Discover Oncology

RESEARCH Open Access

A computational framework integrating multi-omics and machine learning for identifying glycolytic gene markers in breast cancer

Yuxing Liu[1,2†], Chenming Liu[3†], Chunhui Tang[4] and Feng Wang[1*]

†Yuxing Liu and Chenming Liu contributed equally to this work.

*Correspondence:
Feng Wang
wangfeng71012@163.com
[1]Department of General Surgery, Binhai Clinical College, Binhai County People's Hospital, Yangzhou University Medical College, Yancheng 224500, Jiangsu, China
[2]Department of General Surgery, Mingji Hospital, No. 71, Hexi Street, Jianye District, Nanjing 210000, Jiangsu, China
[3]Department of General Surgery, Beijing Friendship Hospital, Capital Medical University, Beijing, China
[4]Department of Infectious Diseases, Jiangbei People's Hospital, No. 552, Geguan Road, Liuhe District, Nanjing 210000, Jiangsu, China

Abstract

Breast cancer remains a leading cause of cancer-related mortality in women worldwide. Metabolic reprogramming, especially enhanced glycolysis, is a hallmark of breast cancer progression and offers promising targets for therapeutic intervention. However, systematic identification of glycolysis-related gene markers at single-cell resolution remains limited.We integrated single-cell RNA sequencing, spatial transcriptomics, and multiple machine learning algorithms to analyze glycolytic heterogeneity in breast cancer. A novel interquartile range (IQR)-based scoring method was developed to quantify glycolytic activity at the single-cell level. Glycolysis-related genes (GRGs) were systematically screened using five machine learning models (RF, Boruta, Lasso, ABESS, and GBM), and their diagnostic and prognostic values were evaluated.We identified significant glycolytic heterogeneity among breast cancer cell subsets, with malignant cells exhibiting the highest glycolytic activity. Seven hub genes, namely PKM, MIF, SOD1, PGK1, APP, LDHB and NUPR1, were consistently identified and validated. These genes showed strong diagnostic potential with high AUC values in ROC analysis, and their elevated expression was confirmed in breast cancer tissues via immunohistochemistry. Spatial and cell communication analyses further revealed distinct metabolic niches and intercellular signaling networks associated with high-glycolytic cells.This study establishes a robust IQR-based glycolysis assessment strategy that overcomes limitations of previous scoring methods. We identified seven core glycolytic regulatory genes that are closely linked to breast cancer progression and poor prognosis. These findings provide novel insights into metabolic reprogramming in breast cancer and offer potential biomarkers for targeted metabolic therapy.

Keywords Breast cancer, Glycolysis, Single-cell RNA sequencing, Spatial transcriptome, Machine learning, CellChat

1 Introduction

Breast cancer is the most common malignant tumor among women worldwide, with both incidence and mortality rates continuing to rise [1]. Current treatment methods, including surgery, chemotherapy, radiotherapy, endocrine therapy and targeted therapy,

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

have all made significant progress. However, limitations such as drug resistance and bone marrow suppression restrict the effectiveness of existing therapies. Moreover, these treatments show minimal efficacy against triple-negative breast cancer (TNBC). A deeper understanding of breast cancer pathogenesis can improve patient outcomes. In recent years, research on abnormal cancer cell metabolism has gained significant attention in the development of cancer treatments. Most normal tissues obtain energy through aerobic glycolysis under aerobic conditions, and only switch to anaerobic glycolysis under hypoxic conditions. However, tumor tissues primarily rely on anaerobic glycolysis for energy production, even when oxygen supply is sufficient. This phenomenon is known as the "Warburg effect" [2]. Similar to most cancer cells, breast cancer cells exhibit significant differences in glucose metabolism, characterized by markedly enhanced glycolysis [3]. Breast cancer cells depend more on ATP produced through glycolysis, providing a biochemical basis for developing drugs that inhibit glycolysis to selectively target these cells. This also makes targeting metabolic enzymes in the glycolytic pathway a viable strategy for anti-cancer drug development.

The glycolytic pathway refers to the process where cells in the cytoplasm break down glucose into pyruvate through a series of enzymatic reactions [4]. Glucose in the cytoplasm is first phosphorylated by hexokinase (HK) into glucose-6-phosphate (G-6-P). G-6-P is then converted into pyruvate by enzymes such as phosphofructokinase (PFK) and pyruvate kinase (PK). The pyruvate produced by the glycolytic pathway can either enter the tricarboxylic acid cycle directly or be converted into lactic acid by lactate dehydrogenase (LDH). In the process of glucose being converted into lactic acid through glycolysis, key metabolic enzymes like HK, PFK, PK, and LDH are associated with tumors and regulated by transcription factors in carcinogenic signaling pathways, such as hypoxia-inducible factor-1α (HIF-1α) [5].

The glycolytic pathway plays a pivotal role in the metabolic reprogramming of breast cancer, with its biological functions in promoting tumor proliferation and invasion being well-established. However, systematic analysis of glycolysis-associated molecular markers at single-cell resolution remains insufficient. In recent years, single-cell and spatial transcriptomics have revealed significant metabolic heterogeneity in breast cancer and its association with poor clinical outcomes [6–9]. However, existing studies still face several key limitations. First, there is insufficiently precise spatial mapping of high-glycolytic malignant cell subpopulations within tumor tissues. Second, there is a lack of a systematic multi-algorithm machine learning framework to robustly identify core regulatory genes from high-dimensional data. Finally, the exploration of intercellular communication networks of highly metabolic cells and their role in shaping the microenvironment remains limited. This study aims to address these gaps by developing a novel interquartile range (IQR)-based single-cell metabolic scoring method, integrating various machine learning algorithms for feature gene selection, and combining spatial transcriptomics with cell communication analysis, with the goal of providing new methodological tools and biological insights.

2 Methods

2.1 Collection of data sets

All single-cell transcriptome data analyzed in this study were retrieved from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/). Specifically, GSE176078

included samples from 20 breast cancer patients, while GSE161529 contained total breast cell samples from three healthy controls. For bulk RNA data analysis, we obtained 1,095 breast cancer samples from the Cancer Genome Atlas (TCGA-BRCA) database. Additionally, six spatial transcriptomic datasets related to breast cancer were acquired from Zenodo [Zenodo (https://zenodo.org/records/4739739)].

2.2 Access to glycolytic pathway genes

To comprehensively identify glycolysis-related genes(GRGs), we systematically searched multiple authoritative databases including KEGG, Reactome, GO, MSigDB Hallmark, bioactome, and MetaPathway, using search terms such as 'glycolysis' and 'glucose metabolism' [10–12]. After integrating findings from these datasets and removing redundancies, we identified a glycolysis-related gene set comprising 341 genes (Supplementary Table 1).

2.3 Preprocessing of data sets

The single-cell RNA sequencing data in this study were generated using Illumina's NextSeq 500 platform on the 10x Genomics platform. For the GSE176078 and GSE161529 datasets, we established stringent quality control criteria: each gene must be expressed in at least five cells, each cell must contain at least 200 detectable genes, and mitochondrial gene expression must account for less than 20% of total genes. Subsequent downstream analyses were performed using the Seurat software package, which included expression normalization based on highly variable genes, principal component analysis (PCA) for dimensionality reduction, and Harmony algorithm correction for batch effects [13].The assessment of glycolytic activity was performed on the batch-corrected, integrated dataset to ensure that the metabolic scores reflect biological differences rather than technical artifacts. We ultimately selected the top 20 principal components based on statistical significance ($P < 0.05$), performed nonlinear dimensionality reduction visualization using the UMAP algorithm, and automated cell subpopulation annotation with SingleR software [14].

