RNA-seq-based elucidation of lactylation in breast cancer

Keywords
Breast cancer, Lactylation, PGK1, Tumor metabolism, Tumor immunity

Abstract

Introduction

Lactylation is the covalent modification of histones using lactate as a small molecule precursor, playing a role in epigenetic regulation. As a novel protein post-translational modification, it has demonstrated significant relevance in the field of cancer diagnosis and therapy. However, the interaction between lactylation and tumor cells in breast cancer has not been extensively investigated.

Material and methods

We acquired breast cancer-related data from the GEO and TCGA databases. Lactylation-related genes were identified from the differentially expressed genes. We utilized COX and LASSO regression to identify genes with significant prognostic value for constructing a prognostic model and assessing its predictive performance. This model was integrated with clinical parameters to create a nomogram. Finally, we conducted immune infiltration analysis, analyzed differences in biological functions, and assessed drug sensitivity.

Results

We ultimately identified 3 lactylation-related genes significantly associated with prognosis. These genes were used to construct a prognostic model and calculate a risk score. Using the median score, patients were divided into high-risk and low-risk groups. Notably, the low-risk group patients exhibited better prognosis and higher levels of immune infiltration. GO/KEGG enrichment analysis revealed that PGK1, the gene with the highest HR among these genes, is widely involved in immune, metabolic, and proliferative signaling pathways. Its high expression also correlates with increased sensitivity to anti-tumor drugs.

Conclusions

We showcased the potential of lactylation-based molecular clustering and prognostic profiling for predicting survival, immune status, and treatment response in breast cancer patients. Additionally, we envision the utilization of PGK1 as a diagnostic marker and therapeutic target in the cancer.

RNA-seq-based elucidation of lactylation in breast cancer

1 Introduction

In the gradual progression from normal cells to cancer, these cells acquire certain acquired functions, including sustaining proliferative signaling, resisting cell death, evading growth suppressors, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis, ultimately leading to tumor formation and deterioration [1]. In recent years, as our understanding of cancer has deepened, additional characteristics of tumors have emerged, such as deregulating cellular energetics, avoiding immune destruction, genome instability and mutation, and tumor-promoting inflammation [2]. Among malignant tumors, breast cancer is the most common in women globally, accounting for 31% of all newly diagnosed cancers. It is a type of malignancy that develops through a multivariate-mediated process involving multiple steps and stages. Importantly, mutations in the BRCA1 and BRCA2 genes not only increase the hereditary nature of breast cancer but also contribute to the complexity of this disease. While early breast cancer screening and treatment advancements have led to reduced mortality rates [3], the rising incidence of breast cancer emphasizes the urgency for targeted interventions. Our study delves into understanding lactylation in breast cancer, aiming to inform tailored treatment strategies and improve patient outcomes.

The deregulation of cellular energetics in cancerous diseases is evident in the downregulation of cell proliferation control and the adaptation of energy metabolism. Under aerobic conditions, normal cells undergo aerobic oxidation of glucose. However, in hypoxic conditions, cells further reduce the pyruvate generated from glycolysis into lactate within the cytoplasm. The Warburg effect indicates that even when oxygen is abundant, cancer cells restructure their energy metabolism by constraining the glucose metabolism process to glycolysis, leading to the production of significant amounts of lactate [4]. Glycolysis-driven energy supply is associated with cancer genes like RAS and MYC, as well as tumor suppressor genes like TP53. Alterations in these genes within cancer cells grant them abilities such as enhanced cell proliferation, resistance to cell death, and evasion of apoptosis, ultimately promoting tumor development [5,6]. Lactate, a metabolic byproduct generated from glucose through glycolysis catalyzed by lactate dehydrogenase (LDH), plays crucial biological roles as an energy source, an immune regulatory molecule, and a participant in gluconeogenesis. LDH exists in two distinct subtypes, LDHA and LDHB, each with specific functions [7]. LDHA is responsible for

converting pyruvate into lactate, and its expression is regulated by proteins like hypoxia inducible factor-1α (HIF1α), c-Myc, and p53[8]. In contrast, LDHB converts lactate back into pyruvate to promote oxidative metabolism, and its loss or downregulation is closely associated with the development and poorer prognosis of cancers like pancreatic and liver cancer [9,10]. Additionally, lactate produced by cancer cells can be secreted into the extracellular environment, serving as a signaling molecule to further promote cancer development [7]. It can stimulate endothelial cells to secrete VEGF protein and activate the NF-κB/IL-8 (CXCL8) pathway, thereby facilitating tumor-related angiogenesis [11,12]. Lactate also plays a vital role in maintaining an acidic environment, regulating the tumor microenvironment (TME) through processes like cell invasion, metastasis, and immune escape, thereby sustaining tumor growth [13]. As a result, lactate has become a potent molecule influencing the behavior of every cell within the TME.

In 2019, Zhang and colleagues introduced a groundbreaking concept called 'lactylation' – a novel post-translational modification. It involves using lactate, a product of cellular metabolism, as a small-molecule precursor to induce lactylation of histone lysine, thereby regulating gene expression. This opened up a new frontier in the study of protein lactylation. They employed mass spectrometric analysis to detect a molecular weight shift of 72.021 Daltons on histone lysine residues in the breast cancer MCF-7 cell line. Through isotopic labeling methods and various in vitro and in vivo experiments, they convincingly demonstrated the widespread presence of lysine lactylation. Furthermore, they found that the abundance of lactylation in MCF-7 cells is positively correlated with lactate concentration, and it is regulated by glycolysis and hypoxia induction [14]. Increasingly, research has shown the close association of lactylation with inflammatory diseases, tumors, neurodegenerative diseases, and more [15-17]. While the research on protein lactylation is still in its early stages, it has opened up new horizons for targeting lactate metabolism, transport, and immune-related anti-cancer strategies. Our study, based on a literature search, revealed limited reports on the functional role of lactylation in breast cancer. Therefore, our research aims to identify differentially expressed genes related to lactylation in breast cancer, construct a prognostic model for more accurate patient prognosis prediction, and explore effective cancer therapies. Our study not only advances our understanding of the interaction between lactylation and cancer but also has the potential to uncover promising cancer immunotherapy targets, contributing to the fight against breast cancer.

2 Methods

2.1 Data download and processing

We obtained breast cancer RNA expression data, CNV files, and corresponding clinicopathological information from the TCGA-BRCA project (GDC (cancer.gov)). Clinical parameters and normalized gene expression data from the GSE162228 (GEO Accession viewer (nih.gov)) breast cancer dataset available in the GEO database, which consists of samples from Taiwanese breast cancer patients [18]. To ensure data integrity, samples lacking essential clinicopathological or survival information were excluded. Lactylation is facilitated by specific enzymes or protein modifiers. Therefore, lactylation-related genes encompass those encoding these enzymes and genes associated with the substrate proteins involved in lactylation. We included a total of 332 lactylation-related genes for subsequent analysis [19]. The lactylation-associated gene Protein-Protein Interaction (PPI) network was constructed using the STRING website (STRING: functional protein association networks (string-db.org)). We calculated the frequency of copy number variations in lactylation-related genes by analyzing changes in gene copy numbers in breast cancer samples from the TCGA database. Subsequently, the "RCircos" package in R language was utilized to create a circular gene copy number map. Finally, COX and co-expression analyses were used to generate the prognostic network of lactylation-related genes.

2.2 Screening of lactylation prognosis-related genes in breast cancer

First, we began by identifying lactylation-related genes with prognostic value through differential expression analysis and univariate COX regression analysis within the entire dataset of breast cancer samples. Subsequently, we narrowed down the list of prognosis-related genes using LASSO regression. Genes with confirmed prognostic significance were then selected through multivariate COX regression analysis, and we proceeded to construct prognostic models. To calculate the risk score for each breast cancer sample, we utilized the accumulation method by multiplying the coefficient with the gene's expression level. Based on the median value, we categorized the samples into high-risk and low-risk groups and examined the prognostic differences between these groups. We employed the Kaplan-Meier method to generate survival curves for breast cancer patients, and these curves were visualized using the "survminer" package. Furthermore, we conducted an in-depth analysis of the clinical data and risk scores for all breast cancer patients, calculating survival times and statuses. This information was used to create a nomogram. Finally, we employed the R package "timeROC" (V0.4) to generate a receiver operating characteristic (ROC) curve for assessing the sensitivity and specificity of the risk model.

2.3 Cluster analysis

We employed the "ConsensusClusterPlus" package to conduct unsupervised clustering of breast cancer samples, based on the expression levels of lactylation-related genes. The results indicated that the samples were most effectively categorized into two distinct classes. Subsequently, we created a heat map to visualize the correlation between the expression patterns of lactylation-related genes in different clusters and the clinical information of patients. We then quantified the expression of immune cells in these distinct clusters using the ssGSEA method and presented the results through box plots. In addition, we obtained the GO/KEGG pathway files from the GSEA website and utilized the "GSEABase" and "GSVA" packages for pathway enrichment analysis and heat map visualization.

