Heliyon 10 (2024) e24777

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Lactylation-related gene signature for prognostic prediction and immune infiltration analysis in breast cancer

Yangchi Jiao [a,1], Fuqing Ji [b,1], Lan Hou [a], Yonggang Lv [b], Juliang Zhang [a,*]

[a] Department of Thyroid, Breast and Vascular Surgery, Xijing Hospital, Air Force Military Medical University, Xi'an, Shaanxi, China
[b] Department of Thyroid Breast Surgery, Xi'an NO.3 Hospital, The Affiliated Hospital of Northwest University, Xi'an, Shaanxi, China

ARTICLE INFO

Keywords:
Breast cancer
Lactylation
Gene signature
Prognostic model
Immune infiltration
Drug response

ABSTRACT

Background: Lactylation is implicated in various aspects of tumor biology, but its relation to breast cancer remains poorly understood. This study aimed to explore the roles of the lactylation-related genes in breast cancer and its association with the tumor microenvironment.
Methods: The expression and mutation patterns of lactylation-related genes were analyzed using the breast cancer data from The Cancer Genome Atlas (TCGA) database and GSE20685 datasets. Unsupervised clustering was used to identify two lactylation clusters. A lactylation-related gene signature was developed and validated using the training and validation cohorts. Immune cell infiltration and drug response were assessed.
Results: We analyzed the mRNA expression, copy number variations, somatic mutations, and correlation networks of 22 lactylation-related genes in breast cancer tissues. We identified two distinct lactylation clusters with different survival outcomes and immune microenvironments. We further classified the patients into two gene subtypes based on lactylation clusters and identified a 7-gene signature for breast cancer survival prognosis. The prognostic score based on this signature demonstrated prognostic value and predicted the therapeutic response.
Conclusion: Lactylation-related genes play a critical role in breast cancer by influencing tumor growth, immune microenvironment, and drug response. This lactylation-related gene signature may serve as a prognostic marker and a potential therapeutic target for breast cancer.

1. Introduction

In 2020, breast cancer became the most prevalent type of cancer worldwide, having surpassed lung cancer in incidence [1]. Despite the heterogeneity of breast cancer diagnostic and therapeutic strategies, effective approaches to improve recurrence-free survival (RFS) and overall survival (OS) rates are lacking [2]. Additionally, the resistance of the disease to chemotherapy, radiotherapy, and endocrine therapy presents a pressing challenge for maintaining survival [3]. Hence, the development of an efficient prognostic model that aids in selecting appropriate treatment approaches and medications for the management of breast cancer is crucial.

Protein lactylation is a newly discovered post-translational modification of proteins involving the attachment of lactate molecules to lysine residues that was first reported in 2019 [4]. Protein lactylation is present in numerous human tissues, and it has garnered significant attention from the scientific community [5–7]. Furthermore, histone lactylation plays a crucial role in various biological

* Corresponding author.
E-mail address: vascularzhang@163.com (J. Zhang).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.heliyon.2024.e24777
Received 28 June 2023; Received in revised form 7 January 2024; Accepted 15 January 2024
Available online 19 January 2024

processes such as glycolysis-related cellular functions [8], macrophage polarization [9], neurodevelopment [10], and regulation of tumor proliferation [11]. This highlights the importance of understanding protein lactylation and its functions in breast cancer. Moreover, numerous studies [12–14] have demonstrated the remarkable prognostic potential of lactylation-related genes in liver cancer and gastric cancer. However, there is currently a lack of scientific inquiry into the potential association between lactylation and breast cancer oncogenesis, the immune microenvironment, or immunotherapy, as well as models that can be employed to assess the prognostic significance of lactylation-related genes in breast cancer. Therefore, investigating the physiological and pathological effects of lactylation is essential. Exploration of lactylation and its potential functions, as well as constructing a predictive model, can lead to novel diagnostic and therapeutic strategies for breast cancer.

This study aimed to comprehensively analyze the role of lactylation-related genes in the breast cancer tumor microenvironment (TME) and immunotherapy. To achieve this goal, we performed a detailed assessment of the expression profiles of 22 lactylation-related genes in patients with breast cancer. Based on these expression levels, the patients were divided into two lactylation clusters, and the differentially expressed genes (DEGs) in these clusters were used to classify the patients into two gene subtypes. In addition, DEGs with prognostic significance were used to construct a lactylation-related gene signature consisting of seven genes (*RAD51*, *NEK10*, *PCP2*, *IDO1*, *CASP14*, *CLSTN2*, and *IGHG1*) and a survival prognostic score was developed. The stability and reliability of the prognostic scores were confirmed in the test cohort. Importantly, this prognostic score proved useful for stratifying patients into high- and low risk groups to predict OS and the immune landscape in breast cancer [15]. Using the prognostic score, we accurately predicted the survival prognosis of patients and their response to treatment.

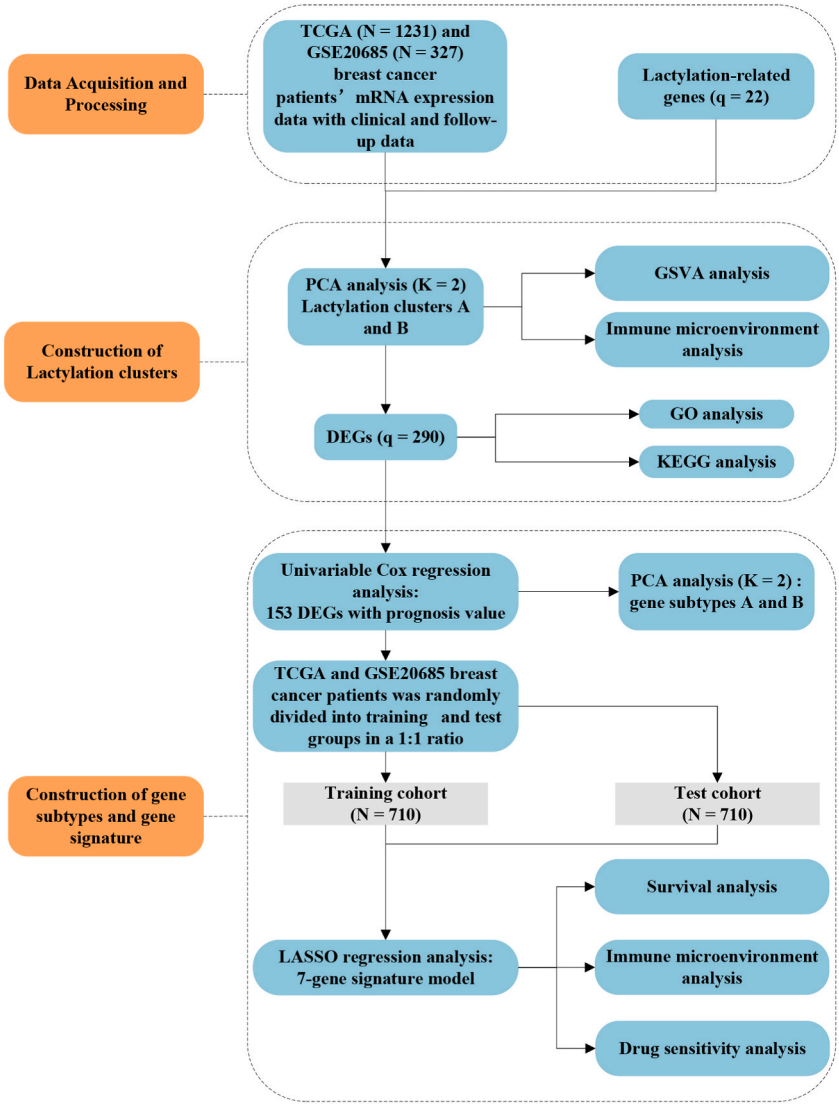

Fig. 1. Flowchart of the study design. N represents the number of patients and q represents the number of genes.

2. Materials and methods

2.1. Data acquisition and preprocessing

The overall design and flowchart of the study are shown in Fig. 1. For this study, in order to obtain training samples and validate the results across different data sets, gene expression profiles and clinical data were compiled from 327 breast cancer tissues available on the GEO data portal (https://www.ncbi.nlm.nih.gov/geo/) (GSE20685, which has a large number of samples in the data set and relatively comprehensive clinical follow-up data of patients) and 1118 breast cancer tissues and 113 normal tissues from the TCGA database (https://portal.gdc.cancer.gov/). In the breast cancer data from TCGA database, survival data were available for 1097 patients, among whom gene expression data were missing for 3 patients and 1 patient had a negative survival time (−7 years), which was subsequently excluded from the analysis. A total of 1093 patients had complete survival and gene expression data, among whom 232 died with a median OS of 3.12 years (1.45–6.80). Cheng et al. reported that the 16 lactylation-related genes had an ability to predict the prognosis of liver cancer [12]. Similarly, Yang et al. found that the 6 lactylation-related genes were significantly associated with the prognosis of gastric cancer, suggesting that lactylation-related genes were associated with tumor prognosis [13]. Therefore, we chose those 22 lactylation-related genes (*ARID3A*, *CCNA2*, *DDX39A*, *EHMT2*, *FABP5*, *G6PD*, *H2AX*, *HMGA1*, *KIF2C*, *MKI67*, *PFKP*, *PKM2*, *RACGAP1*, *RFC4*, *STMN1*, *TKT*, *EFNA3*, *VCAN*, *PLOD2*, *HBB*, *NUP50*, and *STC1*) mentioned in these studies for further analysis.