2.4 Assessment of the active state of the glycolytic pathway

To achieve precise quantification of glycolytic activity in malignant cells, this study overcame the limitations of single algorithms by integrating five complementary single-cell computational methods (AUCell, UCell, singscore, AddModuleScore, and ssgsea) [15–19]. AUCell and UCell provide rank-based scoring resistant to gene dropout; singscore enables directional assessment of gene sets; AddModuleScore calculates the average expression of a gene set relative to background features; and ssgsea projects pathway analysis logic from bulk to single-cell data. Using this integrated approach, we systematically evaluated glycolytic activity in 51,099 tumor cell samples, successfully assigning glycolytic activity scores to each cell.

2.5 Spatial transcriptome analysis

The dataset for this study was sourced from spatial transcriptome data downloaded from the Zenodo database. Following a sequential quality control process, genes detected in fewer than five spots were removed, and spots with fewer than 300 detected features or exhibiting a mitochondrial gene proportion exceeding 20% were filtered out.

Subsequently, data standardization was achieved using the SCTransform method with default parameters in Seurat. Finally, we conducted unsupervised clustering analysis based on PCA across multiple spatial sections through Seurat's RunPCA, FindNeighbors, FindClusters, and RunUMAP functions.

2.6 Quartile screening of high metabolic cell subpopulations based on glycolytic activity

Cell classification is first performed by calculating the percentile of scoring data. Subsequently, we established two key cutoff points based on the distribution: cells scoring above the third quartile (Q3) are designated as HGS, those below the first quartile (Q1) as LGS, and those with scores between Q1 and Q3 are classified as DTGS.To validate that our key findings were not contingent upon this specific threshold, a sensitivity analysis was performed using an alternative stratification strategy.

2.7 High-dimensional weighted gene co-expression network analysis(hdWGCNA)

To analyze the functional associations among genes, we first initiated the analysis process using the hdWGCNA algorithm in R [20]. The process began with preprocessing the expression matrix by removing low-expressing genes and performing rigorous quality control. Subsequently, the optimal soft threshold for constructing the weighted network was determined through the TestSoftPowers function, which enabled generation of the co-expression matrix. Hierarchical clustering was then applied to partition genes into modules, followed by module signature gene analysis to identify eight modules significantly associated with cell clusters for further investigation. Finally, connectivity-based screening identified the top 100 genes, which were designated as core genes.

2.8 Automated selection of glycolysis-related genes through machine learning

Given the limitations of single-machine learning algorithms in genetic screening, this study employs a diversified collaborative analysis strategy. By conducting multi-angle evaluations and optimizing feature selection, we aim to reliably identify disease-associated genes. To ensure reproducibility and fairness, all operations utilize fixed random seeds and employ stratified sampling to partition raw data into balanced training and validation sets.We employed a composite framework of five distinct algorithms to capture complementary aspects of gene relevance and enhance biological robustness. random forest (RF) and Boruta evaluate non-linear interactions, Lasso regression enforces sparsity to mirror parsimonious regulatory networks, ABESS performs optimal subset search, and gradient boosting machine (GBM) captures subtle consistent signals. The consensus across these methodologically independent algorithms identifies genes that are stable and core to the glycolytic program, reducing model-specific artifacts and increasing confidence in their biological significance. First, for broad feature importance evaluation and initial screening, we applied RF, which utilized an ensemble of 500 decision trees with the default number of variables considered at each split. The Boruta algorithm, which internally uses Random Forest, was run for 500 iterations to identify relevant features robustly by comparing them against shuffled shadow features. Subsequently, to enhance the quality and sparsity of the screened features, LASSO regression was introduced to perform high-dimensional regularization. Its penalty strength (λ) was optimized via 10-fold cross-validation, yielding a final λ of 0.000696. In parallel, the ABESS algorithm was used for optimal subset search with its default splicing

settings, where the optimal model size was likewise determined by 10-fold cross-validation. Finally, leveraging a boosting mechanism, GBM was configured with a learning rate of 0.1, a maximum tree depth of 3, and an initial 100 trees. This integrated multiple weak classifiers into a strong predictor to precisely identify hub genes most relevant to the target phenotype from the high-dimensional genetic data. This systematic workflow ensures the robustness and scientific validity of our conclusions.

2.9 Development and validation of machine learning-based benchmark models for glycolysis-related genes

This study employed a multi-model benchmarking approach to establish a computational framework capable of reliably inferring GRGs from single-cell RNA sequencing data. Using R's "mlr3" package, we systematically compared seven models: K-nearest neighbor (KNN), linear discriminant analysis (LDA), logistic regression (LR), naive Bayes (NB), random forest (Ranger), recursive partitioning and regression trees (RPART), and extreme gradient boosting (XGBoost) [21]. In data preprocessing, 80% of samples were randomly allocated for training, with the remaining 20% serving as the test set. We implemented 5-fold cross-validation for hyperparameter optimization and supplemented it with 10-fold cross-validation to rigorously evaluate model generalization performance. Ultimately, based on the average AUC metric, we identified the optimal algorithm for constructing GRGs labels.

2.10 Mapping signaling exchange and lineage progression in cellular ecosystems

To decode intercellular signaling communication, we first employed the CellChat algorithm, which constructs intercellular signaling networks based on known signal molecule-receptor pairing information [22]. By systematically evaluating the expression intensity of relevant molecules in various cell subpopulations, we identified biologically significant signaling pathways and explored their potential connections with clinical manifestations in patients. Building on this foundation, we applied the Monocle algorithm to reconstruct cellular transcription dynamics in temporal sequence [23]. This computational model, through low-dimensional space projection and cell sorting, revealed the dynamic continuity of cells during pathological progression, thereby inferring the relative positions of individual cells within predefined developmental trajectories.

2.11 Statistical analysis and data visualization

All computational analyses in this study were performed in R (version 4.4.1). For single-cell and bulk transcriptomic data, we employed two-tailed hypothesis testing for statistical inference. Specifically, covariability between variables was quantified using Pearson's correlation coefficients, while inter-group comparisons utilized Wilcoxon rank-sum tests. We evaluated three Cox proportional hazards models: a clinical model (age, pT, pN, stage), a glycolytic gene model, and a combined model. Model discrimination was assessed using the concordance index (C-index), and improvement in model fit upon adding the glycolytic signature was tested via likelihood-ratio tests. Independent prognostic associations were examined by multivariable Cox regression, with gene expression standardized (z-score). Results are presented as hazard ratios (HR) with 95%

confidence intervals (CI).We determined statistical significance based on the criterion that the p-value (or adjusted p-value) is less than 0.05.

2.12 Immunohistochemistry (IHC)

The Human Protein Atlas (HPA, https://www.proteinatlas.org) is a Swedish program launched in 2003. It integrates various omics technologies to create a map of all human proteins in cells, tissues, organs, and blood [24]. We obtained immunohistochemistry images of PKM, MIF, SOD1, PGK1, APP and LDHB from the HPA to evaluate the expression of these proteins in tumor and normal tissues of patients.