2.4 GO/KEGG analysis

We conducted the Wilcoxon test to identify DEGs in both groups. The risk score was calculated using the R package "limma," with the criteria of FDR < 0.05 and $|\log2\ FC| \geqslant 1$. For GO/KEGG enrichment analysis, we utilized the R packages "clusterProfiler" and "enrichplot."

2.5 The relationship between lactylation-related molecular patterns and the clinical features and prognosis of breast cancer

To assess the clinical relevance of the clusters generated by consensus clustering, we examined their associations with molecular patterns, clinical characteristics, and survival outcomes. Clinical characteristics encompassed age, gender, tumor staging, and lymph node staging. Furthermore, Kaplan-Meier analyses were conducted using the "survival" and "survminer" packages to evaluate differences in overall survival (OS) among the various models [20].

2.6 Establishment of a predictive nomogram

The nomogram is created to offer meaningful clinical predictions for breast cancer patients, encompassing their risk scores and other clinicopathological characteristics, with a particular focus on the 1-year, 3-year, and 5-year OS rates. We assessed the clinical validity of the established nomogram through calibration curve analysis and decision curve analysis (DCA).

2.7 Lactylation-related molecular patterns and TME in breast cancer

The ESTIMATE algorithm evaluated the StromalScore and ImmuneScore of breast cancer patients, and the CIBERSORT algorithm was employed to calculate the levels of 23 immune cell subtypes for each patient [21,22]. The infiltrating fraction of immune cells was determined using the single sample gene set enrichment analysis (ssGSEA) algorithm [23].

2.8 Drug sensitivity prediction

The half maximal inhibitory concentration (IC50) values for common anti-tumor drugs were computed using the "oncoPredict" R package to predict drug responses in breast cancer patients with varying levels of PGK1 expression.

2.9 Cell culture and transfection

The human breast cancer cell line HS578T, provided by the Medical Laboratory of Yan'an University, was utilized in this study. Cells were cultured in DMEM medium (BI, Israel) supplemented with 10% fetal bovine serum (FBS) (BI, Israel) at 37°C in a constant temperature incubator with 5% carbon dioxide. The siRNA sequence used in this research was PGK1 5'-GAGTCAATCTGCCACAGAA-3' (GenePharma, China)[24]. Previously synthesized siRNA targeting the PGK1 gene was transfected into cells using Lipo 2000 (Invitrogen, USA).

2.10 RNA isolation and quantitative real-time pcr analysis

This study utilized Quantitative RT-PCR to assess the knockdown efficacy of siRNA. Total cellular RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, USA), and RNA concentration was checked. Reverse transcription was performed using Hifair® III 1st Strand cDNA Synthesis SuperMix for qPCR (gDNA digester plus) from Yeasen Biotechnology, China. qPCR was conducted using Hieff® qPCR SYBR Green Master Mix (No Rox) from Yeasen Biotechnology, China, with GAPDH as the reference gene. The primer sequences used in this experiment were as follows: PGK1 (forward, 5'-TCACTCGGGCTAAGCAGATT-3'; reverse, 5'-CAGTGCTCACATGGCTGACT-3'). Amplification reactions were carried out using a qPCR instrument with the following conditions: 95°C for 5 min; 95°C for 10 s; 60°C for 30 s. After 40 cycles of amplification, data analysis was conducted, ensuring correct amplification and melting curves.

2.11 CCK8 assay

In this study, cell viability of HS578T cells was assessed using the Cell Counting Kit-8 (CCK-8) method (IC-1519, InCellGene, Tx. USA). Cells were seeded at a density of 1500 cells per well in a 96-well cell culture plate and then transfected with siRNA. After transfection, cells were placed back in the incubator, and 10 μL of CCK-8 reagent was added at the same time every day for detection at 0h, 24h, 48h, and 72h. Finally, their absorbance at a wavelength of 450nm was measured using a microplate reader (Molecular Devices, USA).

2.12 Scratch wound healing assay

HS578T cells, cultured in a 6-well plate, were transfected with both PGK1 siRNA and NC siRNA at a 70% confluence rate. A sterile 100 μL pipette tip was used to create cell scratches, and images were captured at 0, 12, 24, and 36 hours post-scratch to ensure consistent scratch area. Image acquisition was performed using a Nikon Ti-S fluorescent microscope.

2.13 Statistical analysis

All statistical analyses were conducted using R software (version 4.3.1). The t-test was employed to assess differences between the two groups, while the log-rank test was utilized to examine disparities between the Kaplan-Meier curves. Univariate and multivariate COX regression analyses were performed to identify risk factors associated with breast cancer prognosis. A significance level of $P < 0.05$ was considered indicative of statistical significance.

3 Results

3.1 Lactylation-related genes expression and mutation in breast cancer

We began by identifying lactylation-related genes within the DEGs of the TCGA-BRCA dataset (Figure 1A). Subsequently, we obtained a total of 83 DEGs for further analysis, and their relative expressions are depicted in Figure 1B. Using the STRING website, we conducted a PPI network analysis to elucidate interactions among these DEGs (Figure 1C). Further, we assessed the frequencies of CNV for ten prognosis-related genes through CNV files (Figure 1D). The results hinted at CNV potentially playing a regulatory role in the expression of lactylation prognosis-related genes. In Figure 1E, you can observe the CNV-altered sites on the chromosomes of lactylation prognosis-related genes.

3.2 Lactylation subgroups and their characterization in breast cancer

We combined the TCGA-BRCA and GSE162228 datasets to enhance the sample size and then identified ten prognosis-related genes among the 83 DEGs using univariate COX regression analysis and Kaplan-Meier analysis (CACYBP, G6PD, HSPE1, PGK1, PRDX1, PSMA7, PTMA, RACGAP1, RAN, and WAS). Subsequently, we illustrated the interactions, regulatory relationships, and their significance for survival in breast cancer patients using a network diagram (Figure 2A). A forest plot visually displayed the HR values of the lactylation prognosis-related genes, classifying them as high or low risk (Figure 2B). To gain insights into the connection between lactylation and breast carcinogenesis and to determine how lactylation-related genes correlate with breast cancer expression patterns, we conducted a consensus clustering analysis of breast cancer patients based on

the expression levels of DEGs. The results indicated that the optimal clustering variable was 2 (Figure 2C), and the breast cancer patients in the cohort were well-distributed into these two groups. Principal components analysis (PCA) further affirmed the clear separation between the groups (Figure 2D). Additionally, when comparing the OS of patients in the two groups, we observed that cluster B had a worse prognosis than cluster A (Figure 2E). Furthermore, we investigated the relationship between gene expression and clinicopathologic variables in different clusters, revealing significant differences between the two groups (Figure 2F). We then identified differential pathways between cluster A and cluster B through GSVA analysis. These pathways included "CITRATE-CYCLE-TCA-CYCLE," "MAPK-SIGNALING-PATHWAY," "CELL-CYCLE," "PURINE-METABOLISM," "CYSTEINE-AND-METHIONINE-METABOLISM," and "PYRIMIDINE-METABOLISM" (Figure 2G). Lastly, we analyzed variations in immune cell infiltration levels between different clusters using the ssGSEA algorithm. The results showed that cluster B exhibited relatively higher infiltration of activated CD4 T cells and type 2 T helper cells (Th2). Conversely, cluster A displayed more significant immune cell infiltration, including B cells, natural killer cells, eosinophils, macrophages, mast cells, monocytes, neutrophils, and other cell types (Figure 2H). Consequently, cluster A, characterized by higher immune infiltration levels, displayed a more favorable prognosis compared to cluster B.

3.3 Construction and evaluation of prognostic models

We initially identified genes associated with patient prognosis through univariate COX regression analysis, followed by LASSO regression analysis. The LASSO analysis revealed that, based on the optimal λ-value, gene selection stabilized and minimized partial likelihood bias when including three genes (Figure 3A-B). Consequently, we identified three lactylation-related genes significantly associated with prognosis: Risk score = (0.6409 × PGK1) - (0.3610 × PTMA) - (0.2484 × WAS). Subsequently, we divided the patients into high-risk and low-risk groups using the median risk score. Kaplan-Meier survival curves illustrated that patients in the high-risk group had significantly worse prognosis compared to those in the low-risk group (Figure 3C). Furthermore, we evaluated the predictive performance of this model using ROC curves. The results demonstrated high predictive accuracy with an AUC of 0.721, 0.644, and 0.630 at 1, 3, and 5 years, respectively (Figure 3D). Additionally, the heat map displayed the expression of the selected prognostically relevant genes (Figure 3E). The findings suggested that WAS and PTMA might act as protective factors for breast cancer, while PGK1 could be a risk factor. Notably, we constructed a Sankey diagram to visualize the relationship between different clusters, risk scores, and patients' survival status. These diagrams

revealed that the majority of cluster A corresponded to the low-risk group with a relatively favorable prognosis, while most of cluster B corresponded to the high-risk group with a less favorable prognosis (Fig. 3F). In line with the aforementioned results, Figure 3G indicated that the risk score of cluster B was higher than that of cluster A. Given the strong correlation between the risk score and patient prognosis, we incorporated clinical parameters to construct a nomogram. This nomogram assessed the OS of breast cancer patients at 1, 3, and 5 years (Figure 3H). The calibration curve of the nomogram demonstrated high accuracy between actual observed and predicted values (Figure 3I). Furthermore, the DCA curves showed that the nomogram's prediction of patients' OS at 1, 3, and 5 years outperformed individual clinicopathologic variables (Figure 3J-L). Therefore, our modeled genes exhibited strong predictive efficacy, whether grouped by the risk score derived from COX analysis or unsupervised clustering.