2.2. Construction of lactylation clusters and principal component analysis (PCA)

To identify distinct lactylation patterns in breast cancer, we employed the "ConsensusClusterPlus" R package [16] to perform consensus classification. The number of clusters was determined by examining the tendency and smoothness of the cumulative distribution function (CDF) curve. Additionally, PCA was conducted using the "prcomp" function in the "stats" R package developed by the R Core Team.

2.3. Analysis of clinical characteristics and prognosis of patients in different lactylation clusters

We used the R packages "survival" and "survminer" to perform Kaplan–Meier plot analysis and evaluate the prognostic values of patients with breast cancer in different lactylation clusters. The study also compared clinical characteristics such as age, stage T, and stage N among these clusters.

2.4. Gene set variation analysis of different lactylation clusters

We used the R package "GSVA" to perform a gene set variation analysis (GSVA) to identify the biological processes underlying the unique lactylation patterns observed in the study.

2.5. Estimation of TME in different lactylation clusters

A single-sample gene set enrichment analysis (ssGSEA) algorithm was used to estimate the proportions of 23 human immune cell subsets present in the TME of different lactylation clusters. We also compared the expression levels of 33 critical immune checkpoints across clusters.

2.6. Identification, functional analysis, and gene subtype construction of DEGs in different lactylation clusters

We used the "limma" package in R to identify 290 DEGs with a fold-change of 2 and an adjusted p-value <0.05 between different lactylation clusters. The DEGs were subjected to functional analysis using the "clusterprofiler" R package and a gene set file obtained from the MSigDB database (https://www.gsea-msigdb.org/gsea/msigdb). Univariate Cox regression analysis was performed on the DEGs, and 153 genes with significant prognostic value ($p < 0.05$) were identified for further investigation. Based on these 153 genes, a consensus classification was performed using the "ConsensusClusterPlus" R package, which divided the patients into two gene signature subtypes.

2.7. Development of the DEG model and prognostic score

A dataset of 1420 breast cancer samples from the TCGA database and GSE20685 were used, which were divided 1:1 into a training cohort and a test cohort using the R package "caret" [17] with 710 patients in each group. Univariate Cox regression analysis was performed on the training cohort to identify lactylation-related DEGs associated with OS (the length of time from the start of a clinical study or treatment until death from any cause) [18], followed by the LASSO algorithm using the R package "glmnet" to select the most promising prognostic DEGs. Next, multivariate Cox regression analysis was performed to select independent prognostic DEGs for breast cancer, from which the prognostic signature was constructed. Finally, the prognostic score which was used in cancer research [15] was calculated using a formula based on the selected genes as follows:

$$Prognostic\ Score = \sum_{i=1}^{q} x_i \times \beta_i$$

where x_i represents the expression data of the prognostic DEGs, q represents the number of the selected genes, β_i represents the Cox regression coefficients after normalization and the prognostic score is calculated as the sum of the product of each variable (x) and its corresponding coefficient (β).

2.8. Independent prognosis analysis of the risk model

To assess the independence of the risk model in the training and test cohorts, we performed univariate and multivariate Cox regression analyses using the 'survival' R package. The Cox regression analysis was employed to evaluate the impact of specific factors on survival time, while the Kaplan-Meier method was utilized to estimate the survival function. In both univariate and multivariate analyses, multiple predictive variables were considered for their impact on survival time. Furthermore, the log-rank test is used to compare the survival curves between two prognostic groups, and the resulting p-value assesses the significance of this difference. A nomogram model based on both clinicopathologic characteristics and prognostic score was constructed using the "rms" R package. The c-index was reported to assess the concordance between the prognostic score and survival.

2.9. Relationship between prognostic score and immune cell infiltration

The proportions of immune cells in the low- and high-risk groups were estimated using the CIBERSORT algorithm. In addition, we performed Spearman's rank correlation analysis to explore the potential associations between Prognostic score values and immune-infiltrating cells.

2.10. Analysis of the immune microenvironment landscape and drug sensitivity

We used the ESTIMATE algorithm to evaluate the abundance of stromal and immune cells. The stromal, immune, and ESTIMATE

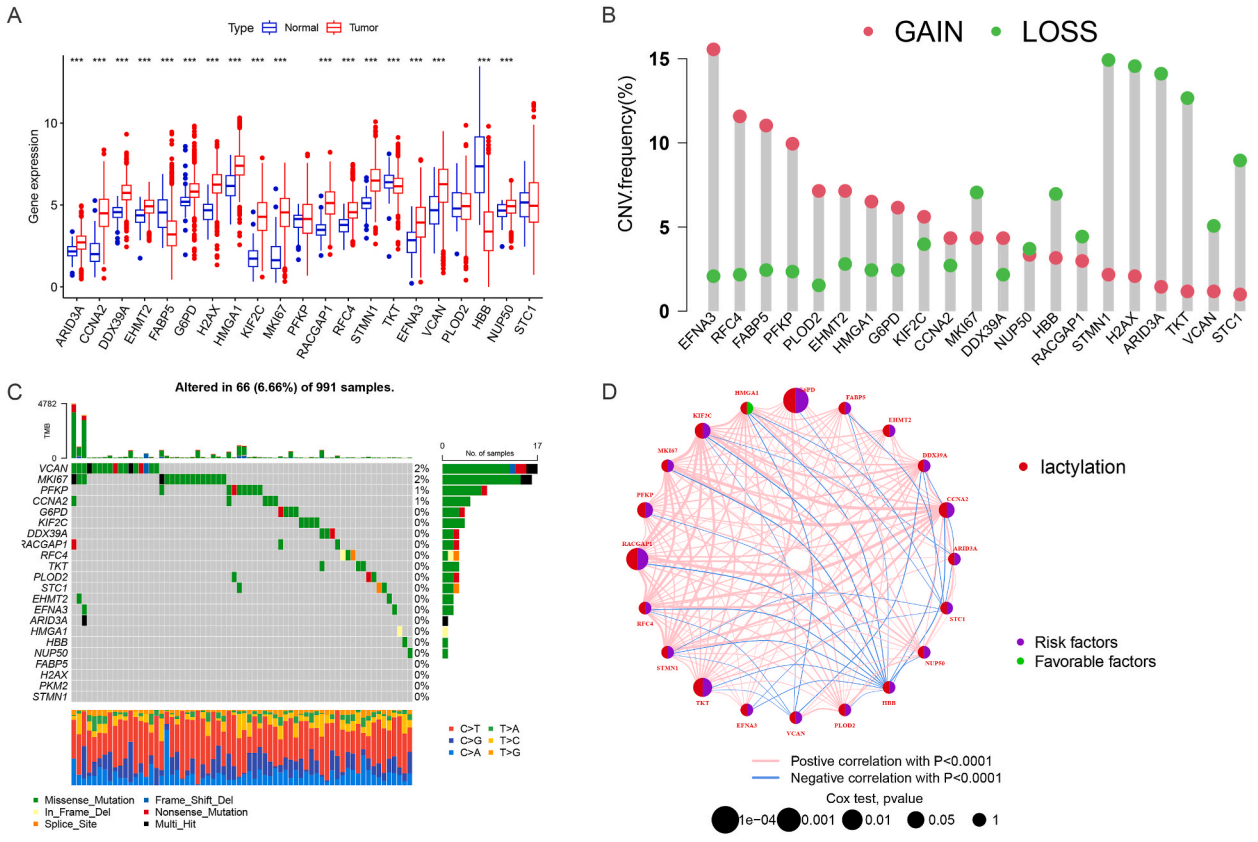

Fig. 2. Landscape of lactylation-related genes in breast cancer. (A) Differential expression of lactylation-related genes in breast cancer compared to normal tissue. (B) Frequency of CNVs in lactylation-related genes. (C) Analysis of genetic changes of lactylation-related genes. (D) Correlation network of the 20 lactylation-related genes. CNV, copy number variation.

scores for breast cancer were calculated using the R package "estimate." To obtain antitumor drug profiles, we accessed the Genomics of Drug Sensitivity in the Cancer (GDSC) portal [19] (https://www.cancerrxgene.org/). Spearman's correlation analysis was performed to determine the correlation between drug sensitivity and prognostic scores. In addition, we used the R package "pRRophetic" to calculate the half-maximum inhibitory concentration (IC50) for drug sensitivity comparisons.