3 Results

3.1 Construction of cell atlas based on single-cell RNA sequencing

This study utilized single-cell RNA sequencing data comprising 20 breast cancer samples and 3 normal controls. Following predefined quality control criteria, 58,908 high-quality cells were selected for subsequent analysis (Supplementary Fig. 1A). To eliminate technical batch effects, we integrated the data using the Harmony algorithm. Quantitative assessment showed an 11.01% reduction in the batch separation ratio and effective local batch mixing, with an average of 10 distinct batches among each cell's 30 nearest neighbors. PCA visualization confirmed successful integration(Supplementary Fig. 1B). UMAP-based nonlinear dimensionality reduction and clustering analysis identified 16 cell subpopulations (Supplementary Fig. 1C). Subsequent SingleR algorithm analysis further classified these subpopulations into seven distinct cell types: CD8 + T cells, epithelial cells, CD4 + T cells, fibroblasts, monocytes, endothelial cells, and B cells (Supplementary Fig. 1D, Fig. 1A). Comparative analysis of normal and tumor tissues revealed significant enrichment of specific immune cell populations in the tumor microenvironment, including CD8 + T cells, CD4 + T cells, and B cells. In contrast, normal tissues exhibit a higher proportion of epithelial cells, fibroblasts, and endothelial cells (Fig. 1B). The annotation of cell types was validated through the expression patterns of marker genes (Supplementary Fig. 1E, Fig. 1C, D). Furthermore, the UMAP spatial distribution of gene expression density visually demonstrates the expression characteristics of key marker genes across different cell populations (Fig. 1E).

3.2 Quantitative analysis of glycolytic activity

To clearly distinguish cell populations with high and low copy number variations (CNVs), this study first employed K-means-based unsupervised clustering analysis, successfully identifying three cell subpopulations (Fig. 2A). Cluster 2, which showed the lowest CNV scores, was defined as normal epithelial cells, while the remaining clusters corresponded to malignant cells (Fig. 2B). Subsequently, we evaluated glycolytic activity using five algorithms: AUCell, UCell, singscore, AddModuleScore, and ssgsea, revealing distinct distribution patterns across different cell types and phenotypes (Fig. 2C). Consistent analysis results demonstrated significantly higher glycolytic activity in tumor tissues compared to normal tissues ($P < 0.05$, Fig. 2D). This trend was corroborated by heatmaps, ridge plots, and UMAP visualizations, showing that malignant cells exhibited markedly higher glycolytic activity and were spatially concentrated (Fig. 2E–H). Spatial mapping based on spot-level resolution indicated an enrichment of transcriptomic signatures associated with malignant cells and high glycolytic activity within the tumor

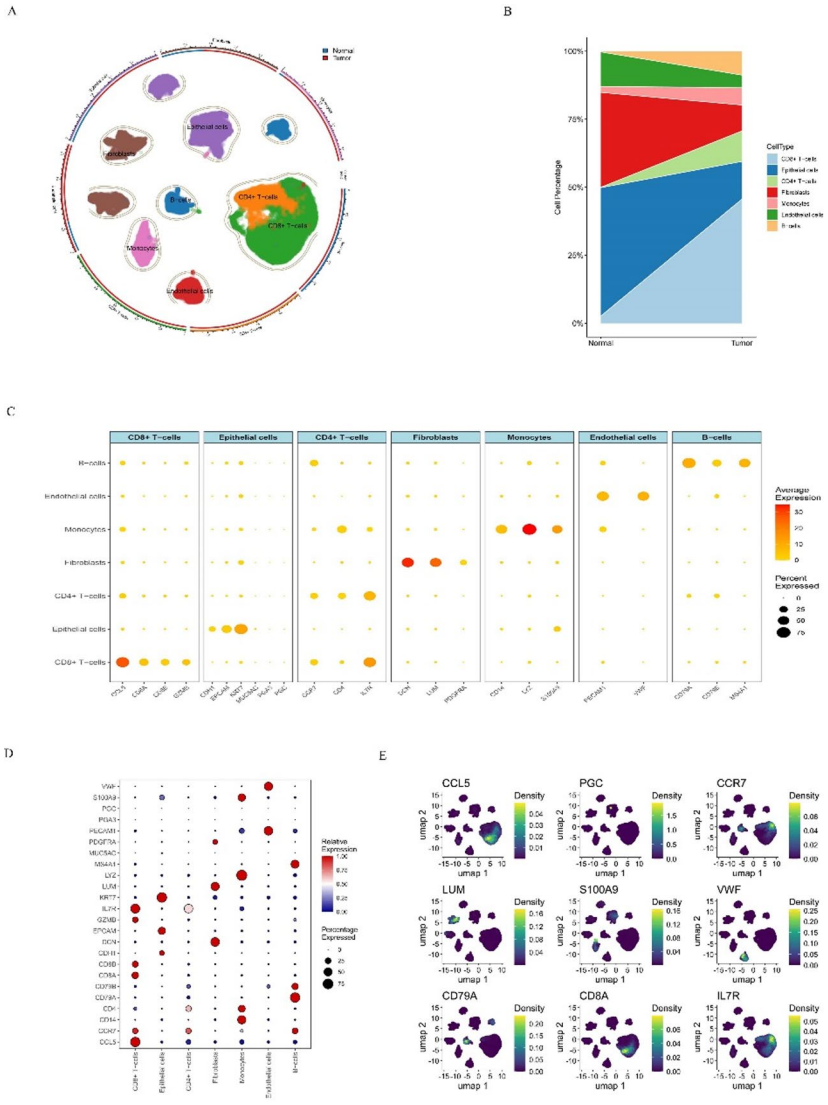

Fig. 1 Analysis of breast cancer tissues using single-cell RNA sequencing. **A** Identification and visualization of heterogeneous cell populations. **B**. Compositional differences in cell types between tumor and normal tissues. **C**, **D** Manual classification of seven unique cell populations based on established marker genes. **E** Expression patterns of canonical marker genes projected onto the UMAP

core region (Fig. 2I–J, Supplementary Fig. 2A–J). Finally, using a robust cell type decomposition algorithm, we successfully mapped the cell types defined in single-cell data to spatial datasets, thereby inferring the composition of malignant cells at key spatial locations (Fig. 2K, Supplementary Fig. 2K–O).

3.3 Phenotypic atlas analysis and subgroup identification of cellular glycolysis

The research findings revealed that cell scores followed a normal distribution across three classification groups: LGS, DTGS, and HGS (Fig. 3A). Enhanced cellular metabolic states were consistent with increased density in the CytoTRACE score, which evaluates cellular differentiation potential (Fig. 3B). Notably, the HGS group exhibited significantly higher scores, suggesting associations with glycolysis or differentiation status (Fig. 3C). UMAP maps clearly demonstrated cell clustering based on glycolysis-related genes (Fig. 3D). Box plots confirmed significantly stronger metabolic activity in the HGS group

