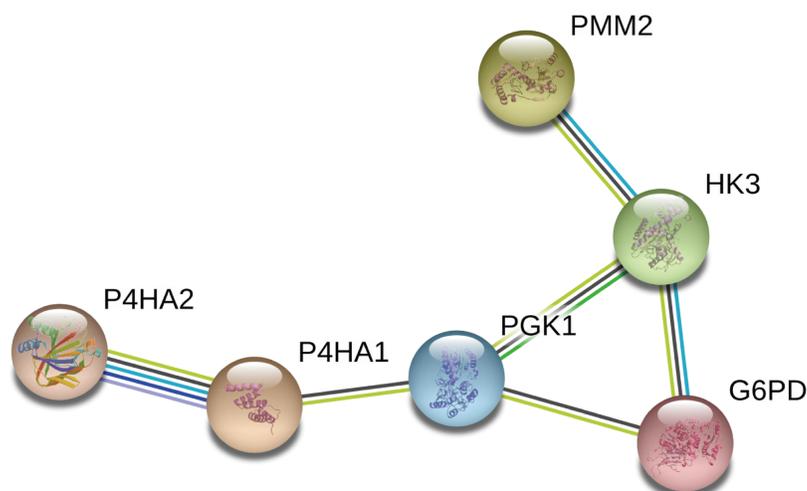


Figure 9 Construction of a network diagram of the interaction between DEGRGs. (A) The analysis of STRING showed that there were interactions among six genes (P4HA2, P4HA1, PGK1, G6PD, HK3, PMM2) in DGRG.

Full-size  DOI: 10.7717/peerj.10249/fig-9

the cell cycle G1-S phase transition, DNA geometric changes, the meiotic cell cycle process, regulation of cellular response to heat, cytokine-mediated signaling pathways, regulation of homotypic cell–cell adhesion, regulation of production of molecular mediators of immune responses, and T cell differentiation (Figs. 8I–8P).

Construction of an interactive network diagram

We constructed a network diagram to visualize the interactions between hub genes to better understand the potential functions of the different DEGRGs on prognosis. The results from STRING showed there were 6 interacting hub genes: P4HA2, P4HA1, PGK1, G6PD, HK3, and PMM2 (Fig. 9).

DISCUSSION

Breast IDC is the most common pathological type of breast tumor, and morbidity and mortality from this cancer continue to increase (Harbeck *et al.*, 2019; Badve & Gokmen-Polar, 2019). There is evidence that changes in glycolysis-related genes have multiple effects on the prognosis of these patients (Li *et al.*, 2020a; Li *et al.*, 2020b; Chen *et al.*, 2019). In particular, tumor cells reprogram the glycolysis pathway to accommodate the increased energy required for malignant transformations, including invasion and metastasis (Shen *et al.*, 2020; Abbaszadeh, Cesmeli & Biray, 2020). Given the importance of glycolysis on tumor prognosis, numerous research groups have developed models based on glycolysis-related genes in their studies of different cancers (Zhang, Zhang & Yu, 2019; Wang *et al.*, 2019a; Wang *et al.*, 2019b; Karasinska *et al.*, 2020). However, no previous study developed a prognostic model for patients with breast IDC based on glycolysis-related genes.

In this research, we identified 13 DEGRGs that were related to prognosis in patients with breast IDC, and then established a risk-scoring model. The results showed that the

OS of patients in the high-risk group was significantly shorter than that of patients in the low-risk group. We also examined the impact of patient age, TMN stage, TNBC status, and risk score to construct a nomogram that predicts the 3-year and 5-year OS of these patients. Our application of various evaluation methods indicated that the nomogram had good performance in the prediction of OS. Our ROC analysis showed that the AUC was 0.842 for 3-year OS and 0.808 for 5-year OS, higher than the AUC values reported in previous models ([Lin et al., 2020](#); [Xie et al., 2019](#)), thus indicating that our model had better predictive ability.