2.11. Statistical analysis

The analyses were performed using R software. Categorical variables were subjected to Pearson's chi-squared test, whereas continuous variables were analyzed using one-way ANOVA, with a p-value of <0.05 indicating statistical significance. OS of patients with breast cancer was analyzed using Kaplan–Meier analysis. Univariate and multivariate Cox regression analyses were performed to

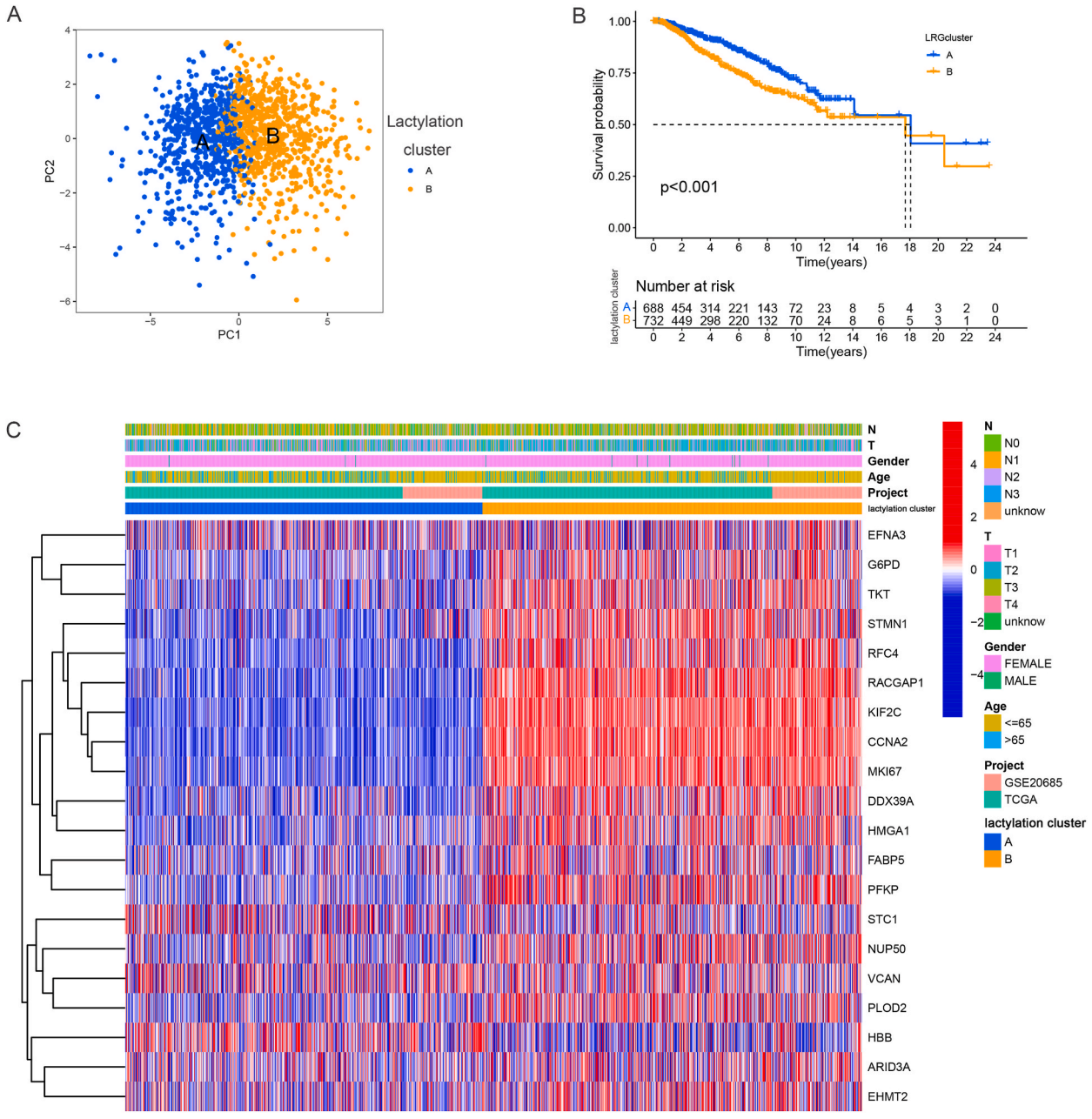

Fig. 3. Identification of lactylation clusters in breast cancer. (A) PCA of the two clusters. (B) Comparison of OS between the two clusters. (C) Heatmap for the correlation between clinicopathologic features and the two lactylation clusters. CDF, cumulative distribution function; PCA, principal component analysis; OS, overall survival.

identify the independent prognostic factors and clinical characteristics that differed significantly between the patient groups. Differences in immune cell infiltration were assessed using the Mann–Whitney *U* test.

3. Results

3.1. Identification of lactylation-related genes

We used the "limma" package to examine the mRNA expression of lactylation-related genes based on TCGA data, which included 113 normal tissues and 1118 breast cancer tissues. Our analysis revealed that *ARID3A*, *CCNA2*, *DDX39A*, *EHMT2*, *G6PD*, *H2AX*, *HMGA1*, *KIF2C*, *MKI67*, *RACGAP1*, *RFC4*, *STMN1*, *EFNA3*, *VCAN*, and *NUP50* were significantly upregulated in breast cancer tissues compared to normal tissues, as shown in Fig. 2A. The expression levels of *FABP5*, *TKT*, and *HBB* were significantly decreased in breast cancer tissues. These results provide novel insights into the differential expression of lactylation-related genes in breast cancer tissues and suggest their potential as diagnostic or therapeutic targets for this disease.

We also investigated the copy number variations (CNVs) of 22 lactylation-related genes in breast cancer to explore how these lactylation-related genes are altered on chromosomes and where they are located. Our analysis revealed that *EFNA3* was the most significantly altered "gain" gene, located on chromosome 1, whereas *STMN1* was mainly "loss" and located on chromosome 1 too (Fig. 2B–S1). We examined the frequency of somatic mutations in these 22 lactylation-related genes and found that 66 (6.66 %) of 991 breast cancer samples had mutations. Specifically, *VCAN* had the highest mutation frequency, followed by *MKI67*, *PFKP*, and *CCNA2*,