3.4 Immune infiltration analysis

As demonstrated in the preceding analysis, there exists a notable disparity in patient prognosis across different risk groups. To delve deeper into the disease's etiology and provide relevant insights for breast cancer immunotherapy, we assessed the correlation between the risk score and immune cell abundance using the CIBERSORT algorithm. The results unveiled variations in the distribution and relative content of immune cells among different risk groups (Figure 4A). Further scrutiny revealed that the risk score exhibited a positive correlation with the infiltration of macrophages M0, macrophages M2, and neutrophils, while displaying a significant negative correlation with the infiltration of naive B cells, activated CD8 T cells, and resting dendritic cells (Figure 4B-G). Subsequently, we conducted a specific analysis of the disparities in immune infiltration levels between distinct risk groups utilizing the ESTIMATE algorithm, which underscored that the low-risk group exhibited higher immune infiltration levels (Figure 4H). Next, we explored the relationship between genes significantly associated with lactylation prognosis and immune cell enrichment. The findings revealed a robust correlation between the two (Figure 4I). Furthermore, we evaluated the association between the risk score and stromal cells as well as immune cells within the TME using the ESTIMATE algorithm. The results indicated that the risk score exhibited a negative correlation with StromalScore, ImmuneScore, and ESTIMATEScore, implying that the low-risk group had a higher infiltration of non-tumor cells within the TME (Figure 4J).

3.5 Prognostic analysis, biological function, and drug sensitivity analysis of PGK1

PGK1 exhibited the highest HR value in both COX and LASSO regression analyses. It was also identified as one of the genes significantly associated with lactylation prognosis. Consequently, we conducted a survival analysis for PGK1. Kaplan-Meier survival curves clearly indicated that variations in PGK1 expression significantly influenced the survival outcomes of breast cancer patients ($P < 0.001$), with patients exhibiting low PGK1 expression demonstrating a more favorable prognosis (Figure 5A). In the GSE124647 dataset, we also observed a significant difference in OS and progression free survival (PFS) rates between patients with high PGK1 expression and those with low expression (Figure 5B-C). Subsequently, we delved into the distinct signaling pathways between the high PGK1 group and the low PGK1 group through GO/KEGG enrichment analysis. Notably, we identified differentially enriched pathways such as "cell cycle," "PPAR signaling pathway," "IL-17 signaling pathway," "tyrosine metabolism," "phenylalanine metabolism," and "ECM-receptor interaction" (Figure 5D-E). Previous research has elucidated that peroxisome proliferator activated receptor (PPAR), aside from regulating energy metabolism, plays a pivotal role in immune cell differentiation and fate determination [25]; interleukin 17 (IL-17) serves as a key player in immune system regulation and is a significant pro-inflammatory factor [26]; and the extracellular matrix can impact immune function by suppressing anti-tumor immune responses [27,28]. Hence, the signaling pathways we identified are extensively implicated in immunoregulation, energy metabolism, and cell proliferation. Lastly, we computed the IC50 values of breast cancer concerning commonly used anti-tumor drugs using the "oncoPredict" tool and compared them between the two groups. The results indicated that patients with high PGK1 expression exhibited increased sensitivity to epirubicin, palbociclib, ribociclib, sorafenib, cytarabine, and gemcitabine (Figure 5F-K).

3.6 Knocking down PGK1 resulted in decreased viability of HS578T cells *in vitro*

We employed the q RT-PCR method to assess the knockdown efficiency of PGK1 siRNA in HS578T breast cancer cells. After 24 hours post-transfection, we examined the expression levels of PGK1 mRNA (Figure 6A) and found a significant decrease induced by the siRNA sequences ($P<0.01$). Subsequently, CCK8 analysis revealed a notable reduction in cell viability following PGK1 gene knockdown (Figure 6B). Finally, a scratch assay was conducted to assess the impact of PGK1 knockdown on the migration capability of HS578T cells. The results indicated a significantly slower scratch closure in the PGK1 knockdown group compared to the siRNA negative control (NC) group (Figure 6C), suggesting that PGK1 knockdown may be an effective strategy to inhibit breast cancer cell proliferation and migration.

4 Discussion

Protein post-translational modification refers to the chemical alterations of proteins after translation, which can regulate protein activity, localization, folding, and interactions with other biomolecules. Proteins can undergo various forms of modification, such as acetylation, methylation, ubiquitination, and, with the advancement of high-sensitivity mass spectrometry, modifications stemming from cellular metabolites like lactylation are also gradually being discovered. In the human embryonic kidney HEK293T cell line, over expression of histone acetyltransferase p300 (p300) has been observed to enhance lysine lactylation levels, while the absence of p300 results in a reduction in histone lysine lactylation levels in HEK293T and similar cell lines [14]. Similarly, in the lactate-induced mouse macrophage system RAW 264.7, the levels of lactylation can be significantly reduced by knocking down p300 or CREB-binding protein (CBP) [29]. Additionally, Class I and Class III histone deacetylases (HDACs) play a role in de-lactylation within cells [30].

The discovery of lactylation has not only opened up new frontiers in the study of protein post-translational modification but has also proposed potential regulatory mechanisms for the role of lactate in physiological and pathological processes such as cancer, inflammation, and metabolism. Lactylation levels exhibit dynamic changes in mouse oocytes and pre-implantation embryos, and in vitro hypoxic culture reduces lactylation levels, impairing the developmental potential of pre-implantation embryos [31]. Metabolic remodeling induced by GLIS family zinc finger 1 (GLIS1) involves the generation of abundant lactate and an increase in lactylation levels on pluripotency gene promoters, enhancing reprogramming efficiency and even reprogramming of aging cells [32]. Macrophages specific expression of B-cell adapter for phosphoinositide 3-kinase (BCAP) affects the expression of repair genes by regulating lactylation levels, aiding the body in mitigating inflammatory responses [33]. Additionally, neuronal excitation in the brain elevates lactate content and lactylation levels in brain cells [34]. In Alzheimer's disease (AD) patients' brain samples, lactylation levels rise and enrich at the promoters of glycolysis genes, activating their transcription. This ultimately forms a "glycolysis/histone lactylation/pyruvate kinase M2 (PKM2)" positive feedback loop, promoting the development of AD [17]. In ocular melanoma, elevated lactylation levels upregulate YTH N6-methyladenosine RNA-binding protein 2 (YTHDF2) expression, leading to the degradation of period circadian regulator 1 (PER1) and TP53 mRNA, ultimately driving tumor initiation, progression, and unfavorable outcomes [16]. Thus, exploring the role of lactylation in breast cancer becomes highly intriguing. This not only offers insights into protein post-translational

modification in breast cancer research but also paves the way for new directions in the treatment of breast cancer patients.

In breast cancer, lactylation is closely associated with tumor growth, the immune microenvironment, and drug response [35]. In this study, we initiated our investigation by delving into breast cancer data from the GEO and TCGA databases to identify lactylation-related genes. Utilizing an unsupervised clustering approach, we categorized breast cancer patients into two clusters, namely cluster A and cluster B. Among them, cluster A displayed a more favorable prognosis and higher levels of immune infiltration. Further analysis through COX regression and LASSO regression led us to identify three lactylation-related genes (PTMA, WAS, PGK1) that hold significant prognostic value. Among these, prothymosin alpha (PTMA) shows progressively upregulated expression in esophageal squamous cell carcinoma, with significantly higher expression levels between tumors and adjacent normal tissues as the disease progresses [36]. Circ-0004277 participates in colorectal cancer cell proliferation by upregulating PTMA expression [37]. Additionally, studies indicate that levels of PTMA in tumor samples from breast cancer patients are significantly higher than in normal breast tissue, and these PTMA levels correlate positively with certain indicators of cancer progression [38]. However, in bladder cancer, PTMA exerts its tumor-suppressive role by upregulating PTEN and coordinating the nuclear factor erythroid 2-related factor 2 (NRF2) signaling pathway through tripartite motif-containing protein 21 (TRIM21) [39]. The WAS gene belongs to the Wiskott-Aldrich syndrome protein family, and N-WASP exhibits significantly downregulated expression in breast cancer, correlating with poor prognosis [40]. Similarly, WASP acts as a tumor suppressor in T cell lymphoma [41], while in prostate cancer, it enhances cancer cell invasion and metastasis [42]. WASP and its family can also regulate actin polymerization in breast cancer, promoting cell invasion and migration, thus exhibiting oncogenic functions [43]. PGK1 is an essential enzyme in the glycolysis pathway and is involved in various biological processes. In hepatocellular carcinoma, PGK1 promotes cancer cell metastasis through pathways like HIF-1α/PGK1 and MYC/PGK1 [44,45]. In colon cancer, PGK1 fosters cancer metastasis by upregulating the expression of early growth response 1 (EGR1) and cysteine-rich 61 (CYR61) [46]. In papillary thyroid carcinoma, Sirtuin 6 (SIRT6) enhances tumor invasiveness by increasing PGK1 expression to promote the Warburg effect [47]. Subsequently, we constructed a prognostic model using these three genes and assessed its efficacy. Patients were stratified into high-risk and low-risk groups based on the median risk score, revealing significant differences in patient outcomes between the groups. Furthermore, we analyzed the differences in immune infiltration levels between different risk groups using the CIBERSORT and