Among the 13 DEGRGs we used to construct the risk model, increased levels of P4H2A, NUP155, ALDH3B1, SDC1, G6PD, COPB2, B3GNT3, PMM2, and PGK1 were associated with poor prognosis and increased levels of HK3, AGRN, P4HA1, and ISG20 were associated with favorable prognosis ([Table S2](#)). Previous research reported increased levels of P4HA2 and P4HA1 (the two isomers of collagen prolyl 4-hydroxylase) in several types of human cancers and that both enzymes promoted glycolysis in tumor cells ([Li et al., 2019](#)). PGK1 is the first key enzyme to produce ATP in the glycolytic pathway, PGK1 is not only a metabolic enzyme but also a protein kinase, which mediates the tumor growth, migration and invasion through phosphorylation some important substrates ([Fu & Yu, 2020](#)). There is also evidence that SDC1 can promote the migration of breast cancer cells across the blood–brain barrier by regulating the expression of cytokines, thus promoting brain metastasis ([Sayyad et al., 2019](#)). Mele et al. found that the overexpression of G6PD can induce lapatinib resistance in breast cancer and also found a significant correlation between high expression of G6PD and tumor recurrence ([Mele et al., 2019](#)). Sauter et al. found that the level of HK3 in the nipple aspirate of patients with breast cancer was significantly lower than that of healthy women, and considered an elevated HK3 level as a possible sign of early breast cancer ([Mannello & Gazzanelli, 2001](#)). Thus, these previous studies are consistent with our finding that glycolysis-related genes were closely related to the occurrence and development of breast cancer and the prognosis of patients.

To further characterize the potential roles of the 13 DEGRGs in our risk model, we performed GO and KEGG enrichment analysis. The results showed that the gene set was enriched in cell cycle, DNA replication, glycolysis gluconeogenesis, RNA degradation, arachidonic acid metabolism, cytokine cytokine receptor interaction, cytosolic DNA sensing pathway and ribosome. Previous studies showed that under hypoxic conditions, metabolic reprogramming of breast tumor stem cells helped to maintain their growth and proliferation ([Peng et al., 2018a](#); [Peng et al., 2018b](#)). Breast cancer occurs in patients with impaired immune function, in which cytokines function as growth signals for tumor cells. Many studies showed that interactions between the immune system and cancer cells, which are mediated by cytokines and chemokines, affect the initiation and progression of breast cancer and the response to treatment ([Lim et al., 2018](#); [King, Mir & Singh, 2017](#); [Fabre et al., 2018](#)). DNA replication is a fundamental biological process, and replication disorders can lead to genomic instability, an important feature of breast cancer. Many experimental and clinical studies have identified disorders of DNA replication during the development and progression of breast cancer ([Kitao et al., 2018](#)). Aerobic glycolysis pathway includes hexokinase, phosphofructokinase (PFK), and other genes ([Wu et al., 2020](#)). These previous

findings therefore support our conclusion that the 13 glycolysis-related genes identified here play an important role in the progression of breast tumors.

Our study has some limitations. Firstly, the predictive model lacks information about patient treatments, thus limiting its predictive performance. Secondly, this study was based on bioinformatics analysis, and further studies are needed to determine the potential functional mechanisms.

CONCLUSIONS

In conclusion, our prediction model, which is based on 13 DEGRGs and the clinical characteristics of patients, can reliably predict the OS of patients with breast IDC. These 13 DEGRGs and several related miRNAs thus appear to play an important role in the development and progression of breast IDC.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Guangdong Medical Research Foundation (Grant No. A2020622) and the Elite Young Scholars Program of Jiangmen Central Hospital (Grant No. J201905). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

Guangdong Medical Research Foundation: A2020622.

Elite Young Scholars Program of Jiangmen Central Hospital: J201905.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Xiaoping Li and Qihe Yu conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Jishang Chen and Hui Huang performed the experiments, prepared figures and/or tables, and approved the final draft.
- Zhuangsheng Liu performed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Chengxing Wang, Yaoming He, Xin Zhang, Weiwen Li and Chao Li analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.
- Jinglin Zhao performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Wansheng Long conceived and designed the experiments, performed the experiments, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The RNA-seq transcriptome data and corresponding clinicopathological information of the data are available at the TCGA project: TCGA-BRCA.