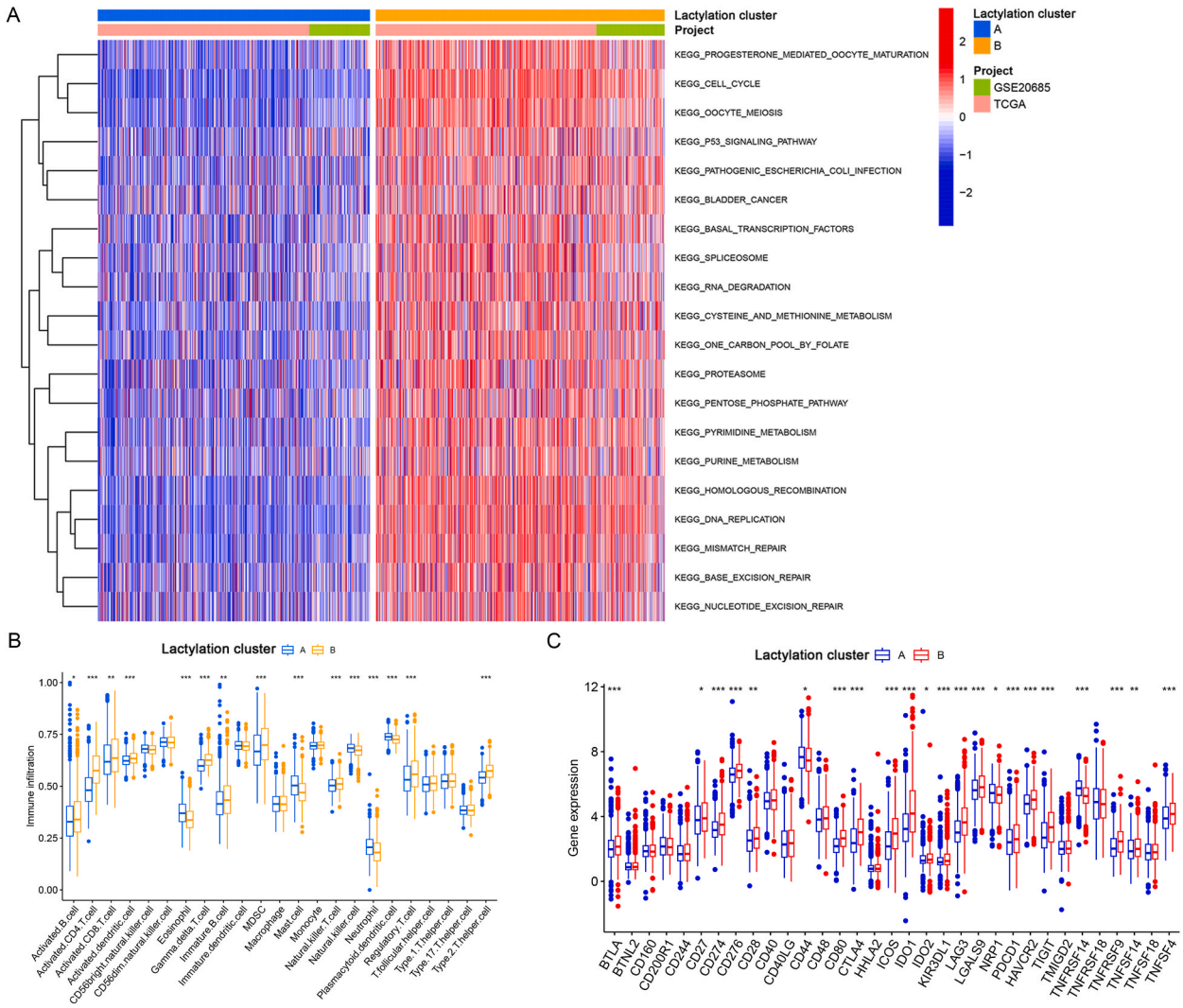

Fig. 4. Features of the TME in the lactylation clusters identified in breast cancer. (A) Comparison of the GSVA of biological pathways between the two lactylation clusters in breast cancer. Abundance of 23 infiltrating immune cell types (B) and the expression levels of 32 immune checkpoints (C) in the two lactylation clusters. TME, tumor microenvironment; GSVA, gene set variation analysis.

whereas the other lactylation-related genes did not show significant mutations (Fig. 2C).

We also constructed a correlation network using 20 lactylation-related genes (*H2AX* and *PKM2* were not included in TCGA or GSE20685 datasets), as shown in Fig. 2D. In this network, the red lines indicate positive correlations between lactylation-related genes, whereas the blue lines indicate negative correlations. In addition, univariate Cox regression analysis revealed significant differences in OS between patients with high and low lactylation-related gene expression (Table S1). Specifically, almost all lactylation-related genes were identified as "risk" factors.

3.2. Identification of lactylation clusters in breast cancer

To further investigate the expression patterns of lactylation-related genes involved in breast cancer tumorigenesis, we integrated the breast cancer data from TCGA database and GSE20685 into the TCGA-GSE cohort (N = 1420). Using the R package ConsensusClusterPlus, we performed unsupervised clustering analysis and selected k = 2 based on the empirical CDF plots. These plots show that k = 2 had the highest within-group correlations and the lowest between-group correlations compared to other values (Figs. S1 and S2). Two distinct lactylation-related gene expression patterns, lactylation clusters A and B, were observed. Furthermore, patients with breast cancer in the TCGA-GSE cohort were completely separated (Fig. 3A). To evaluate survival outcomes, we performed Kaplan–Meier (K-M) survival analysis for the two clusters. This analysis revealed poor OS in patients in cluster B (Fig. 3B). Finally, we examined the clinical and pathological characteristics of both breast cancer clusters and lactylation-related gene expression levels (Fig. 3C).

3.3. Characteristics of the TME related to lactylation clusters of breast cancer

To comprehensively analyze the role of lactylation-related genes in breast cancer TME, we performed GSVA enrichment analysis. As shown in Fig. 4A, cluster B was significantly enriched in various pathways and processes, including progesterone-mediated oocyte maturation, cell cycle, and oocyte meiosis. Furthermore, using ssGSEA, we observed significant differences in the infiltration of

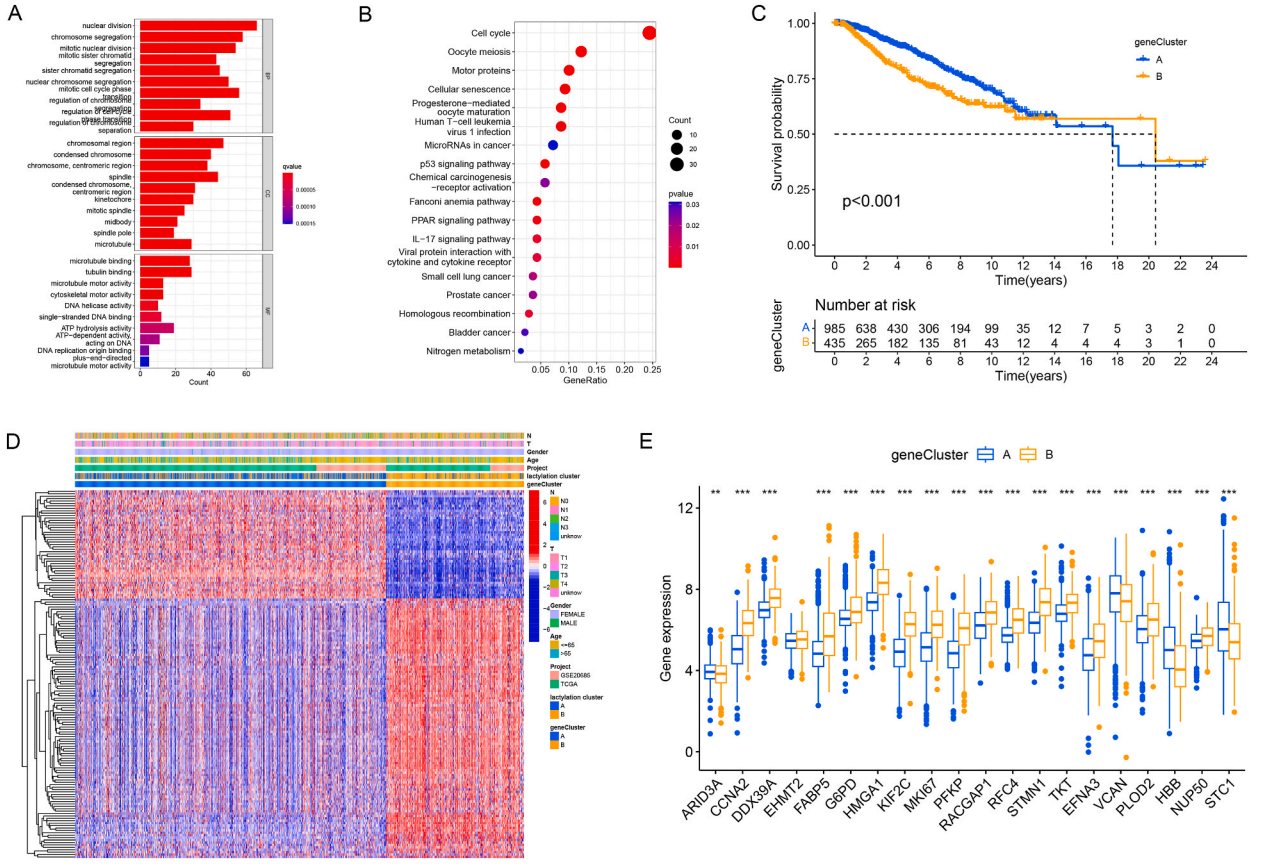

Fig. 5. Identification of gene subtypes based on lactylation clusters of breast cancer. GO (A) and KEGG (B) enrichment analyses of DEGs among the two lactylation clusters. (C) OS of the two gene subtypes. (D) Heatmap for the connections between clinicopathologic features and the two gene subtypes. (E) Differences in the expression of 20 lactylation-related genes among the two gene subtypes. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differentially expressed genes.

different innate immune cell populations between lactylation clusters A and B (Fig. 4B). We also evaluated the expression of 32 immune checkpoints in the two clusters and found that the expression of most genes related to immune checkpoints (*BTLA*, *CD27*, *CD274*, *CD276*, *CD28*, *CD80*, *CTLA4*, *ICOS*, *IDO1*, *IDO2*, *KIR3DL1*, *LAG3*, *LGALS9*, *PDCD1*, *HAVCR2*, *TIGIT*, *TNFRSF9*, *TNFSF14*, and *TNFSF4*)