ESTIMATE algorithms. The results indicated a close correlation between the low-risk group and immune cell infiltration. Following this, we developed a nomogram by incorporating the risk score and clinical-pathological parameters. The calibration curve and DCA curve both demonstrated the high accuracy of this nomogram in predicting survival rates. Given that PGK1 exhibited the highest HR in the COX regression analysis, we delved deeper into its role. The results revealed that patients with high PGK1 expression had significantly worse prognosis than those with low PGK1 expression. Prior research has indicated that high intracellular expression of PGK1 leads to increased tumor cell proliferation and can enhance the progression and metastasis of breast cancer through the promotion of HIF-1α-mediated EMT [48,49]. Furthermore, PGK1's involvement in various protein post-translational modifications such as acetylation, phosphorylation, ubiquitination, and succinylation plays a crucial role in regulating tumor metabolism and growth [50-53]. Consistent with these findings, our GO/KEGG enrichment analysis identified PGK1's extensive involvement in immune, metabolic, and proliferative signaling pathways. Furthermore, PGK1 has been associated with chemotherapy resistance in cancer patients [54]. Finally, our drug sensitivity analysis revealed that patients with high PGK1 expression exhibited high sensitivity to anti-tumor drugs such as epirubicin and palbociclib. Other studies have also shown that inhibiting PGK1 can increase gastric cancer cell sensitivity to 5-FU and mitomycin [55], and breast cancer patients with high PGK1 expression had shorter overall survival when treated with paclitaxel [56].

Finally, we confirmed through *in vitro* cell experiments that the knockout of the PGK1 gene in human breast cancer HS578T cells significantly inhibits both proliferation and migration of breast cancer cells. This underscores the pivotal role of the PGK1 gene in the development of breast cancer and its potential as a promising therapeutic target for the future. Additionally, assessing PGK1 expression before chemotherapy could predict patients' sensitivity to chemotherapy drugs, and reducing PGK1 expression presents a new strategy to overcome drug resistance. However, there is an urgent need for specific inhibitors targeting PGK1 to target cancer cells and develop therapeutic drugs, which holds significant importance. Despite some remaining questions about lactylation, the progress in related research has opened up an entirely new field in protein post-translational modification. We hope to elucidate the specific roles and regulatory mechanisms of lactylation in diseases in the near future.

5 References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000;100(1):57-70. doi:10.1016/s0092-8674(00)81683-9

2. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013
3. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023.*CA Cancer J Clin*. 2023;73(1):17-48. doi:10.3322/caac.21763
4. WARBURG O. On the origin of cancer cells. *Science*. 1956;123(3191):309-314. doi:10.1126/science.123.3191.309
5. DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab*. 2008;7(1):11-20. doi:10.1016/j.cmet.2007.10.002
6. Jones RG, Thompson CB. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev*. 2009;23(5):537-548. doi:10.1101/gad.1756509
7. Certo M, Tsai CH, Pucino V, Ho PC, Mauro C. Lactate modulation of immune responses in inflammatory versus tumor microenvironments. *Nat Rev Immunol*. 2021;21(3):151-161. doi:10.1038/s41577-020-0406-2
8. Brooks GA. The Science and Translation of Lactate Shuttle Theory. *Cell Metab*. 2018;27(4):757-785. doi:10.1016/j.cmet.2018.03.008
9. Cui J, Quan M, Jiang W, et al. Suppressed expression of LDHB promotes pancreatic cancer progression via inducing glycolytic phenotype. *Med Oncol*. 2015;32(5):143. doi:10.1007/s12032-015-0589-8
10. Kim JH, Kim EL, Lee YK, et al. Decreased lactate dehydrogenase B expression enhances claudin 1-mediated hepatoma cell invasiveness via mitochondrial defects. *Exp Cell Res*. 2011;317(8):1108-1118. doi:10.1016/j.yexcr.2011.02.011
11. Végran F, Boidot R, Michiels C, Sonveaux P, Feron O. Lactate influx through the endothelial cell monocarboxylate transporter MCT1 supports an NF-κB/IL-8 pathway that drives tumor angiogenesis. *Cancer Res*. 2011;71(7):2550-2560. doi:10.1158/0008-5472.CAN-10-2828
12. Polet F, Feron O. Endothelial cell metabolism and tumor angiogenesis: glucose and glutamine as essential fuels and lactate as the driving force. *J Intern Med*. 2013;273(2):156-165. doi:10.1111/joim.12016
13. Ippolito L, Morandi A, Giannoni E, Chiarugi P. Lactate: A Metabolic Driver in the Tumour Landscape. *Trends Biochem Sci*. 2019;44(2):153-166. doi:10.1016/j.tibs.2018.10.011
14. Zhang D, Tang Z, Huang H, et al. Metabolic regulation of gene expression by histone lactylation. *Nature*. 2019;574(7779):575-580. doi:10.1038/s41586-019-1678-1
15. Chu X, Di C, Chang P, et al. Lactylated Histone H3K18 as a Potential Biomarker for the Diagnosis and Predicting the Severity of Septic Shock. *Front Immunol*. 2022;12:786666. Published 2022 Jan 6. doi:10.3389/fimmu.2021.786666
16. Yu J, Chai P, Xie M, et al. Histone lactylation drives oncogenesis by facilitating m^6A reader protein YTHDF2 expression in ocular melanoma. *Genome Biol*. 2021;22(1):85. Published 2021 Mar 16. doi:10.1186/s13059-021-02308-z
17. Pan RY, He L, Zhang J, et al. Positive feedback regulation of microglial glucose metabolism by histone H4 lysine 12 lactylation in Alzheimer's disease. *Cell Metab*. 2022;34(4):634-648.e6. doi:10.1016/j.cmet.2022.02.013
18. Chen YJ, Huang CS, Phan NN, et al. Molecular subtyping of breast cancer intrinsic taxonomy with oligonucleotide microarray and NanoString nCounter. *Biosci Rep*. 2021;41(8):BSR20211428. doi:10.1042/BSR20211428
19. Cheng Z, Huang H, Li M, Liang X, Tan Y, Chen Y. Lactylation-Related Gene Signature Effectively Predicts Prognosis and Treatment Responsiveness in Hepatocellular Carcinoma. *Pharmaceuticals (Basel)*. 2023;16(5):644. Published 2023 Apr 25. doi:10.3390/ph16050644
20. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to

understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg*. 2010;143(3):331-336. doi:10.1016/j.otohns.2010.05.007
21. Meng Z, Ren D, Zhang K, Zhao J, Jin X, Wu H. Using ESTIMATE algorithm to establish an 8-mRNA signature prognosis prediction system and identify immunocyte infiltration-related genes in Pancreatic adenocarcinoma. *Aging (Albany NY)*. 2020;12(6):5048-5070. doi:10.18632/aging.102931
22. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243-259. doi:10.1007/978-1-4939-7493-1_12
23. Huang L, Wu C, Xu D, Cui Y, Tang J. Screening of Important Factors in the Early Sepsis Stage Based on the Evaluation of ssGSEA Algorithm and ceRNA Regulatory Network. *Evol Bioinform Online*. 2021;17:11769343211058463. Published 2021 Nov 26. doi:10.1177/11769343211058463
24. Pei S, Zhang P, Yang L, et al. Exploring the role of sphingolipid-related genes in clinical outcomes of breast cancer. *Front Immunol*. 2023;14:1116839. Published 2023 Feb 13. doi:10.3389/fimmu.2023.1116839
25. Christofides A, Konstantinidou E, Jani C, Boussiotis VA. The role of peroxisome proliferator-activated receptors (PPAR) in immune responses. *Metabolism*. 2021;114:154338. doi:10.1016/j.metabol.2020.154338
26. Mills KHG. IL-17 and IL-17-producing cells in protection versus pathology. *Nat Rev Immunol*. 2023;23(1):38-54. doi:10.1038/s41577-022-00746-9
27. Gordon-Weeks A, Yuzhalin AE. Cancer Extracellular Matrix Proteins Regulate Tumour Immunity. *Cancers (Basel)*. 2020;12(11):3331. Published 2020 Nov 11. doi:10.3390/cancers12113331
28. Pires A, Burnell S, Gallimore A. Exploiting ECM remodelling to promote immune-mediated tumour destruction. *Curr Opin Immunol*. 2022;74:32-38. doi:10.1016/j.coi.2021.09.006
29. Yang K, Fan M, Wang X, et al. Lactate promotes macrophage HMGB1 lactylation, acetylation, and exosomal release in polymicrobial sepsis. *Cell Death Differ*. 2022;29(1):133-146. doi:10.1038/s41418-021-00841-9
30. Moreno-Yruela C, Zhang D, Wei W, et al. Class I histone deacetylases (HDAC1-3) are histone lysine delactylases. *Sci Adv*. 2022;8(3):eabi6696. doi:10.1126/sciadv.abi6696
31. Yang W, Wang P, Cao P, et al. Hypoxic in vitro culture reduces histone lactylation and impairs pre-implantation embryonic development in mice. *Epigenetics Chromatin*. 2021;14(1):57. Published 2021 Dec 21. doi:10.1186/s13072-021-00431-6
32. Li L, Chen K, Wang T, et al. Glis1 facilitates induction of pluripotency via an epigenome-metabolome-epigenome signalling cascade [published correction appears in Nat Metab. 2020 Oct;2(10):1179]. *Nat Metab*. 2020;2(9):882-892. doi:10.1038/s42255-020-0267-9
33. Irizarry-Caro RA, McDaniel MM, Overcast GR, Jain VG, Troutman TD, Pasare C. TLR signaling adapter BCAP regulates inflammatory to reparatory macrophage transition by promoting histone lactylation. *Proc Natl Acad Sci U S A*. 2020;117(48):30628-30638. doi:10.1073/pnas.2009778117
34. Hagihara H, Shoji H, Otabi H, et al. Protein lactylation induced by neural excitation. *Cell Rep*. 2021;37(2):109820. doi:10.1016/j.celrep.2021.109820
35. Jiao Y, Ji F, Hou L, Lv Y, Zhang J. Lactylation-related gene signature for prognostic prediction and immune infiltration analysis in breast cancer. *Heliyon*. 2024;10(3):e24777. Published 2024 Jan 19. doi:10.1016/j.heliyon.2024.e24777
36. Zhu Y, Qi X, Yu C, et al. Identification of prothymosin alpha (PTMA) as a biomarker for esophageal squamous cell carcinoma (ESCC) by label-free quantitative proteomics and