The raw measurements are available in the [Supplemental Tables](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.10249#supplemental-information>.

REFERENCES

- Abbaszadeh Z, Cesmeli S, Biray AC. 2020. Crucial players in glycolysis: cancer progress. *Gene* 726:144158 DOI 10.1016/j.gene.2019.144158.
- Ahmad A. 2019. Breast cancer statistics: recent trends. *Advances in Experimental Medicine and Biology* 1152:1–7 DOI 10.1007/978-3-030-20301-6_1.
- Badve SS, Gokmen-Polar Y. 2019. Ductal carcinoma in situ of breast: update 2019. *Pathology* 51(6):563–569 DOI 10.1016/j.pathol.2019.07.005.
- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Research* 41(Database issue):D991–D995 DOI 10.1093/nar/gks1193.
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* 68(6):394–424 DOI 10.3322/caac.21492.
- Cao L, Wang M, Dong Y, Xu B, Chen J, Ding Y, Qiu S, Li L, Karamfilova ZE, Zhou X, Xu Y. 2020. Circular RNA circRNF20 promotes breast cancer tumorigenesis and Warburg effect through miR-487a/HIF-1alpha/HK2. *Cell Death and Disease* 11(2):145 DOI 10.1038/s41419-020-2336-0.
- Chen F, Chen J, Yang L, Liu J, Zhang X, Zhang Y, Tu Q, Yin D, Lin D, Wong PP, Huang D, Xing Y, Zhao J, Li M, Liu Q, Su F, Su S, Song E. 2019. Extracellular vesicle-packaged HIF-1alpha-stabilizing lncRNA from tumour-associated macrophages regulates aerobic glycolysis of breast cancer cells. *Nature Cell Biology* 21(4):498–510 DOI 10.1038/s41556-019-0299-0.
- Dai S, Peng Y, Zhu Y, Xu D, Zhu F, Xu W, Chen Q, Zhu X, Liu T, Hou C, Wu J, Miao Y. 2020. Glycolysis promotes the progression of pancreatic cancer and reduces cancer cell sensitivity to gemcitabine. *Biomedicine & Pharmacotherapy* 121:109521 DOI 10.1016/j.biopha.2019.109521.
- Fabre J, Giustinniani J, Garbar C, Merrouche Y, Antonicelli F, Bensussan A. 2018. The interleukin-17 family of cytokines in breast cancer. *International Journal of Molecular Sciences* 19(12):3880 DOI 10.3390/ijms19123880.