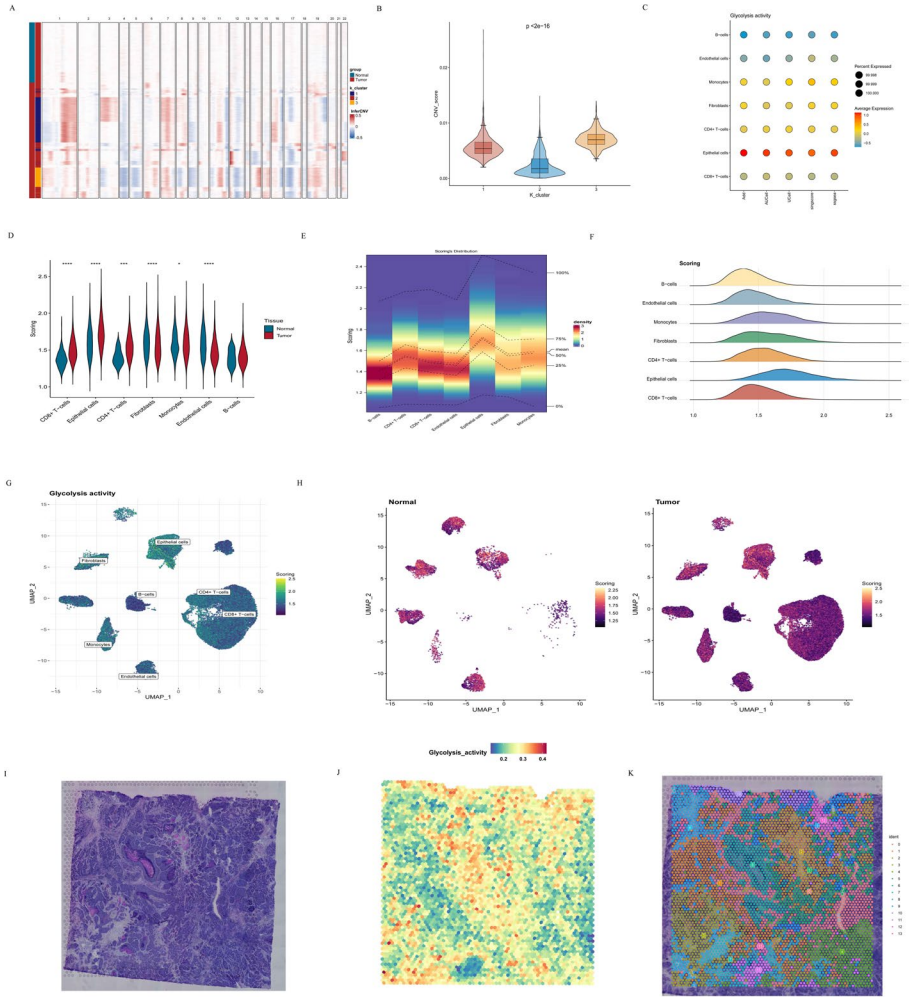

Fig. 2 Assessment of glycolytic activity. **A** The K-means algorithm was employed to analyze inferred copy number variations, enabling the classification of malignant cell populations. **B** A violin plot was used to display the distribution of CNV scores among the clusters identified by K-means grouping. **C** Glycolytic activity across various cell subsets was assessed using five computational tools: AUCell, Ucell, singscore, AddModuleScore, and ssgsea. **D** Glycolysis levels were compared between diverse cell types from normal tissues and tumor specimens. **E**–**G** The glycolytic activity scores for seven major cell types were visualized. **H** Spatial distribution of glycolytic activity in normal and tumor samples was depicted in a two-dimensional UMAP space. **I**. Histological regions within spatial transcriptomics samples were distinguished by hematoxylin and eosin (HE) staining. **J** A spatial map illustrates the organizational pattern of glycolysis activity across tissue structures. K. Tissue architecture in breast cancer was reconstructed through unsupervised pattern recognition methods

(Fig. 3E). The positive correlation between the two was further validated by a Pearson correlation coefficient of 0.54 ($p < 0.001$) (Fig. 3F). Trajectory analysis revealed intratumoral transcriptional heterogeneity in malignant cells (Fig. 3G–I). In the high glycolysis group, pathways such as glycolysis and gluconeogenesis showed significant enrichment and upregulation, collectively indicating enhanced metabolic activity (Fig. 3J).

3.4 Screening for HGS-related genes using hdWGCNA

To conduct in-depth analysis of GRGs, this study employed hdWGCNA for systematic evaluation. By balancing model fit ($R^2 > 0.9$) with network connectivity, we determined a soft power threshold of 6, a decision supported by the hierarchical clustering dendrogram (Fig. 4A). Using this threshold, co-expression analysis successfully clustered genes

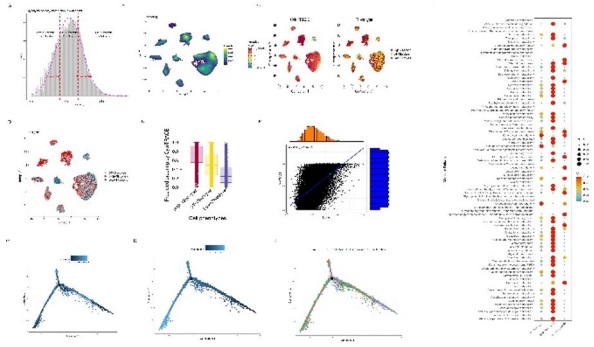

Fig. 3 CytoTRACE reveals glycolytic heterogeneity and cell states. **A** Normality assessment of glycolytic activity distribution. **B**–**D** UMAP projection of CytoTRACE scores. **E** Inter-group comparison of CytoTRACE scores. **F**. Scatter plot for correlation verification. **G** Pseudotime reconstruction of malignant cell states. **H** Metabolic pathway enrichment across glycolytic phenotypes