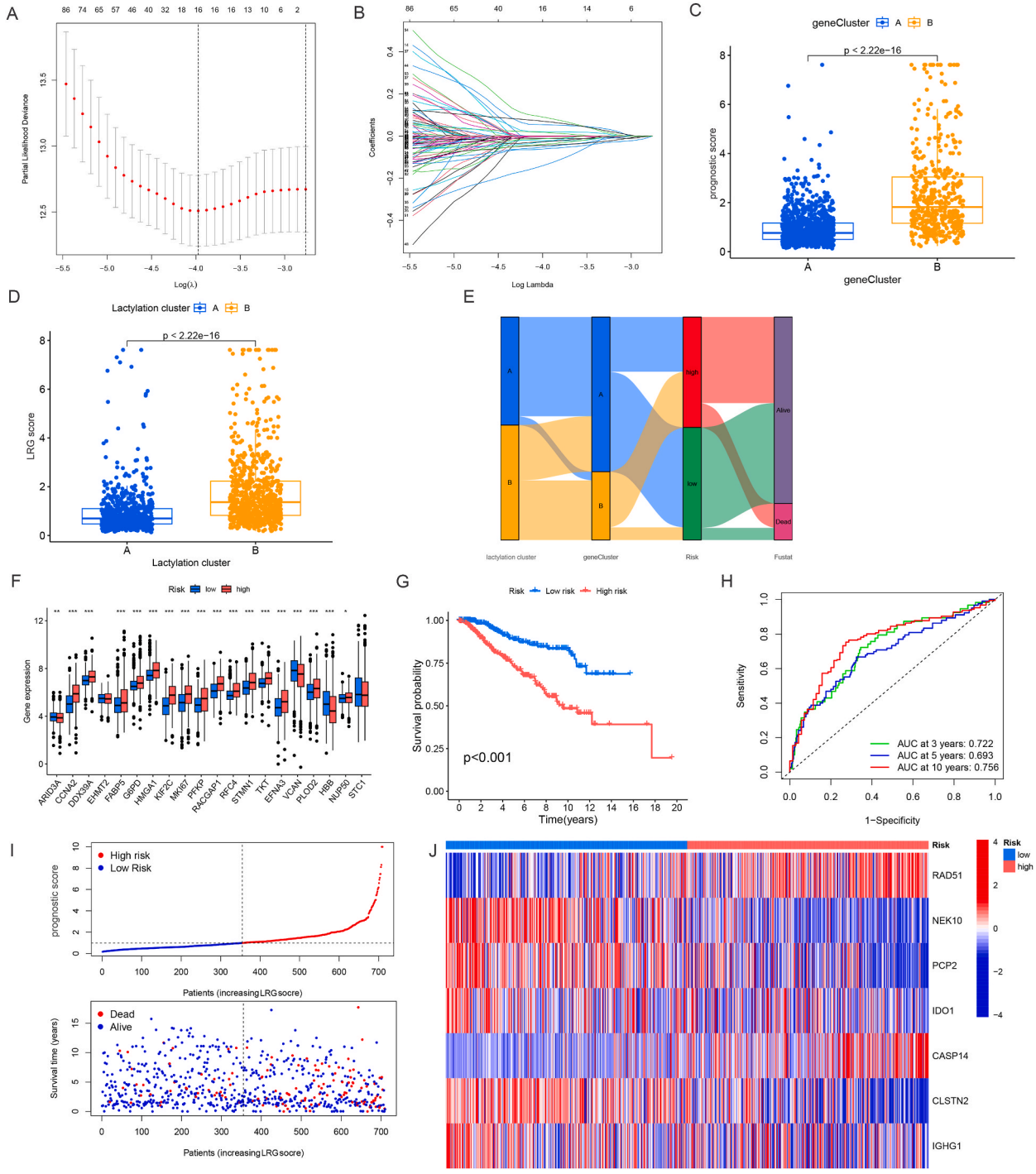

Fig. 6. Identification of a lactylation-related DEG signature and risk model in the training cohort. (A) LASSO regression for 16 candidate genes. (B) Cross-validation for 16 OS-related genes in the LASSO regression. Risk score in the different gene (C) and lactylation (D) clusters. (E) Alluvial diagram of the connection between lactylation cluster, genecluster, and prognostic score. (F) Lactylation-related gene expression in the two risk groups. (G) Kaplan–Meier curve analysis for OS. (H) ROC curve analysis demonstrated the predictive efficiency of the prognostic score. (I) Ranked dot and scatter plots of prognostic score distribution and patient survival status. (J) Heatmap of the expression of the seven OS-related genes.

was high in cluster B, whereas *CD44*, *NRP1*, and *TNFRSF14* were highly expressed in cluster A (Fig. 4C). Based on the results, we identified two clusters characterized by distinct immunological and metabolic features, suggesting that lactylation may influence the immune microenvironment and metabolic processes involved in breast cancer progression.

3.4. Identifying gene subtypes based on breast cancer lactation clusters

Subsequently, we identified 290 lactylation cluster-related DEGs by further analysis using the "limma" package to explore the distinct biological behaviors of each cluster (see Table S2). First, functional enrichment, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed for the lactylation cluster-related DEGs (Fig. 5A and B, and Tables S3 and S4). Subsequently, univariate Cox regression analysis was used to identify 153 genes with significant prognostic value ($p < 0.05$) for further investigation (Table S5). Furthermore, a consensus approach was used to classify the patients into two gene subtypes based on 153 prognostic genes (Fig. S2). As shown in Fig. 5C, patients belonging to subtype B had the most favorable OS, whereas those belonging to cluster A had the poorest OS ($p < 0.001$). A comparison of the clinicopathological characteristics and expression of DEGs

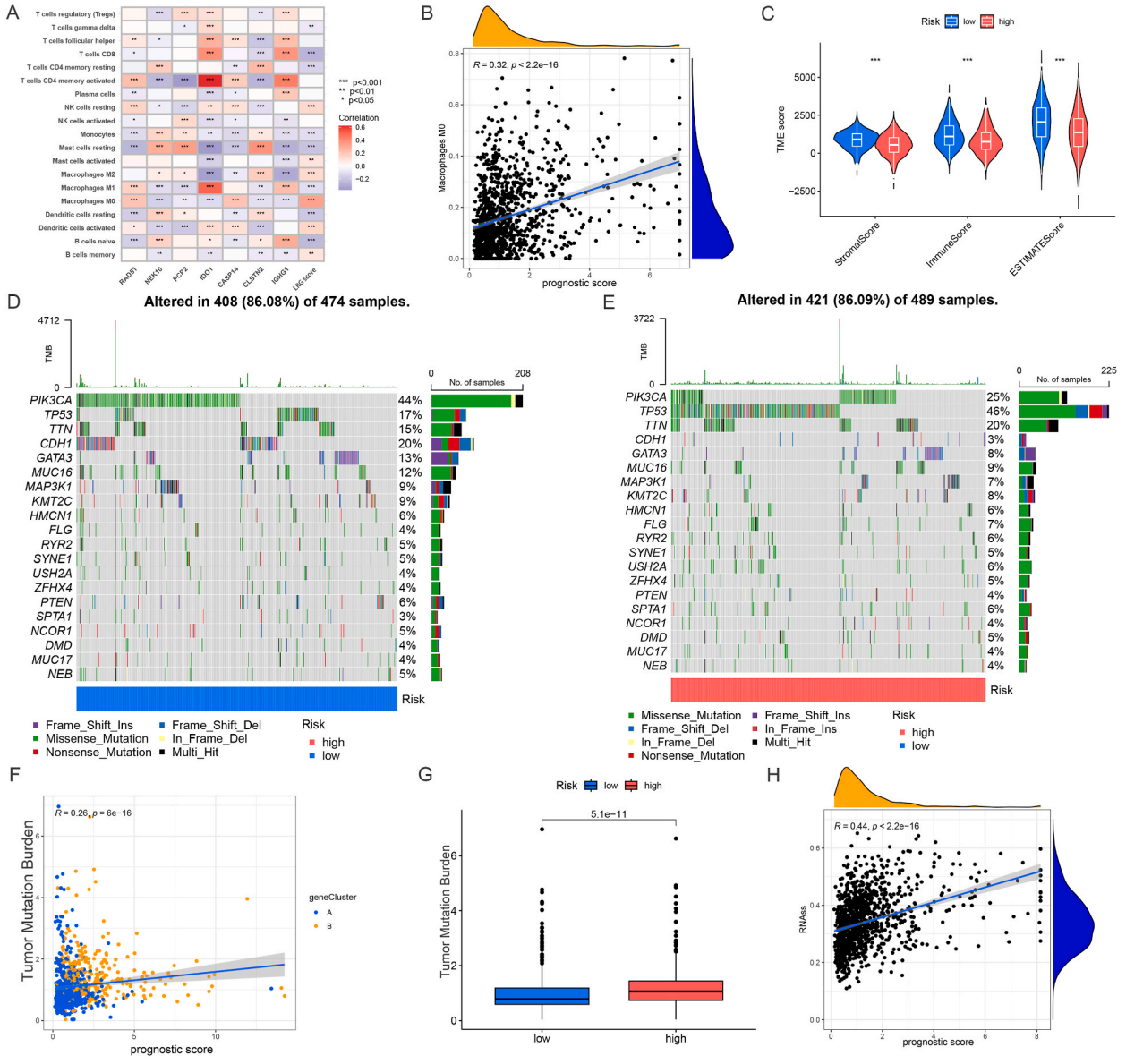

Fig. 7. Immune microenvironment and TMB of breast cancer tissues with different prognostic scores. (A) Correlation between immune cells with seven OS-related genes and prognostic score. (B) Correlation between prognostic score and M0 macrophages. (C) TME score of the two risk groups. Somatic mutation frequency rate based on the low- (D) and high- (E) risk groups. (F) Correlation between TMB and prognostic score. (G) TMB of the two risk groups. (H) Relationship between prognostic score and RNAss. TMB, tumor mutation burden.