Quantitative Dot Blot (QDB). *Clin Proteomics*. 2019;16:12. Published 2019 Apr 5. doi:10.1186/s12014-019-9232-6
37. Yang L, Sun H, Liu X, et al. Circular RNA hsa_circ_0004277 contributes to malignant phenotype of colorectal cancer by sponging miR-512-5p to upregulate the expression of PTMA [published online ahead of print, 2020 Jan 21]. *J Cell Physiol*. 2020;10.1002/jcp.29484. doi:10.1002/jcp.29484
38. Dominguez F, Magdalena C, Cancio E, et al. Tissue concentrations of prothymosin alpha: a novel proliferation index of primary breast cancer. *Eur J Cancer*. 1993;29A(6):893-897. doi:10.1016/s0959-8049(05)80433-2
39. Tsai YS, Jou YC, Tsai HT, Shiau AL, Wu CL, Tzai TS. Prothymosin-α enhances phosphatase and tensin homolog expression and binds with tripartite motif-containing protein 21 to regulate Kelch-like ECH-associated protein 1/nuclear factor erythroid 2-related factor 2 signaling in human bladder cancer. *Cancer Sci*. 2019;110(4):1208-1219. doi:10.1111/cas.13963
40. Martin TA, Pereira G, Watkins G, Mansel RE, Jiang WG. N-WASP is a putative tumour suppressor in breast cancer cells, in vitro and in vivo, and is associated with clinical outcome in patients with breast cancer. *Clin Exp Metastasis*. 2008;25(2):97-108. doi:10.1007/s10585-007-9120-8
41. Menotti M, Ambrogio C, Cheong TC, et al. Wiskott-Aldrich syndrome protein (WASP) is a tumor suppressor in T cell lymphoma. *Nat Med*. 2019;25(1):130-140. doi:10.1038/s41591-018-0262-9
42. Mughees M, Bano F, Wajid S. Mechanism of WASP and WAVE family proteins in the progression of prostate cancer. *Protoplasma*. 2021;258(4):683-693. doi:10.1007/s00709-021-01608-2
43. Frugtniet B, Jiang WG, Martin TA. Role of the WASP and WAVE family proteins in breast cancer invasion and metastasis. *Breast Cancer (Dove Med Press)*. 2015;7:99-109. Published 2015 Apr 24. doi:10.2147/BCTT.S59006
44. Ai J, Huang H, Lv X, et al. FLNA and PGK1 are two potential markers for progression in hepatocellular carcinoma. *Cell Physiol Biochem*. 2011;27(3-4):207-216. doi:10.1159/000327946
45. Xie H, Tong G, Zhang Y, Liang S, Tang K, Yang Q. PGK1 Drives Hepatocellular Carcinoma Metastasis by Enhancing Metabolic Process. *Int J Mol Sci*. 2017;18(8):1630. Published 2017 Jul 27. doi:10.3390/ijms18081630
46. Ahmad SS, Glatzle J, Bajaeifer K, et al. Phosphoglycerate kinase 1 as a promoter of metastasis in colon cancer. *Int J Oncol*. 2013;43(2):586-590. doi:10.3892/ijo.2013.1971
47. Yu W, Yang Z, Huang R, Min Z, Ye M. SIRT6 promotes the Warburg effect of papillary thyroid cancer cell BCPAP through reactive oxygen species. *Onco Targets Ther*. 2019;12:2861-2868. Published 2019 Apr 15. doi:10.2147/OTT.S194256
48. Lay AJ, Jiang XM, Kisker O, et al. Phosphoglycerate kinase acts in tumour angiogenesis as a disulphide reductase. *Nature*. 2000;408(6814):869-873. doi:10.1038/35048596
49. Fu D, He C, Wei J, et al. PGK1 is a Potential Survival Biomarker and Invasion Promoter by Regulating the HIF-1α-Mediated Epithelial-Mesenchymal Transition Process in Breast Cancer. *Cell Physiol Biochem*. 2018;51(5):2434-2444. doi:10.1159/000495900
50. Hu H, Zhu W, Qin J, et al. Acetylation of PGK1 promotes liver cancer cell proliferation and tumorigenesis. *Hepatology*. 2017;65(2):515-528. doi:10.1002/hep.28887
51. Zhang Y, Yu G, Chu H, et al. Macrophage-Associated PGK1 Phosphorylation Promotes Aerobic Glycolysis and Tumorigenesis. *Mol Cell*. 2018;71(2):201-215.e7. doi:10.1016/j.molcel.2018.06.023
52. Zhang N, Gao R, Yang J, et al. Quantitative Global Proteome and Lysine Succinylome Analyses Reveal the Effects of Energy Metabolism in Renal Cell Carcinoma. *Proteomics*. 2018;18(19):e1800001. doi:10.1002/pmic.201800001

53. Dong W, Li H, Wu X. Rab11-FIP2 suppressed tumor growth via regulation of PGK1 ubiquitination in non-small cell lung cancer. *Biochem Biophys Res Commun*. 2019;508(1):60-65. doi:10.1016/j.bbrc.2018.11.108
54. He Y, Luo Y, Zhang D, et al. PGK1-mediated cancer progression and drug resistance. *Am J Cancer Res*. 2019;9(11):2280-2302. Published 2019 Nov 1.
55. Schneider CC, Archid R, Fischer N, et al. Metabolic alteration--Overcoming therapy resistance in gastric cancer via PGK-1 inhibition in a combined therapy with standard chemotherapeutics. *Int J Surg*. 2015;22:92-98. doi:10.1016/j.ijsu.2015.08.020
56. Sun S, Liang X, Zhang X, et al. Phosphoglycerate kinase-1 is a predictor of poor survival and a novel prognostic biomarker of chemoresistance to paclitaxel treatment in breast cancer. *Br J Cancer*. 2015;112(8):1332-1339. doi:10.1038/bjc.2015.114

Figure.1| Lactylation-related genes expression and mutation in breast cancer (A) Difference analysis of gene expression in breast cancer. **(B)** Differential expression of lactylation-related genes in breast cancer. **(C)** PPI network of lactylation-related genes. **(D)** Frequency of copy number variation of lactylation-related genes. **(E)** Chromosomal distribution circle diagram of lactylation-related genes.

Figure.2 | Lactylation subgroups and their characteristic in breast cancer. (A) Network diagram of lactylation-related genes. **(B)** Forest plot of lactylation-related genes at different risks. **(C)** Unsupervised clustering of lactylation-related genes. **(D)** PCA analysis among different clusters. (**E**) Prognostic analysis among different clusters. (**F**) Relationship between clinicopathological features and expression levels of lactylation-related genes among different clusters. **(G)** GSVA analysis of signaling pathways between different clusters. **(H)** Immune infiltration levels between different clusters.

527

Figure.3 | Construction and evaluation of prognostic models. (A-B) LASSO regression screening of prognostic genes. **(C)** K-M curves of different risk groups. **(D)** ROC curves of different risk groups. **(E)** Expression of modeling genes in different risk groups. **(F)** Relationships between different clusters and risk scores, and survival status. **(G)** Differences in risk scores between different clusters. **(H)** Nomogram for predicting the probability of OS at 1, 3, and 5 years in breast cancer patients. **(I)** Calibration curve for nomogram. **(J-L)** DCA curves for nomogram.