- Fu Q, Yu Z. 2020.** Phosphoglycerate kinase 1 (PGK1) in cancer: a promising target for diagnosis and therapy. *Life Sciences* **256**:117863 DOI [10.1016/j.lfs.2020.117863](https://doi.org/10.1016/j.lfs.2020.117863).
- Hanker AB, Sudhan DR, Arteaga CL. 2020.** Overcoming endocrine resistance in breast cancer. *Cancer Cell* **37**(4):496–513 DOI [10.1016/j.ccell.2020.03.009](https://doi.org/10.1016/j.ccell.2020.03.009).
- Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, Ruddy K, Tsang J, Cardoso F. 2019.** Breast cancer. *Nature Reviews Disease Primers* **5**(1):66 DOI [10.1038/s41572-019-0111-2](https://doi.org/10.1038/s41572-019-0111-2).
- Hu D, Jiang L, Luo S, Zhao X, Hu H, Zhao G, Tang W. 2020.** Development of an autophagy-related gene expression signature for prognosis prediction in prostate cancer patients. *Journal of Translational Medicine* **18**(1):160 DOI [10.1186/s12967-020-02323-x](https://doi.org/10.1186/s12967-020-02323-x).
- Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. 2019.** New approach for understanding genome variations in KEGG. *Nucleic Acids Research* **47**(D1):D590–D595 DOI [10.1093/nar/gky962](https://doi.org/10.1093/nar/gky962).
- Karasinska JM, Topham JT, Kalloger SE, Jang GH, Denroche RE, Culibrk L, Williamson LM, Wong HL, Lee M, O’Kane GM, Moore RA, Mungall AJ, Moore MJ, Warren C, Metcalfe A, Notta F, Knox JJ, Gallinger S, Laskin J, Marra MA, Jones S, Renouf DJ, Schaeffer DF. 2020.** Altered gene expression along the glycolysis-cholesterol synthesis axis is associated with outcome in pancreatic cancer. *Clinical Cancer Research* **26**(1):135–146 DOI [10.1158/1078-0432.CCR-19-1543](https://doi.org/10.1158/1078-0432.CCR-19-1543).
- King J, Mir H, Singh S. 2017.** Association of cytokines and chemokines in pathogenesis of breast cancer. *Progress in Molecular Biology and Translational Science* **151**:113–136 DOI [10.1016/bs.pmbts.2017.07.003](https://doi.org/10.1016/bs.pmbts.2017.07.003).
- Kitao H, Iimori M, Kataoka Y, Wakasa T, Tokunaga E, Saeki H, Oki E, Maehara Y. 2018.** DNA replication stress and cancer chemotherapy. *Cancer Science* **109**(2):264–271 DOI [10.1111/cas.13455](https://doi.org/10.1111/cas.13455).
- Li B, Geng R, Wu Q, Yang Q, Sun S, Zhu S, Xu Z, Sun S. 2020a.** Alterations in immune-related genes as potential marker of prognosis in breast cancer. *Frontiers in Oncology* **10**:333 DOI [10.3389/fonc.2020.00333](https://doi.org/10.3389/fonc.2020.00333).
- Li Q, Wang Q, Zhang Q, Zhang J, Zhang J. 2019.** Collagen prolyl 4-hydroxylase 2 predicts worse prognosis and promotes glycolysis in cervical cancer. *American Journal of Translational Research* **11**(11):6938–6951.
- Li W, Xu M, Li Y, Huang Z, Zhou J, Zhao Q, Le K, Dong F, Wan C, Yi P. 2020b.** Comprehensive analysis of the association between tumor glycolysis and immune/inflammation function in breast cancer. *Journal of Translational Medicine* **18**(1):92 DOI [10.1186/s12967-020-02267-2](https://doi.org/10.1186/s12967-020-02267-2).
- Lim B, Woodward WA, Wang X, Reuben JM, Ueno NT. 2018.** Inflammatory breast cancer biology: the tumour microenvironment is key. *Nature Reviews Cancer* **18**(8):485–499 DOI [10.1038/s41568-018-0010-y](https://doi.org/10.1038/s41568-018-0010-y).
- Lin QG, Liu W, Mo YZ, Han J, Guo ZX, Zheng W, Wang JW, Zou XB, Li AH, Han F. 2020.** Development of prognostic index based on autophagy-related genes analysis in breast cancer. *Aging* **12**(2):1366–1376 DOI [10.18632/aging.102687](https://doi.org/10.18632/aging.102687).