between the two gene subtypes is shown in a heatmap (Fig. 5D). Finally, we observed the differential expression patterns of lactylation-related genes between the two gene subtypes (Fig. 5E).

3.5. Development and validation of the lactylation-related DEG signature and prognostic score

To investigate the mechanisms underlying breast cancer progression, we selected 16 prognostic subtype-related genes identified by Lasso-penalized Cox regression analysis (Fig. 6A and B). The final risk score was based on seven prognostic subtype-related gene signatures. The correlation coefficients are listed in Supplementary Table S6. The prognostic score was calculated based on the expression levels of these genes as follows:

$$Prognostic\ score = 0.254 \times RAD51 + 0.121 \times CASP14 - 0.168 \times NEK10 - 0.142 \times PCP2 - 0.262 \times IDO1 - 0.130 \times CLSTN2 - 0.069 \times IGHG1$$

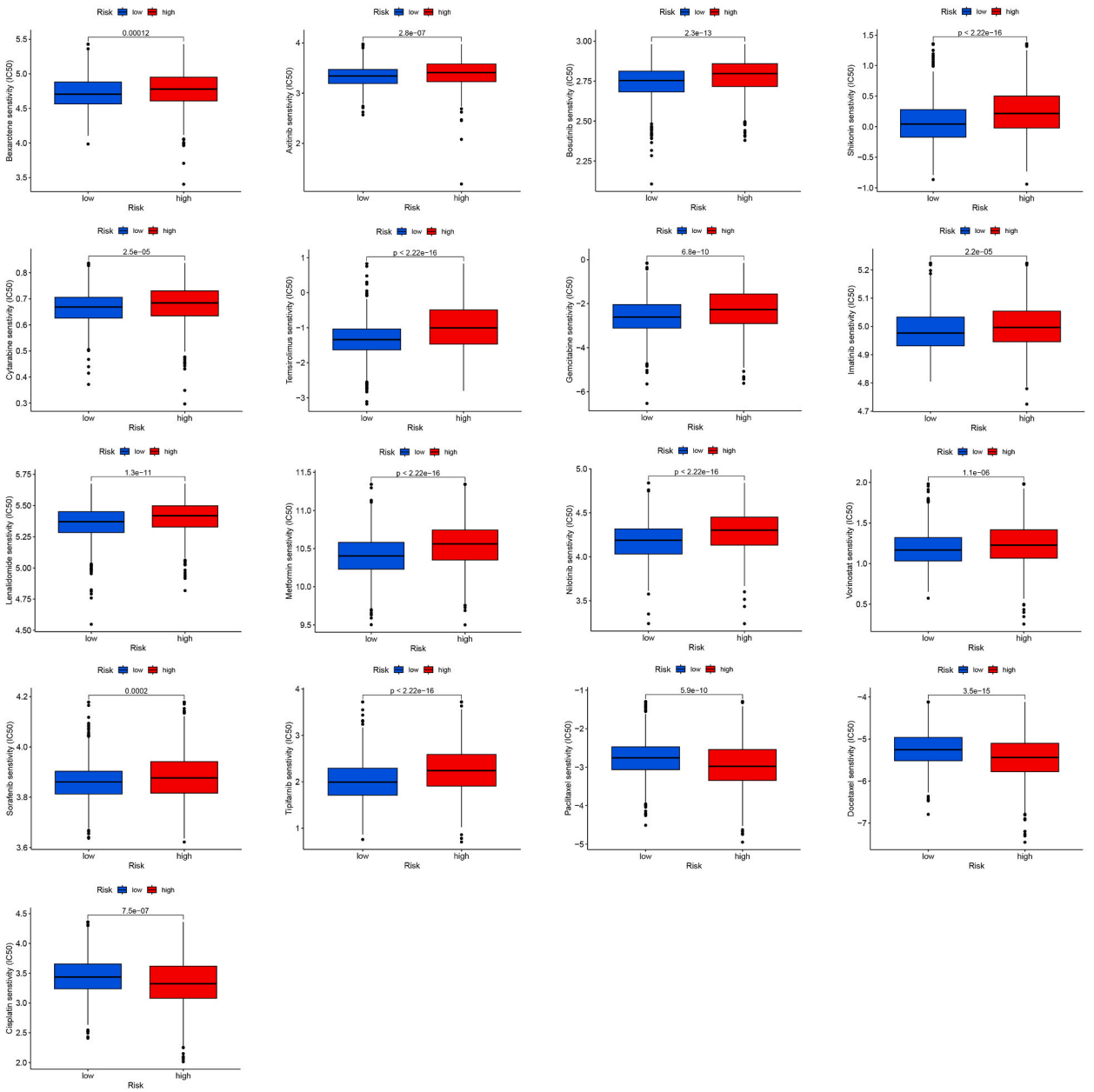

Fig. 8. Drug sensitivity analysis of patients with breast cancer in the low- and high-risk groups.

The concordance between the prognostic score and survival is highly favorable, as evidenced by a c-index of 0.652. To understand the characteristics of patients with different prognostic scores, we divided the breast cancer training cohort into high and low prognostic scores based on the median prognostic score. A scatter plot was created to visualize the distribution of patients within lactylation clusters, gene clusters, and prognostic scores (Fig. 6C and D). Cluster B had significantly higher prognostic scores than cluster A, and the lactylation cluster B had higher prognostic scores than cluster A. The alluvial plot shows the distribution of patients with breast cancer within the two lactylation clusters, two gene clusters, and two prognostic score groups (Fig. 6E). There were also significant differences in the expression of lactylation-related genes between the high and low prognostic score groups (Fig. 6F). Kaplan–Meier analysis confirmed a worse prognosis in the high prognostic score group than in the low prognostic score group (Fig. 6G). In addition, time-dependent receiver operating characteristic (ROC) analysis showed that the lactylation-related DEG signature had good predictive ability for 3-year, 5-year, and 10-year survival (Fig. 6H), which was further verified in the test cohort (Fig. S3). Additionally, patients in the high prognostic score group had a higher probability of death, as shown in the survival distribution plot (Fig. 6I). The heat map shows the expression of the seven genes in relation to the prognostic score (Fig. 6J).

3.6. Immune activity and tumor mutation burden (TMB) analysis with different prognostic scores

The immune microenvironment plays a crucial role in tumorigenesis and response to immunotherapy. Therefore, our study aimed to investigate the TME landscape of patients with breast cancer categorized into high- and low-risk groups. Initially, we observed a strong association between the prognostic score and infiltration of multiple immune cells (Fig. 7A). As the prognostic score increased, there was a gradual rise in the proportion of macrophage M0 cells (Fig. 7B–S4) with a correlation coefficient of 0.32 ($R = 0.32$). These findings suggest that M0 macrophages may be a significant contributing factor to the unfavorable prognosis observed in patients with breast cancer [20]. Additionally, by examining the expression profile, we estimated the stromal and immune scores for both the high- and low-risk groups (Fig. 7C). Furthermore, we analyzed the 20 most common gene mutations in these groups and observed higher *PIK3CA* mutation rates in the low-risk group than in the high-risk group, whereas *TP53* expression exhibited the opposite trend (Fig. 7D and E). Moreover, there was a notable positive correlation between the prognostic score and TMB, which was significantly higher in the high-risk group than in the low-risk group (Fig. 7F and G). Furthermore, as the prognostic score increased, tumor cell stemness showed an upward trend (Fig. 7H).