528

529

530

Figure.4| Immune infiltration analysis. (A) Distribution and relative content of immune cells in different risk groups. **(B-G)** Correlation between risk scores and immune cell types. **(H)** Level of immune infiltration in different risk groups. **(I)** Correlation between modeling genes and immune cell abundance. **(J)** Correlation between risk scores and StromalScore, ImmuneScore.

531

Figure.5 |Prognostic analysis, biological function, and drug sensitivity analysis of PGK1. (A) Survival curves of PGK1. **(B-C)** Differences in OS and PFS among patients with varying levels of PGK1 expression in the GSE124647 dataset. **(D-E)** GO/KEGG analysis of biological functions and signaling pathways of differentially expressed genes between the high PGK1 group and low PGK1 group. **(F-K)** Comparison of the IC50 values of common antitumor drugs between the high PGK1 group and low PGK1 group, including epirubicin, palbociclib, ribociclib, sorafenib, cytarabine, gemcitabine.

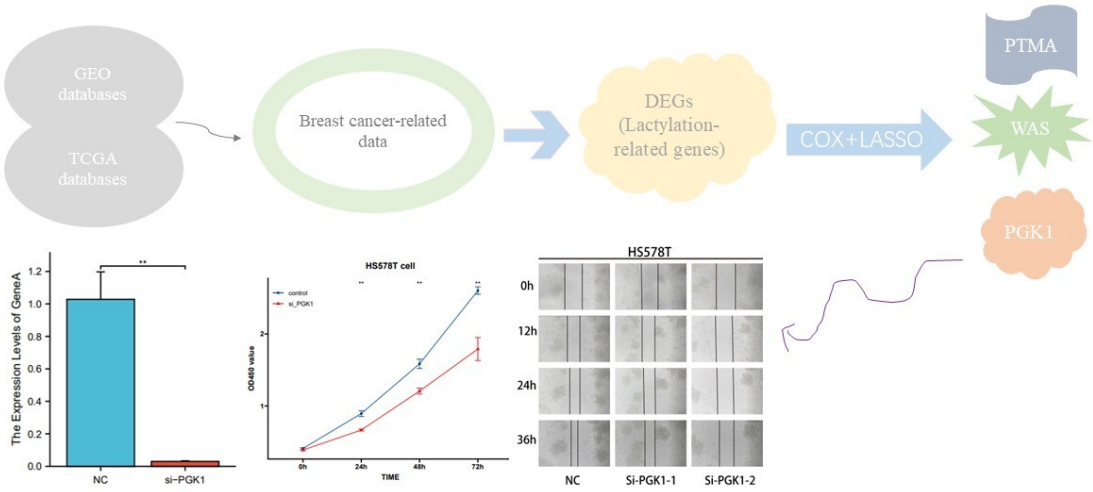
532

533

534

535

Figure.6| Cell experiment. (A) qRT-PCR assessment of PGK1 mRNA levels after 24 hours of transfection. siRNA sequence could lead to a significant decrease in PGK1 mRNA expression (P<0.01). **(B)** CCK8 assay. The viability of the cells was significantly reduced after PGK1 Knockdown. **(C)** Scratch wound healing assay**.** A significantly slower scratch closure in the PGK1 knockdown group compared to the siRNA NC group. All data are expressed as the means ± SD of the three experimental groups. *P <0.05, **P <0.01, ***P <0.001 were considered statistically significant.


GEO databases
TCGA databases
Breast cancer-related data
DEGs (Lactylation-related genes)
COX+LASSO
PTMA
WAS
PGK1
The Expression Levels of GeneA
NC
si-PGK1
HS578T cell
control
si_PGK1
OD450 value
TIME
0h
24h
48h
72h
HS578T
0h
12h
24h
36h
NC
Si-PGK1-1
Si-PGK1-2

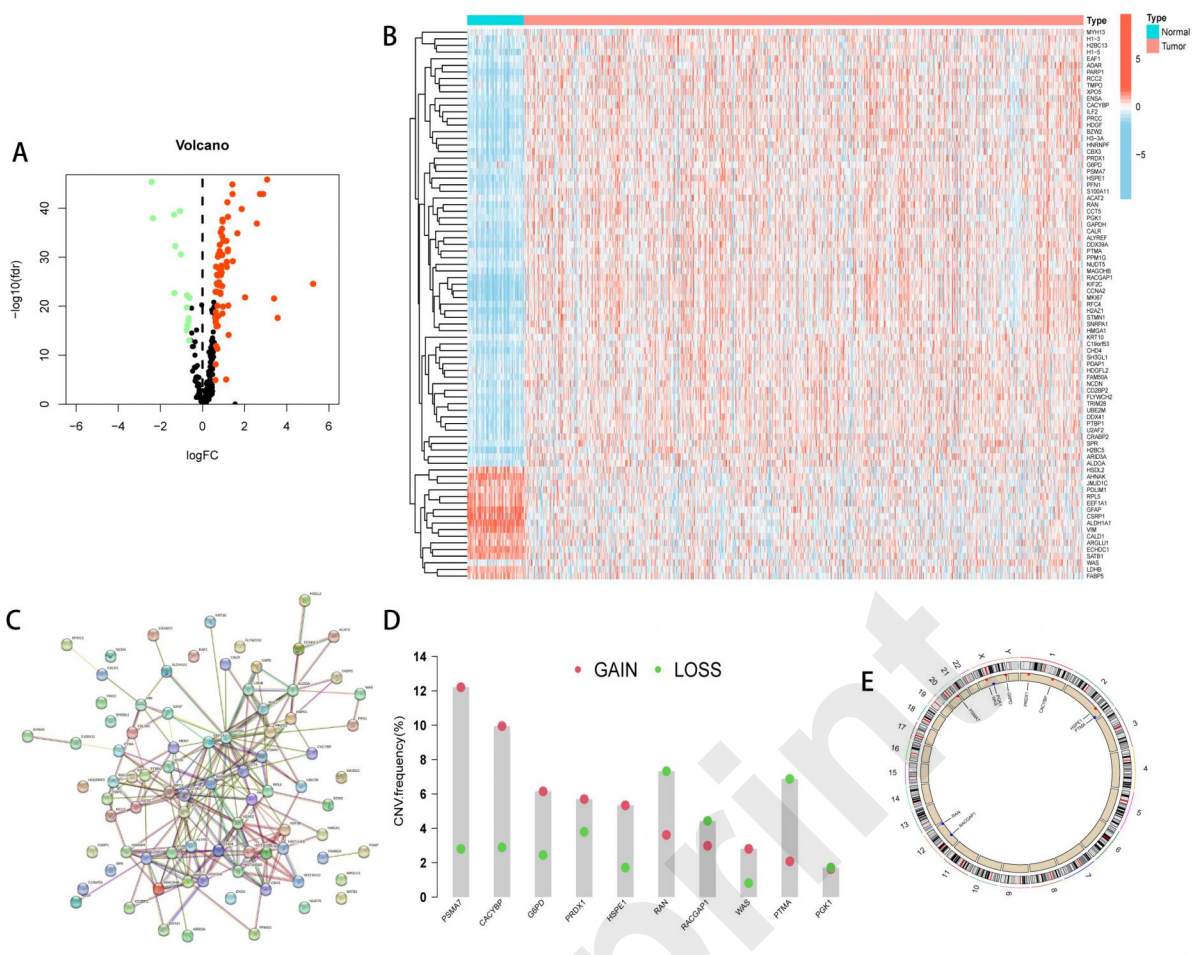


Figure.1| Lactylation-related genes expression and mutation in breast cancer.