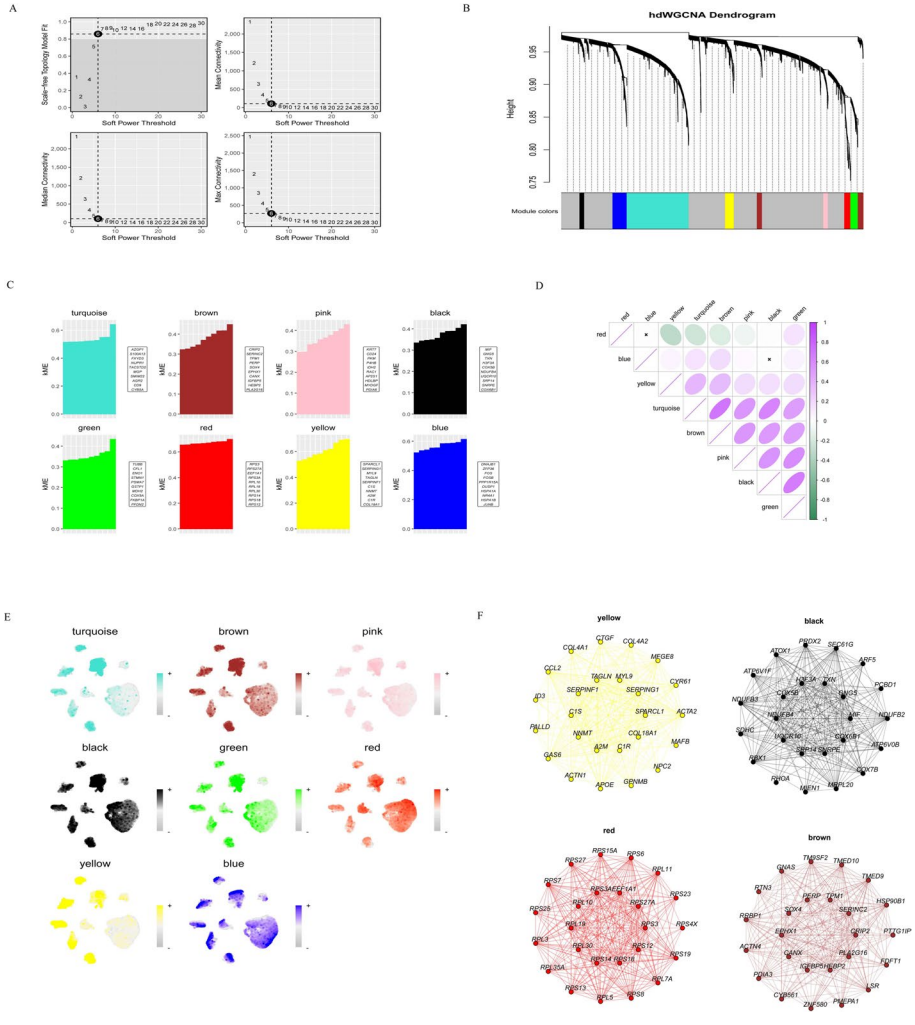

Fig. 4 Systematic identification of core GRGs via hdWGCNA. **A** Determination of the optimal soft power threshold. **B** Identification of co-expression modules, resulting in eight distinct gene clusters. **C** Prioritization of core genes within each module based on kME values. **D** Correlation heatmap depicting associations among the identified gene modules. **E** Feather plot visualization of module scores in malignant cells. **F** Network architecture of four representative hub modules

into eight core modules based on expression profile similarity (Fig. 4B). The study further identified the top 10 most interconnected genes within each module and visualized inter-module relationships (Fig. 4C, D). UMAP analysis further confirmed the functional distinctiveness of these modules (Fig. 4E). Ultimately, 720 cell-associated core genes were identified in the HGS group (Fig. 4F, Supplementary Table 2).

3.5 Screening and functional annotation of glycolytic hub genes

This study first identified 808 genes with significantly different expression levels (Fig. 5A, Supplementary Table 3) by comparing HGS and LGS groups. Cross-annotation of these differentially expressed genes (DEGs) with glucose regulation-related genes revealed 293 overlapping genes (Fig. 5B, Supplementary Table 4). Subsequent Pearson correlation analysis identified 228 genes showing significant positive correlation with glycolytic activity ($Cor > 0$, adjusted P-value < 0.05, Fig. 5C, Supplementary Table 5). Functional annotation further clarified their biological significance: Disease Ontology (DO) analysis identified associations with critical diseases such as breast cancer and melanoma (Fig. 5D, Supplementary Table 6), while the Gene Ontology (GO) analysis revealed significant enrichment in specific functions including regulation of apoptotic signaling pathways and cellular respiration (Fig. 5E, Supplementary Table 7).

3.6 Hub gene screening based on machine learning

This study developed an integrated computational framework to precisely identify hub genes through comparative analysis of HGS and LGS. The framework runs five machine learning algorithms in parallel, with differentiated screening criteria based on their characteristics: models providing feature importance scores retain only the top twenty genes, while results from other models are fully incorporated into the candidate pool. Specifically, the random forest and Lasso algorithms identified 20 and 197 candidate genes respectively (Fig. 6A–C). Subsequently, Boruta and GBM algorithms each identified 20 genes (Fig. 6D–F). The ABESS algorithm independently screened 20 genes (Fig. 6G). Through intersection analysis of these results, we successfully identified seven shared core characteristic genes: PKM, MIF, SOD1, PGK1, APP, LDHB, and NUPR1 (Fig. 6H).A gene was defined as a high-confidence hub gene only if it was selected by all five algorithms. All seven identified genes met this stringent criterion.The feature importance scores or selection coefficients for these genes across each individual algorithm

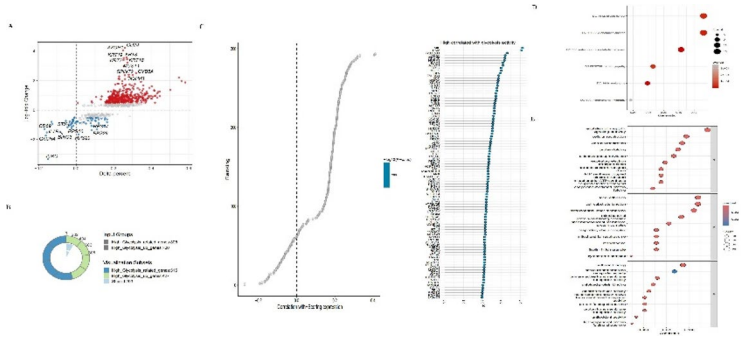

Fig. 5 Hub genes linked to glycolytic activity were pinpointed. **A** Differential gene expression analysis pertaining to glycolysis. **B** Venn diagram illustrating the overlap between GRGs and DEGs. **C** The 100 genes most correlated with GRGs scores. **D**, **E** GO and DO analyses for the 228 genes connected to GRGs activity

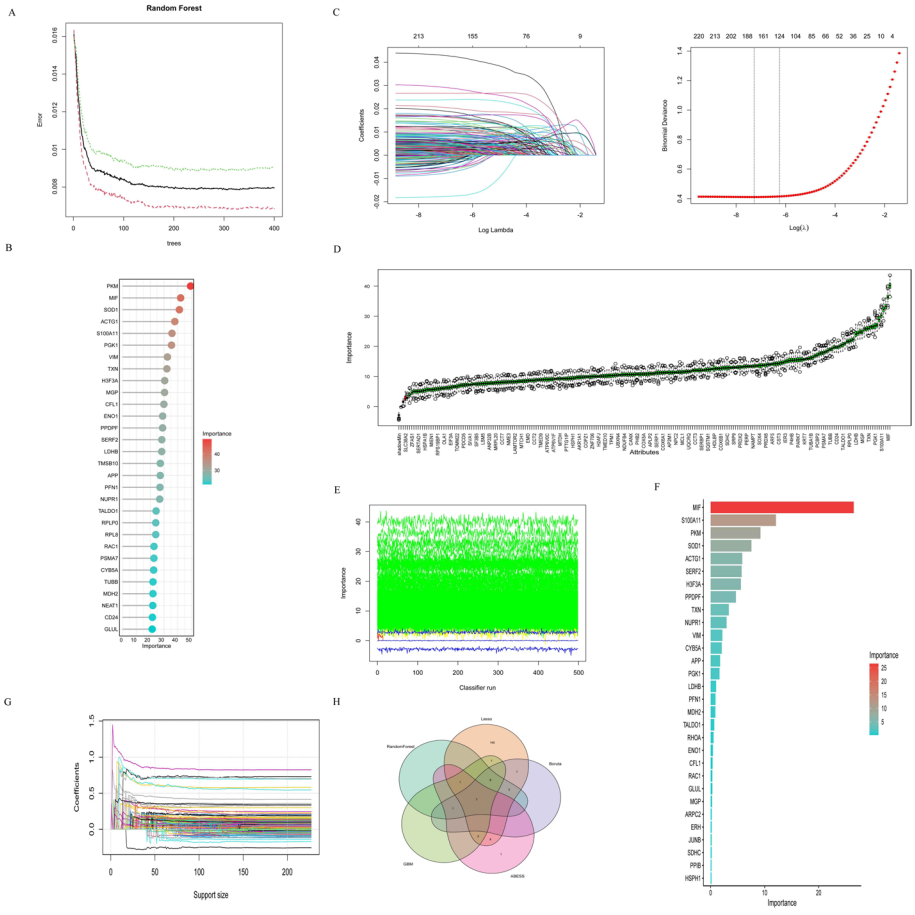

Fig. 6 Hub genes were identified via machine learning approaches. **A**, **B** RF algorithm identified hub genes by assessing feature importance across an ensemble of decision trees. **C** The LASSO algorithm performs feature selection by applying an L1 regularization penalty to constrain regression coefficients. **D**, **E** The Boruta algorithm evaluates gene importance by comparing it against the maximum importance of randomized shadow features. **F** GBM algorithm was subsequently applied to refine the signature gene set. **G** The ABESS algorithm was employed for further gene screening and identification. **H** A Venn diagram displays the consensus set of nine differentially expressed genes derived from the five machine learning algorithms

are comprehensively provided in Supplementary Table 8, demonstrating their consistent and high-ranking prioritization by all methodologically distinct models.