- Mannello F, Gazzanelli G. 2001.** Prostate-specific antigen (PSA/hK3): a further player in the field of breast cancer diagnostics? *Breast Cancer Research* 3(4):238–243 DOI 10.1186/bcr302.
- Mele L, La Noce M, Paino F, Regad T, Wagner S, Liccardo D, Papaccio G, Lombardi A, Caraglia M, Tirino V, Desiderio V, Papaccio F. 2019.** Glucose-6-phosphate dehydrogenase blockade potentiates tyrosine kinase inhibitor effect on breast cancer cells through autophagy perturbation. *Journal of Experimental & Clinical Cancer Research* 38(1):160 DOI 10.1186/s13046-019-1164-5.
- O’Neal J, Clem A, Reynolds L, Dougherty S, Imbert-Fernandez Y, Telang S, Chesney J, Clem BF. 2016.** Inhibition of 6-phosphofructo-2-kinase (PFKFB3) suppresses glucose metabolism and the growth of HER2+ breast cancer. *Breast Cancer Research and Treatment* 160(1):29–40 DOI 10.1007/s10549-016-3968-8.
- Patra KC, Wang Q, Bhaskar PT, Miller L, Wang Z, Wheaton W, Chandel N, Laakso M, Muller WJ, Allen EL, Jha AK, Smolen GA, Clasquin MF, Robey B, Hay N. 2013.** Hexokinase 2 is required for tumor initiation and maintenance and its systemic deletion is therapeutic in mouse models of cancer. *Cancer Cell* 24(2):213–228 DOI 10.1016/j.ccr.2013.06.014.
- Peng F, Li Q, Sun JY, Luo Y, Chen M, Bao Y. 2018a.** PFKFB3 is involved in breast cancer proliferation, migration, invasion and angiogenesis. *International Journal of Oncology* 52(3):945–954 DOI 10.3892/ijo.2018.4257.
- Peng F, Wang JH, Fan WJ, Meng YT, Li MM, Li TT, Cui B, Wang HF, Zhao Y, An F, Guo T, Liu XF, Zhang L, Lv L, Lv DK, Xu LZ, Xie JJ, Lin WX, Lam EW, Xu J, Liu Q. 2018b.** Glycolysis gatekeeper PDK1 reprograms breast cancer stem cells under hypoxia. *Oncogene* 37(8):1062–1074 DOI 10.1038/onc.2017.368.
- Sayyad MR, Puchalapalli M, Vergara NG, Wangensteen SM, Moore M, Mu L, Edwards C, Anderson A, Kall S, Sullivan M, Dozmorov M, Singh J, Idowu MO, Koblinski JE. 2019.** Syndecan-1 facilitates breast cancer metastasis to the brain. *Breast Cancer Res Treat* 178(1):35–49 DOI 10.1007/s10549-019-05347-0.
- Shen C, Xuan B, Yan T, Ma Y, Xu P, Tian X, Zhang X, Cao Y, Ma D, Zhu X, Zhang Y, Fang JY, Chen H, Hong J. 2020.** m(6)A-dependent glycolysis enhances colorectal cancer progression. *Mol Cancer* 19(1):72 DOI 10.1186/s12943-020-01190-w.
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, Mering CV. 2019.** STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research* 47(D1):D607–D613 DOI 10.1093/nar/gky1131.
- Tomczak K, Czerwinska P, Wiznerowicz M. 2015.** The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp Oncol* 19(1A):A68–A77 DOI 10.5114/wo.2014.47136.
- Wang Y, Lu JH, Wu QN, Jin Y, Wang DS, Chen YX, Liu J, Luo XJ, Meng Q, Pu HY, Wang YN, Hu PS, Liu ZX, Zeng ZL, Zhao Q, Deng R, Zhu XF, Ju HQ, Xu RH. 2019a.** LncRNA LINRIS stabilizes IGF2BP2 and promotes the aerobic glycolysis in colorectal cancer. *Molecular Cancer* 18(1):174 DOI 10.1186/s12943-019-1105-0.

- Wang ZH, Zhang YZ, Wang YS, Ma XX. 2019b.** Identification of novel cell glycolysis related gene signature predicting survival in patients with endometrial cancer. *Cancer Cell International* **19**:Article 296 DOI [10.1186/s12935-019-1001-0](https://doi.org/10.1186/s12935-019-1001-0).
- Wu Z, Wu J, Zhao Q, Fu S, Jin J. 2020.** Emerging roles of aerobic glycolysis in breast cancer. *Clinical and Translational Oncology* **22(5)**:631–646 DOI [10.1007/s12094-019-02187-8](https://doi.org/10.1007/s12094-019-02187-8).
- Xie P, Ma Y, Yu S, An R, He J, Zhang H. 2019.** Development of an immune-related prognostic signature in breast cancer. *Frontiers in Genetics* **10**:1390 DOI [10.3389/fgene.2019.01390](https://doi.org/10.3389/fgene.2019.01390).
- Zhang L, Zhang Z, Yu Z. 2019.** Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *Journal of Translational Medicine* **17(1)**:423 DOI [10.1186/s12967-019-02173-2](https://doi.org/10.1186/s12967-019-02173-2).