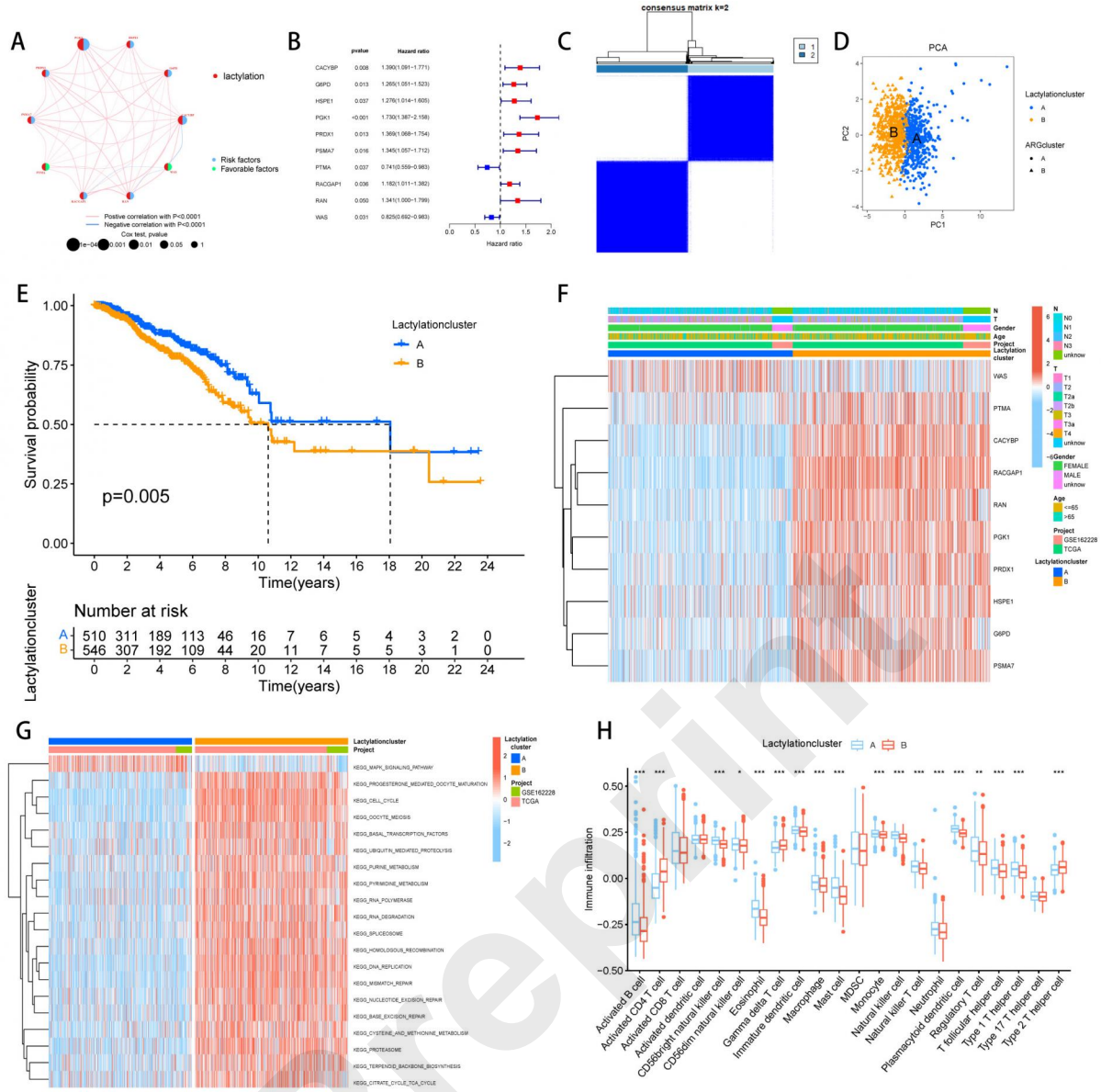


Figure.2 | Lactylation subgroups and their characteristic in breast cancer.

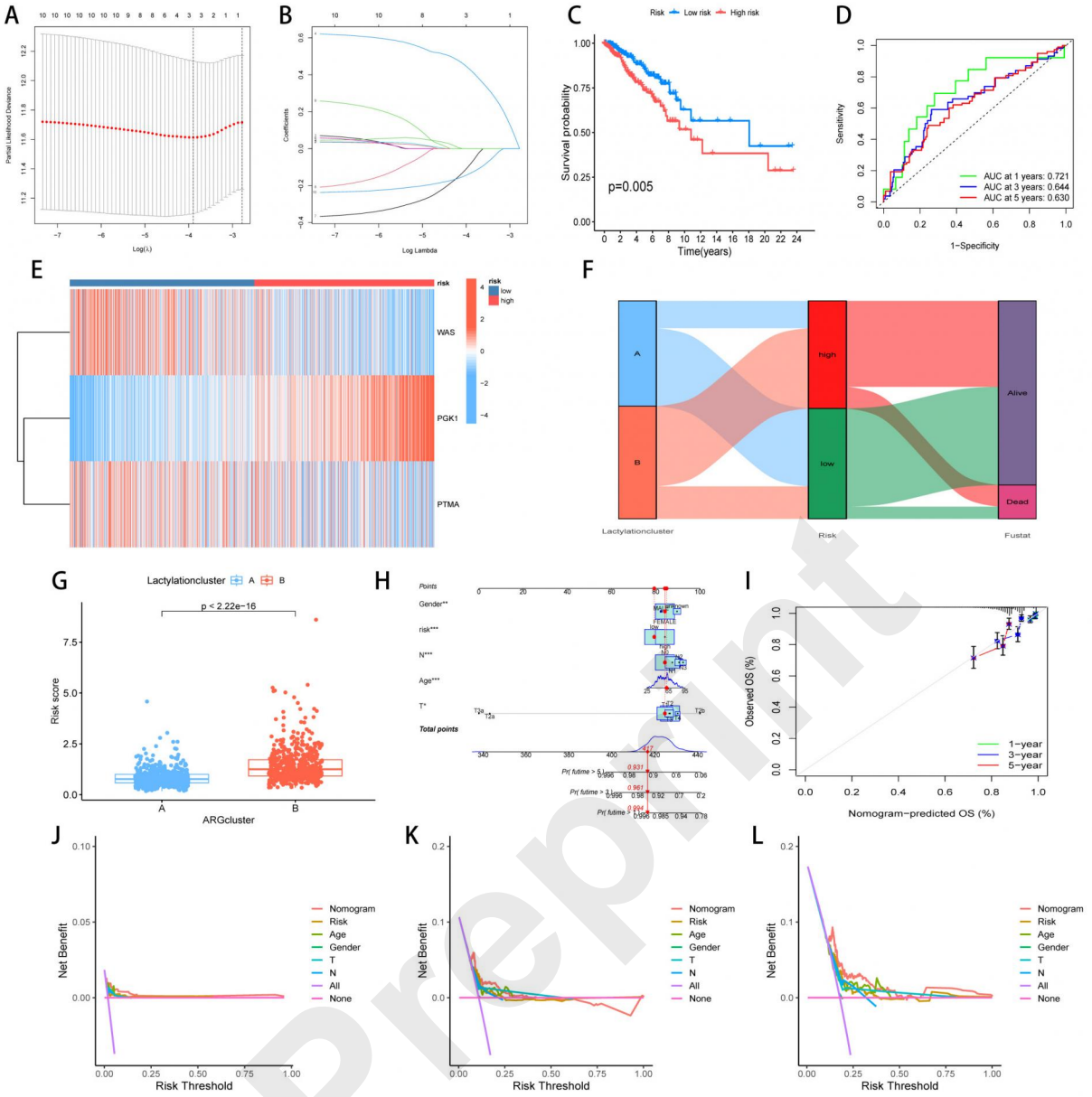


Figure.3 | Construction and evaluation of prognostic models.

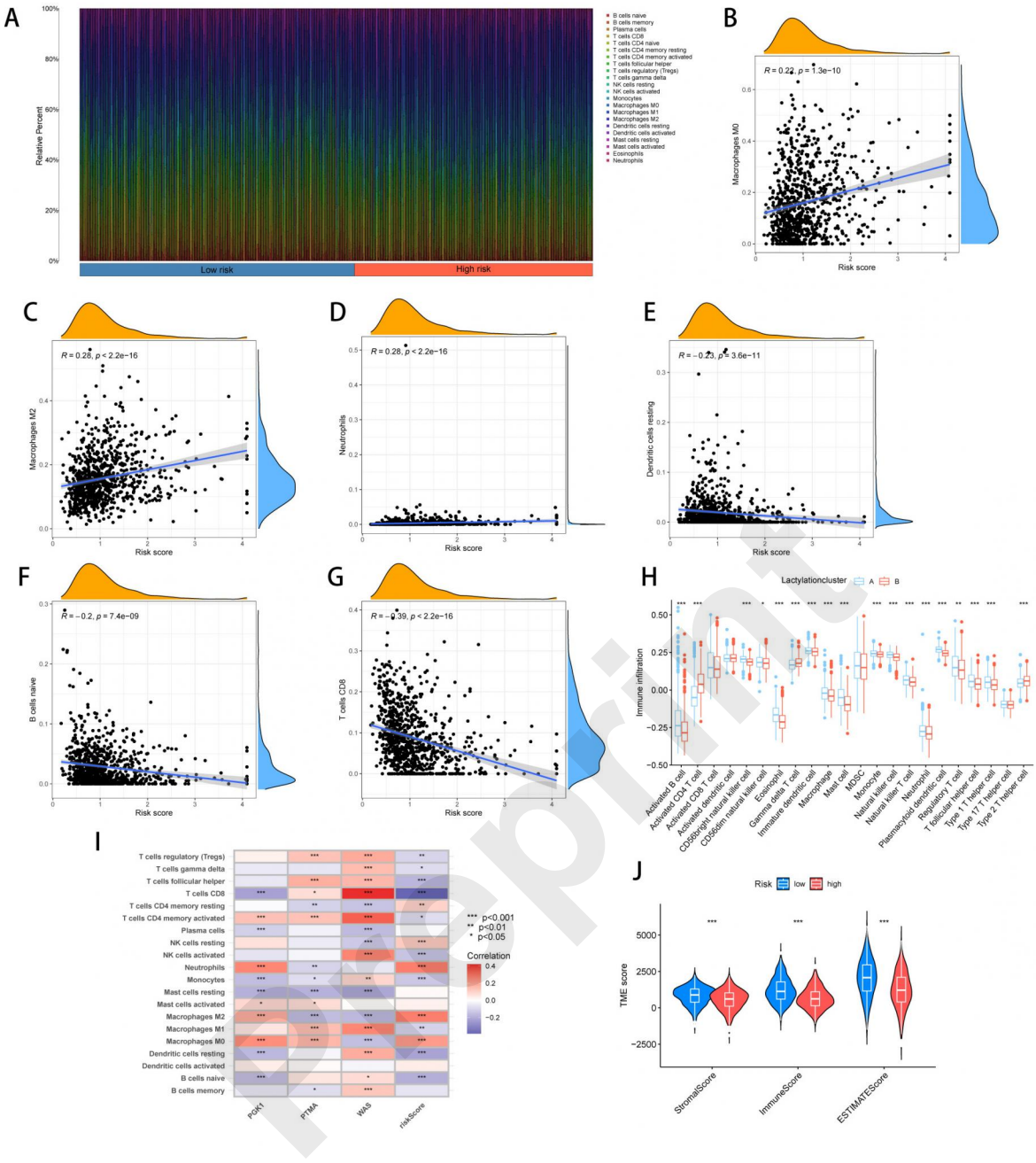


Figure.4| Immune infiltration analysis.