3.7 Sensitivity analysis validates the robustness of hub genes

To validate the robustness of our findings, we performed a sensitivity analysis using an alternative stratification method (the top and bottom 10% of cells by glycolytic score). The cell populations identified as metabolically extreme showed partial overlap with our original IQR-based groups (Jaccard index = 0.400). Crucially, all seven hub glycolytic genes (PKM, MIF, SOD1, PGK1, APP, LDHB, and NUPR1) remained significantly upregulated in the high-glycolytic groups defined by the stricter 10% threshold, with concordantly stronger fold-changes, confirming that our core discovery is robust and not an artifact of the specific classification boundary (Supplementary Tables 9–10).

3.8 Performance evaluation of hub GRGs under multi-model perspectives

This study evaluated the classification performance of multiple biomarkers, with ROC curves demonstrating their sensitivity-specificity relationships. The area under the curve

(AUC) values, serving as an indicator of overall discriminative accuracy, were: PKM: 0.796 (95% CI 0.791 – 0.800), MIF: 0.843 (95% CI 0.839 – 0.848), SOD1: 0.732 (95% CI 0.727 – 0.738), PGK1: 0.693 (95% CI 0.688 – 0.699), APP: 0.644 (95% CI 0.640 – 0.649), LDHB: 0.628 (95% CI 0.622 – 0.634), and NUPR1: 0.653 (95% CI 0.649 – 0.657) (Fig. 7A). Subsequently, we systematically compared the performance of various machine learning models (KNN, LDA, LR, NB, Ranger, RPART, and XGBoost) using their average AUC values as benchmarks to assess their relative effectiveness in this specific task (Fig. 7B). Figure 7C displays the precision-recall curves for each classifier. The model's comprehensive performance was evaluated using ROC curves in Fig. 7D–E, with an overall AUC of 0.959, demonstrating the algorithm's reliability in both specificity and sensitivity. We validated core gene expression patterns using bulk RNA sequencing data and further assessed their correlation with survival status (Fig. 7F–L, Supplementary Fig. 3A–G).To further investigate the prognostic value of the identified genes, we analyzed their expression in relation to overall survival (OS) in luminal A, luminal B, and TNBC subtypes. In the luminal A subtype, high expression of PKM and SOD1 predicts better OS, whereas high expression of NUPR1 is associated with poorer prognosis. In the luminal B subtype, high expression of PGK1 and LDHB is significantly correlated with worse survival outcomes. In TNBC, high expression of NUPR1 similarly serves as a risk factor for poor prognosis, further underscoring the subtype-dependent nature of its biological function(Supplementary Fig. 4, 5 and 6).To evaluate the prognostic value of the glycolytic gene signature beyond conventional clinical parameters, we performed an analysis in the TCGA-BRCA cohort. Comparison of model discrimination showed

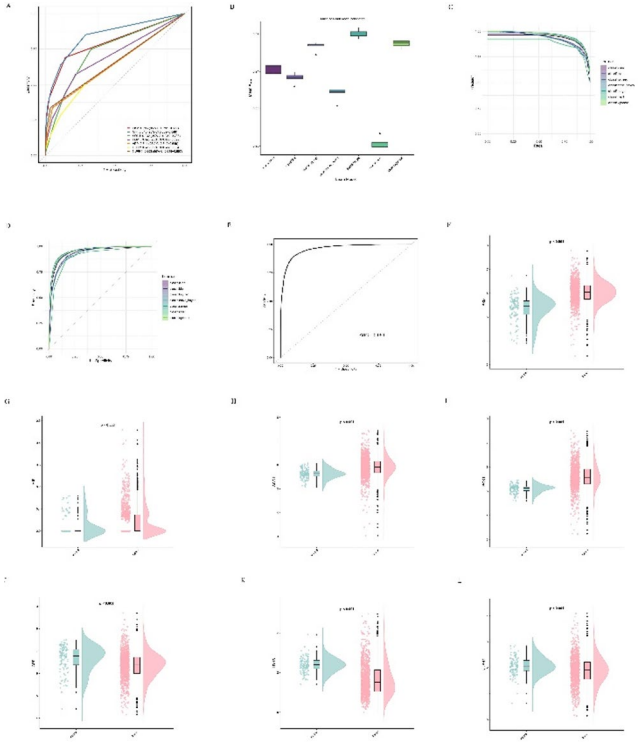

Fig. 7 Multi-model benchmarking of GRGs. **A** ROC analysis was used to evaluate the discriminatory power of the hub genes. **B** Performance of benchmark machine learning models. **C** Precision-recall analysis aids in model optimization to minimize false positives and false negatives. **D**, **E** Comparative performance of various machine learning models based on ROC curves. **F**–**L** Expression patterns of the feature genes in bulk RNA-seq data

that the combined model integrating clinical variables and glycolytic genes achieved a C-index of 0.824 (95% CI 0.673–0.933), which was significantly higher than that of the clinical-only model (C-index = 0.754, 95% CI 0.628–0.864). The improvement in model fit was highly statistically significant ($P < 0.001$) (Supplementary Table 11). Multivariable Cox regression adjusted for clinical covariates confirmed that four of the seven glycolytic genes remained independent prognostic factors(Supplementary Table 12).These results demonstrate that the glycolytic gene signature provides significant incremental prognostic information.

3.9 Hub GRGs pathway enrichment landscape

To thoroughly analyze the functions of hub gene sets, we conducted gene set enrichment analysis (GSEA), revealing significant associations between seven hub gene sets and the glycolysis pathway (Fig. 8A–G). To evaluate their diagnostic performance, we plotted ROC curves in bulk RNA-seq datasets. Quantifying performance through AUC, the results showed the following AUC values: PKM: 0.854 (95% CI 0.821 – 0.887), MIF: 0.570 (95% CI 0.530 – 0.611), SOD1: 0.752 (95% CI 0.717 – 0.786), PGK1: 0.854 (95% CI 0.832 – 0.876), APP: 0.665 (95% CI 0.609 – 0.722), LDHB: 0.804 (95% CI 0.778 – 0.830),