3.7. Prediction of anticancer therapy response in high- and low prognostic score groups in breast cancer

We assessed the responses of the high and low prognostic score groups to both conventional and innovative anticancer agents. The low prognostic score group displayed increased sensitivity to bexarotene, axitinib, bosutinib, shikonin, cytarabine, temsirolimus, gemcitabine, zmatinib, lenalidomide, metformin, nilotinib, vorinostat, sorafenib, and tipifarnib. Conversely, the high prognostic score group exhibited a greater responsiveness to paclitaxel, docetaxel, and cisplatin (Fig. 8). These results suggest that the prognostic score has the potential to serve as a predictor of anticancer therapy sensitivity, indicating that targeted small-molecule inhibitors may be more suitable for low-risk patients, whereas chemotherapy may be more appropriate for high-risk patients.

4. Discussion

Recently, increasing evidence has highlighted the various roles of lactate in tumor biology. Lactate not only serves as an important source of nutrients for tumor cells but also contributes to tumor growth, proliferation, metastasis, drug resistance, and immune suppression. For example, it can create an acidic immune microenvironment and increase the expression of proteins that confer tumor resistance [21,22]. Lactate also plays a significant role in epigenetic modifications by influencing macrophage polarization via histone lactylation [4]. Consequently, the role of lactylation in tumorigenesis and cancer progression has become an area of growing interest. However, few studies have explored the relationship between lactation and breast cancer. Therefore, comprehensive studies in this field are crucial.

In our study, we first summarized the expression and mutation patterns of lactylation-related genes based on the breast cancer data from TCGA database and GSE20685 datasets. Although the global mutation rate was only 6.66 %, most lactylation-related genes exhibited significant differences in expression and prognosis when comparing breast cancer and normal samples. Using an unsupervised clustering algorithm, we divided the patients with breast cancer into two lactylation clusters: clusters A and B. Patients in cluster B demonstrated more advanced clinicopathological characteristics and worse OS than those in cluster A. Examination of the TME revealed that cluster B was enriched with a large number of immune cells and important immune checkpoints. This indicates a close association between these clusters and the TME in breast cancer, suggesting that lactylation-related genes play a critical role in immune regulation within breast cancer.

We identified two gene subtypes based on the DEGs of the two lactylation clusters. To further investigate the role of lactylation in breast cancer progression and TME, we developed a lactylation-related DEG signature and calculated the prognostic score using seven prognostic genes (*RAD51*, NEK10, *PCP2*, *IDO1*, *CASP14*, *CLSTN2*, and *IGHG1*).

RAD51 is involved in homologous DNA recombination and repair. Its suppression sensitizes cancer cells to DNA-damaging drugs, whereas high levels promote resistance and metastasis in triple-negative breast cancer (TNBC) [23]. *RAD51* is a potential therapeutic target in TNBC. *NEK10*, a NIMA-related kinase, that regulates cell cycle progression. It also promotes *MEK1* activation, leading to G2/M phase arrest and *ERK1/2* phosphorylation. *NEK10* knockdown inhibits *MEK1* and *ERK1/2* phosphorylation. *NEK10* is highly expressed in luminal breast cancer tissues, correlating with the tumor stage and indicating improved survival outcomes [24]. *IDO1*

encodes the indoleamine 2,3-dioxygenase (IDO) enzyme, which contributes to the immunosuppressive TME when expressed at higher levels [25]. *CASP14* belongs to the caspase family and is involved in apoptosis. Caspase activation is essential for cell death. High *CASP14* expression in breast cancer, particularly triple-negative breast cancer, is associated with aggressive phenotypes and poor survival outcomes [26]. Overexpression of IgG1 heavy chain (*IGHG1*) is associated with poor survival in breast cancer. Silencing *IGHG1* decreases the viability, invasion, proliferation, and epithelial-mesenchymal transition of breast cancer cells and promotes apoptosis. Additionally, *IGHG1* silencing enhances the sensitivity of breast cancer cells to paclitaxel and cisplatin. Activation of the *AKT* pathway by *IGHG1* contributes to the malignant progression of breast cancer. Therefore, targeting *IGHG1* may be a promising therapeutic strategy for breast cancer [27]. We validated the expression of these prognosis-related genes encoding proteins in normal and breast cancer tissues using the ualcan database. Among them, *RAD51*, *PCP2*, *IDO1*, *CASP14*, and *IGHG1* exhibited significant differences in expression between breast cancer and normal breast tissues (Fig. S5), suggesting their important roles in breast cancer development.

Additionally, we validated the signature and prognostic scores in the testing cohorts, and patients with high prognostic scores exhibited a worse prognosis. We created a quantitative nomogram based on prognostic score and tumor stage to aid in the prognostic stratification of patients with breast cancer and to potentially guide clinical decisions.

Tumorigenesis is attributed not only to genetic abnormalities but also to immune dysregulation. Disruption of the immune function and abnormal immune factors can impair immune surveillance, thereby promoting the initiation and progression of malignant tumors [28]. It is important to note that the TME plays a critical role in tumor metastasis and can influence the efficacy of targeted therapies. To explore the immune cell landscape across different risk subtypes, we analyzed the relationship between 22 immune cell types and the prognostic scores. In the high-risk group with poor survival, we observed a significant increase in M0 macrophage infiltration, suggesting their crucial role in breast cancer progression. Furthermore, we found that each of the seven risk genes correlated with the level of M0 macrophage infiltration, with *CASP14* showing the highest correlation. This indicated a potential association between *CASP14* and M0 macrophages, which is an interesting direction for further research. We also analyzed the expression levels of these prognosis-related genes in different cell types using the TISCH database. Among them, *CASP14* and *PCP2* showed significantly increased expression in breast cancer tissues, while *IDO1* exhibited abundant expression in mono/macro cells. *RAD51* exhibited upregulation in Tprolif cells (Fig. S6), suggesting the important roles of these genes in immune regulation. Moreover, patients with different prognostic scores exhibited significantly different responses to antitumor drugs, with those with low prognostic scores being more sensitive to targeted small-molecule inhibitors, whereas patients with high prognostic scores were more suitable for chemotherapy.

To the best of our knowledge, this is the first study to investigate lactylation-related genes as prognostic signatures of breast cancer. However, our study has certain limitations. First, the results need to be further validated by analyzing clinical samples and experimental data, which will provide more substantial support for the conclusions of this study. Second, the biological functions and mechanisms of the seven risk genes involved in breast cancer progression require further investigation. Despite these limitations, our findings shed light on potential prognostic markers and therapeutic targets for breast cancer, paving the way for future studies.

5. Conclusion

In summary, our study identified 22 lactylation-related genes associated with the prognosis of patients with breast cancer. Using these genes, we successfully classified patients into two distinct subtypes and developed a prognostic signature based on seven of these genes, which accurately predicted patient outcomes. Furthermore, our analysis uncovered potential drug targets for high-risk patients and provided insights into the tumor immune microenvironment. Our findings suggest that incorporating lactylation-related gene-based scores into clinical practice could be a valuable tool for the prognostic prediction of breast cancer. However, further studies are necessary to validate our findings and investigate the underlying mechanisms by which these genes contribute to breast cancer progression.

Funding

This work was supported by the Shaanxi Key Research and Development Program of China (No. 2021SF-101) and the Foundation of the Health Commission of Shaanxi Province (2022B003).

Ethics declarations

Informed consent was not required because the data used in this study were obtained from public databases.

CRediT authorship contribution statement

Yangchi Jiao: Writing – original draft, Visualization, Validation, Software, Investigation, Conceptualization. **Fuqing Ji:** Writing – original draft, Funding acquisition. **Lan Hou:** Conceptualization. **Yonggang Lv:** Conceptualization. **Juliang Zhang:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all those who provided assistance in this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e24777.