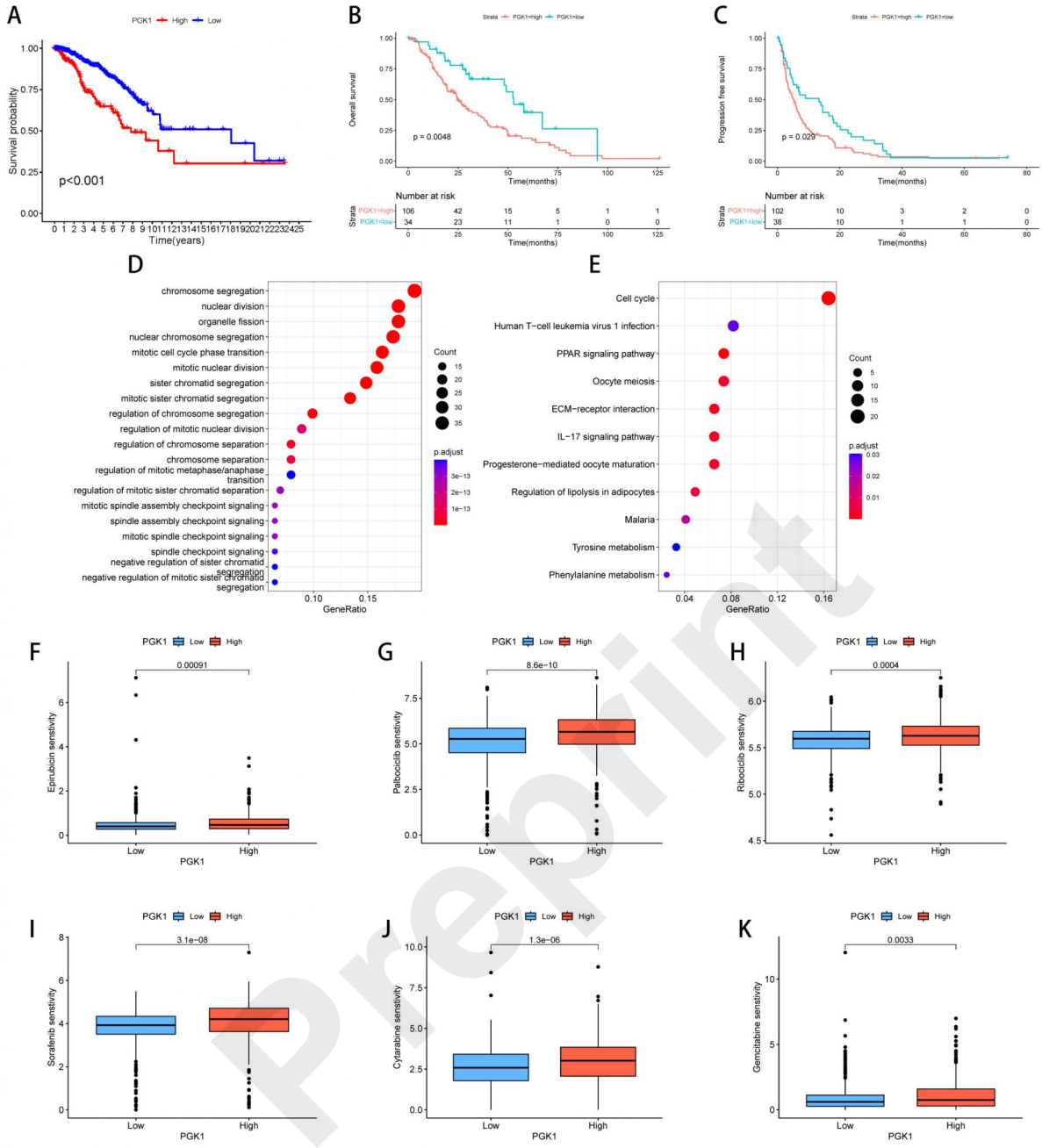


Figure.5 |Prognostic analysis, biological function, and drug sensitivity analysis of PGK1.

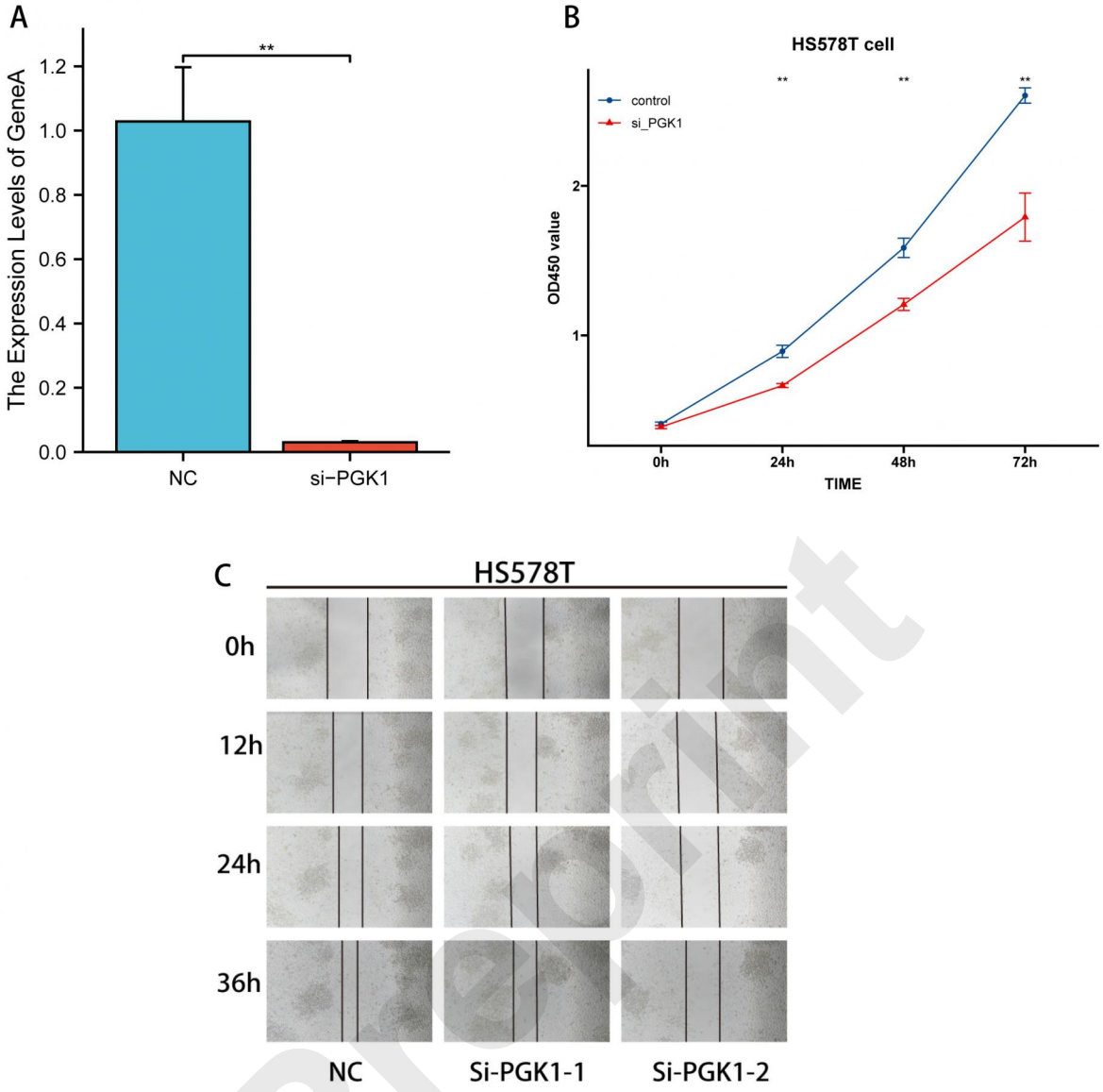


Figure.6| Cell experiment.