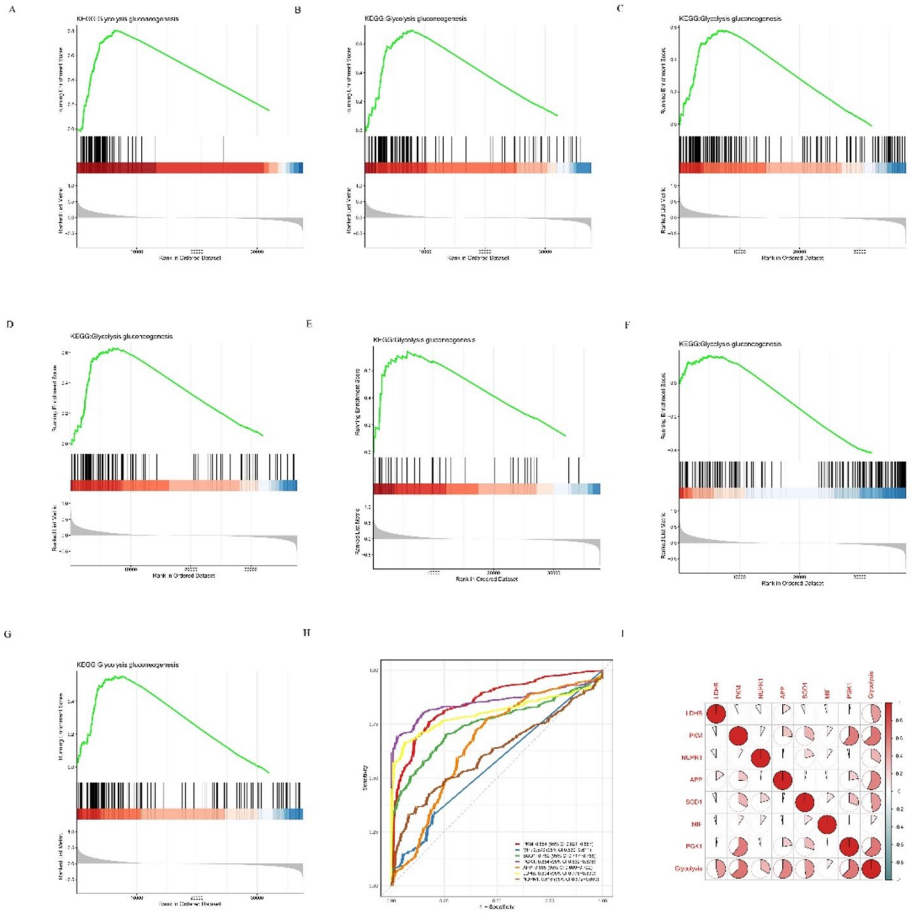

Fig. 8 **A**–**G** GSEA was applied to hub GRGs in malignant cells. **H** Profiling the diagnostic power of the seven hub genes using ROC analysis on bulk data. **I** Correlation analysis depicting the relationships between all seven hub genes

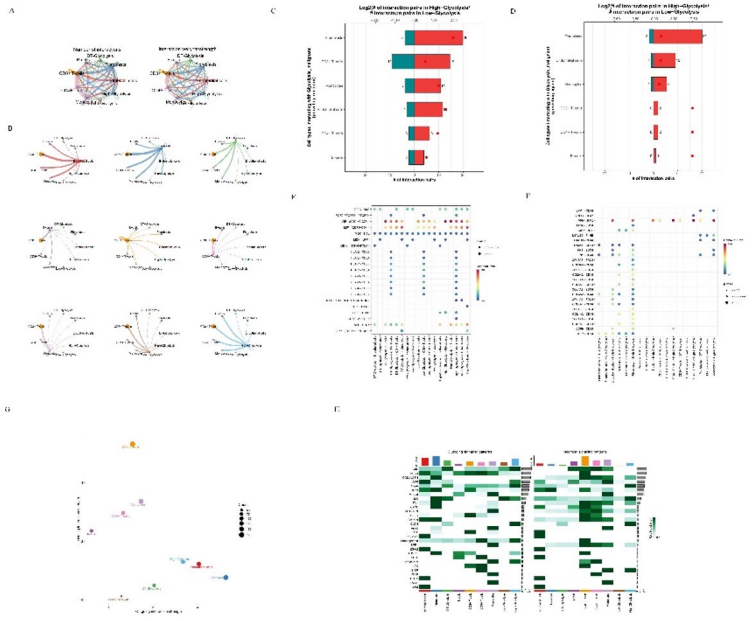

Fig. 9 Characterization of intercellular signaling networks. **A** Depicts the quantity of interactions and their corresponding intensity. **B** Specific interactions between ligands and their respective receptors are displayed. **C**, **D** Identification of cell populations that communicate with glycolytic malignant cells via receptor or ligand provision. **E**, **F** Bubble plots illustrate ligand-receptor pairs involved in signaling to HGS, DTGS, and LGS cell types. **G** Scatter plot visualizes variations in incoming and outgoing interaction strengths across different cell populations. **H** Comparative analysis of afferent and efferent signaling patterns among all cell types

and NUPR1: 0.616 (95% CI 0.572 – 0.660) (Fig. 8H). Subsequently, we demonstrated the interactions and correlation strengths among the seven core GRGs (Fig. 8I).

3.10 A panoramic view of intercellular communication

To elucidate the impact of glycolysis on the function of highly malignant cells, we focused on comparing the cell adhesion behaviors of HGS ($n = 4429$) and LGS ($n = 4429$) tumor cells. Inferred cell-cell communication analysis suggested potential signaling interactions between spatial domains enriched for high-glycolytic signatures and domains characterized by fibroblast, CD8 + T-cell, and endothelial cell markers (Fig. 9A–D). A key finding was that HGS cells exhibited significantly higher signal outflow compared to LGS cells (Fig. 9G). We further examined whether the identified seven core glycolytic genes (PKM, MIF, SOD1, PGK1, APP, LDHB, and NUPR1) were directly involved in these ligand-receptor interactions. Our analysis revealed that MIF, a key core gene, functions as a major ligand in the enhanced signaling output from HGS cells, particularly in the MIF-(CD74 + CXCR4) signaling axis. This pathway showed significantly higher communication probability from HGS cells to fibroblasts ($P < 0.01$) and CD8 + T cells ($P < 0.01$), suggesting that MIF-mediated signaling may serve as a metabolic-immune-stromal communication bridge in the tumor microenvironment. (Fig. 9E, F, H). Subsequent correlation analyses of markers (PKM, MIF, SOD1, PGK1, APP, LDHB, and NUPR1) showed PKM to be the most strongly correlated marker (Supplementary Fig. 7A–G).

3.11 Validation of the GRGs using clinical tissue samples

We used immunohistochemical staining to detect the expression of GRGs proteins, namely PKM, MIF, SOD1,PGK1,APP and LDHB, in the tumors and normal tissues of

patients. The results showed that the expression of these proteins in tumor tissues was higher than those in normal tissues(Supplementary Fig. 8).

4 Discussion

Metabolic reprogramming is a hallmark of cancer [25]. Aerobic glycolysis serves as the primary mechanism for tumor cells to reprogram glucose metabolism, creating a highly acidic microenvironment that facilitates cancer cell proliferation, invasion, and immune evasion [26]. Although glycolysis has garnered significant attention in oncology, systematic exploration of regulatory genes involved in this process in breast cancer remains limited. Therefore, a comprehensive understanding of the GRGs functional network will significantly advance precision therapeutic strategies for breast cancer.