References

[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 J. Clin. 71 (2021) 209–249, https://doi.org/10.3322/caac.21660.
[2] N. Harbeck, F. Penault-Llorca, J. Cortes, M. Gnant, N. Houssami, P. Poortmans, K. Ruddy, J. Tsang, F. Cardoso, Breast cancer, Nat. Rev. Dis. Prim. 5 (2019) 66, https://doi.org/10.1038/s41572-019-0111-2.
[3] S. Song, M. Zhang, P. Xie, S. Wang, Y. Wang, Comprehensive analysis of cuproptosis-related genes and tumor microenvironment infiltration characterization in breast cancer, Front. Immunol. 13 (2022) 978909, https://doi.org/10.3389/fimmu.2022.978909.
[4] D. Zhang, Z. Tang, H. Huang, G. Zhou, C. Cui, Y. Weng, W. Liu, S. Kim, S. Lee, M. Perez-Neut, J. Ding, D. Czyz, R. Hu, Z. Ye, M. He, Y.G. Zheng, H.A. Shuman, L. Dai, B. Ren, R.G. Roeder, L. Becker, Y. Zhao, Metabolic regulation of gene expression by histone lactylation, Nature 574 (2019) 575–580, https://doi.org/10.1038/s41586-019-1678-1.
[5] N. Wan, N. Wang, S. Yu, H. Zhang, S. Tang, D. Wang, W. Lu, H. Li, D.G. Delafield, Y. Kong, X. Wang, C. Shao, L. Lv, G. Wang, R. Tan, N. Wang, H. Hao, H. Ye, Cyclic immonium ion of lactyllysine reveals widespread lactylation in the human proteome, Nat. Methods 19 (2022) 854–864, https://doi.org/10.1038/s41592-022-01523-1.
[6] Z. Yang, C. Yan, J. Ma, P. Peng, X. Ren, S. Cai, X. Shen, Y. Wu, S. Zhang, X. Wang, S. Qiu, J. Zhou, J. Fan, H. Huang, Q. Gao, Lactylome analysis suggests lactylation-dependent mechanisms of metabolic adaptation in hepatocellular carcinoma, Nat. Metab. 5 (2023) 61–79, https://doi.org/10.1038/s42255-022-00710-w.
[7] D. Yang, J. Yin, L. Shan, X. Yi, W. Zhang, Y. Ding, Identification of lysine-lactylated substrates in gastric cancer cells, iScience 25 (2022) 104630, https://doi.org/10.1016/j.isci.2022.104630.
[8] L. Li, K. Chen, T. Wang, Y. Wu, G. Xing, M. Chen, Z. Hao, C. Zhang, J. Zhang, B. Ma, Z. Liu, H. Yuan, Z. Liu, Q. Long, Y. Zhou, J. Qi, D. Zhao, M. Gao, D. Pei, J. Nie, D. Ye, G. Pan, X. Liu, Glis1 facilitates induction of pluripotency via an epigenome-metabolome-epigenome signalling cascade, Nat. Metab. 2 (2020) 882–892, https://doi.org/10.1038/s42255-020-0267-9.
[9] R.A. Irizarry-Caro, M.M. McDaniel, G.R. Overcast, V.G. Jain, T.D. Troutman, C. Pasare, TLR signaling adapter BCAP regulates inflammatory to reparatory macrophage transition by promoting histone lactylation, Proc. Natl. Acad. Sci. U. S. A. 117 (2020) 30628–30638, https://doi.org/10.1073/pnas.2009778117.
[10] H. Hagihara, H. Shoji, H. Otabi, A. Toyoda, K. Katoh, M. Namihira, T. Miyakawa, Protein lactylation induced by neural excitation, Cell Rep. 37 (2021) 109820, https://doi.org/10.1016/j.celrep.2021.109820.
[11] J. Yu, P. Chai, M. Xie, S. Ge, J. Ruan, X. Fan, R. Jia, Histone lactylation drives oncogenesis by facilitating m6A reader protein YTHDF2 expression in ocular melanoma, Genome Biol. 22 (2021) 85, https://doi.org/10.1186/s13059-021-02308-z.
[12] Z. Cheng, H. Huang, M. Li, X. Liang, Y. Tan, Y. Chen, Lactylation-related gene signature effectively predicts prognosis and treatment responsiveness in hepatocellular carcinoma, Pharmaceuticals 16 (2023) 644, https://doi.org/10.3390/ph16050644.
[13] H. Yang, X. Zou, S. Yang, A. Zhang, N. Li, Z. Ma, Identification of lactylation related model to predict prognostic, tumor infiltrating immunocytes and response of immunotherapy in gastric cancer, Front. Immunol. 14 (2023) 1149989, https://doi.org/10.3389/fimmu.2023.1149989.
[14] X. Zhang, Y. Li, Y. Chen, Development of a comprehensive gene signature linking hypoxia, glycolysis, lactylation, and metabolomic insights in gastric cancer through the integration of bulk and single-cell RNA-seq data, Biomedicines 11 (2023) 2948, https://doi.org/10.3390/biomedicines11112948.
[15] Y. Ct, E.T. L Gy, Sensitivity analysis for survival prognostic prediction with gene selection: a copula method for dependent censoring, Biomedicines 11 (2023), https://doi.org/10.3390/biomedicines11030797.
[16] M.D. Wilkerson, D.N. Hayes, ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking, Bioinformatics 26 (2010) 1572–1573, https://doi.org/10.1093/bioinformatics/btq170.
[17] G. Tsiliki, C.R. Munteanu, J.A. Seoane, C. Fernandez-Lozano, H. Sarimveis, E.L. Willighagen, RRegrs: an R package for computer-aided model selection with multiple regression models, J. Cheminf. 7 (2015) 46, https://doi.org/10.1186/s13321-015-0094-2.
[18] T. Emura, S. Matsui, H.-Y. Chen, compound.Cox: Univariate feature selection and compound covariate for predicting survival, Comput. Methods Progr. Biomed. 168 (2019) 21–37, https://doi.org/10.1016/j.cmpb.2018.10.020.
[19] W. Yang, J. Soares, P. Greninger, E.J. Edelman, H. Lightfoot, S. Forbes, N. Bindal, D. Beare, J.A. Smith, I.R. Thompson, S. Ramaswamy, P.A. Futreal, D.A. Haber, M.R. Stratton, C. Benes, U. McDermott, M.J. Garnett, Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells, Nucleic Acids Res. 41 (2013) D955–D961, https://doi.org/10.1093/nar/gks1111.
[20] Z. Zhang, R. Hao, Q. Guo, S. Zhang, X. Wang, TP53 mutation infers a poor prognosis and is correlated to immunocytes infiltration in breast cancer, Front. Cell Dev. Biol. 9 (2021) 759154, https://doi.org/10.3389/fcell.2021.759154.
[21] J.D. Rabinowitz, S. Enerbäck, Lactate: the ugly duckling of energy metabolism, Nat. Metab. 2 (2020) 566–571, https://doi.org/10.1038/s42255-020-0243-4.
[22] K.G. de la Cruz-López, L.J. Castro-Muñoz, D.O. Reyes-Hernández, A. García-Carrancá, J. Manzo-Merino, Lactate in the regulation of tumor microenvironment and therapeutic approaches, Front. Oncol. 9 (2019) 1143, https://doi.org/10.3389/fonc.2019.01143.
[23] J.O. Lee, M.J. Kang, W.S. Byun, S.A. Kim, I.H. Seo, J.A. Han, J.W. Moon, J.H. Kim, S.J. Kim, E.J. Lee, S. In Park, S.H. Park, H.S. Kim, Metformin overcomes resistance to cisplatin in triple-negative breast cancer (TNBC) cells by targeting RAD51, Breast Cancer Res. 21 (2019) 115, https://doi.org/10.1186/s13058-019-1204-2.
[24] W.-L. Gao, L. Niu, W.-L. Chen, Y.-Q. Zhang, W.-H. Huang, Integrative analysis of the expression levels and prognostic values for NEK family members in breast cancer, Front. Genet. 13 (2022) 798170, https://doi.org/10.3389/fgene.2022.798170.
[25] J.-L. Wei, S.-Y. Wu, Y.-S. Yang, Y. Xiao, X. Jin, X.-E. Xu, X. Hu, D.-Q. Li, Y.-Z. Jiang, Z.-M. Shao, GCH1 induces immunosuppression through metabolic reprogramming and Ido1 upregulation in triple-negative breast cancer, J Immunother Cancer 9 (2021) e002383, https://doi.org/10.1136/jitc-2021-002383.

[26] T. Handa, A. Katayama, T. Yokobori, A. Yamane, J. Horiguchi, R. Kawabata-Iwakawa, S. Rokudai, P. Bao, N. Gombodorj, B. Altan, K. Kaira, T. Asao, H. Kuwano, M. Nishiyama, T. Oyama, Caspase14 expression is associated with triple negative phenotypes and cancer stem cell marker expression in breast cancer patients, J. Surg. Oncol. 116 (2017) 706–715, https://doi.org/10.1002/jso.24705.
[27] Y. Jin, L. Qiu, W. Bao, M. Lu, F. Cao, H. Ni, B. Zhao, High expression of IGHG1 promotes breast cancer malignant development by activating the AKT pathway, Cell Cycle 22 (2023) 718–731, https://doi.org/10.1080/15384101.2022.2147141.
[28] P.-Y. Chu, Y.-D.T. Tzeng, Y.-H. Chiu, H.-Y. Lin, C.-H. Kuo, M.-F. Hou, C.-J. Li, Multi-omics reveals the immunological role and prognostic potential of mitochondrial ubiquitin ligase MARCH5 in human breast cancer, Biomedicines 9 (2021) 1329, https://doi.org/10.3390/biomedicines9101329.