This study establishes a novel methodology for tumor research by integrating single-cell transcriptome sequencing, machine learning, and spatial transcriptomic analysis. The multimodal research strategy systematically elucidates glycolytic metabolic characteristics of breast cancer from dual perspectives of cell type and spatial distribution, revealing diverse metabolic remodeling patterns among malignant cells and expanding our understanding of tumor metabolic complexity. Based on single-cell transcriptomic data, we identified seven distinct cell categories. Computational methods such as AUCell and ssgsea were employed to assess sample glycolytic activity, revealing that cancer cells exhibit the highest metabolic activity. By combining single-cell sequencing with interquartile range stratification, this study achieved precise analysis of glycolysis-related gene expression changes in breast cancer cells, demonstrating significant metabolic heterogeneity. Compared to conventional expression analysis methods that struggle to capture intercellular metabolic differences, the IQR classification technique quantifies metabolic states at single-cell level by categorizing cells into distinct glycolytic activity tiers, thereby significantly enhancing the reliability of subsequent analyses. Using this cell clustering approach, we successfully identified specific high-expressing genes within high glycolytic activity subpopulations. This precise cell classification provides crucial insights into key metabolic perturbations affecting tumor proliferation and invasion. The strategy not only revolutionizes traditional differential gene identification approaches but also substantially improves analytical rigor while effectively controlling biases caused by metabolic heterogeneity. Comprehensive analysis revealed significant differences in glycolytic characteristics among various cell types, leading to the selection of the HGS subpopulation as the primary focus for subsequent research. Through spatial transcriptomics and cell interaction analysis, this study identified unique metabolic activity patterns and cellular communication networks in breast cancer cells within their microenvironment. To date, no research has systematically integrated single-cell transcriptome sequencing, spatial transcriptomics technology, and machine learning models to identify glycolysis-related biomarkers in breast cancer.

To enhance the robustness of gene screening, this study employed five machine learning algorithms based on different mathematical principles: Random Forest (RF), Boruta, Lasso regression, ABESS, and Gradient Boosting Machine (GBM). This multi-algorithm fusion strategy effectively suppressed overfitting while improving reproducibility through complementary effects among models. By constructing a composite model framework, we systematically identified seven core genes: PKM, MIF, SOD1, PGK1, APP, LDHB, and NUPR1, significantly enhancing the screening system's anti-interference

capability. PKM, a key rate-limiting enzyme in aerobic glycolysis of tumor cells, promotes glycolytic flux through high expression. It enters the nucleus to form complexes with other proteins, activating signaling pathways like HIF1-α/β-catenin, thereby driving glycolytic reprogramming and tumor progression [27]. In breast cancer, PKM exhibits strong expression in numerous clinical tissue samples, with its abnormal overexpression correlating with poor prognosis and resistance to neoadjuvant chemotherapy [28]. Our analysis suggested enhanced signaling from high-glycolytic cells, notably via MIF-related pathways. This finding aligns with known breast cancer biology: MIF, secreted by cancer stem cells, can activate pathways such as WNT/β-catenin that promote glycolysis and suppress immunity. Thus, glycolytic reprogramming in malignant cells may not only meet bioenergetic demands but also, through secreted factors like MIF, actively remodel the microenvironment by potentially influencing immune cells and fibroblasts. These inferred interactions support the concept of "metabolically triggered signaling," in which metabolites and secreted factors from active tumor cells drive pro-tumorigenic communication networks [29]. The clinical relevance of our core glycolytic genes (PKM, MIF, SOD1, PGK1, APP, LDHB, NUPR1) is shaped by their context within breast cancer heterogeneity. While glycolysis is a cancer hallmark, key genes exhibit breast cancer-specific roles. For instance, PKM has unique non-metabolic functions in breast cancer nuclei, while LDHB and PGK1 are particularly overexpressed and drive progression in aggressive TNBC [30–32]. Furthermore, MIF-CD74 signaling actively shapes the TNBC immunosuppressive microenvironment, and SOD1 maintains stemness in therapy-resistant cells [33, 34]. This collective evidence suggests our gene panel may not only mark general glycolytic flux but could also reflect the distinct metabolic dependencies of aggressive breast cancer subtypes, informing the development of subtype-specific therapeutic strategies. As a transcriptional regulator, NUPR1 induces autophagy by activating TFE3 transcription, a process closely linked to breast cancer cell proliferation and migration [35]. Our immunohistochemical studies confirmed elevated expression of PKM, MIF, SOD1, PGK1, APP and LDHB in breast cancer tumor tissues compared to normal tissues, further demonstrating these glycoproteins' critical roles in breast cancer development, metabolic reprogramming and poor clinical prognosis.

Our integrated multi-algorithm framework advances beyond conventional differential expression analysis by prioritizing consistently important genes, such as the seven hub genes (PKM, MIF, SOD1, PGK1, APP, LDHB, NUPR1), while filtering out biologically peripheral signals. The convergence of five distinct algorithms indicates that these genes form a coherent functional module within the glycolytic network. This is supported by their known roles in glycolysis, stress response, and cancer progression, as well as their coordinated expression and interactions shown in our cell communication analysis. Thus, our methodology distills complex transcriptional data into a focused, biologically interpretable core gene set.

While this study has yielded significant findings, its conclusions require cautious interpretation. First, although retrospective data analysis suggests GRGs hold diagnostic potential, their clinical value remains to be validated through large-scale prospective cohort studies. Second, the spatial transcriptomics data, while informative for mapping regional heterogeneity, are constrained by spot-level resolution. This limitation necessitates cautious interpretation of the inferred cell-cell communication networks, as signals are assigned to spatial domains containing multiple cells rather than to individual cells,

and restricts our ability to precisely map glycolytic metabolic domains at single-cell resolution. Higher-resolution spatial technologies are needed to reveal molecular details. Finally, three critical challenges demand urgent resolution: direct experimental evidence demonstrating GRGs' regulation of glycolytic pathways, mechanisms mediating cellular interactions in tumor microenvironments, and the clinical translational efficacy of GRGs as standalone biomarkers.We acknowledge that the mechanistic links between the identified hub genes and glycolytic flux remain correlative. Future functional experiments are needed to establish causality. However, our study provides a strongly prioritized target list for such validation, offering a clear translational roadmap.

5 Conclusion

By integrating single-cell transcriptomics, spatial transcriptomics, and multiple machine learning algorithms, this study systematically analyzed the glycolytic metabolic heterogeneity in breast cancer and successfully identified seven core glycolytic regulatory genes: PKM, MIF, SOD1, PGK1, APP, LDHB, and NUPR1. These genes exhibit significantly elevated expression in tumor tissues and are closely associated with enhanced glycolytic activity, malignant progression, and poor prognosis. The research not only constructed a spatial distribution atlas of highly metabolic cell subpopulations but also revealed their unique cellular communication patterns, providing novel candidate biomarkers and potential intervention strategies for targeted metabolic therapy in breast cancer.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s12672-026-04826-3.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements
This work benefits from the collective contributions of researchers and consortia across multiple platforms, particularly acknowledging data resources from TCGA and GEO.

Author contributions
Y.X.L. designed the study and wrote the manuscript, C.H.T. performed the literature search and analyzed the data, F.W. and C.M.L. revised the first draft, and F.W. critically revised the manuscript and provided final approval. Y.X.L. and C.M.L. contributed equally to this work, and all authors read the manuscript and approved its content.

Funding
Not applicable.

Data availability
The datasets analyzed during the current study are available in public repositories. The Gene Expression Omnibus (GEO) datasets (GSE176078 and GSE161529) are available from the GEO database (https://www.ncbi.nlm.nih.gov/geo/) via the following accession links: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176078 and https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161529. The TCGA-BRCA dataset is available from the TCGA database managed by the Genomic Data Commons (GDC) Data Portal at https://portal.gdc.cancer.gov/projects/TCGA-BRCA. The spatial transcriptomics dataset is available from the Zenodo repository at https://zenodo.org/records/4739739. All these datasets are publicly available as per the data access policies of their respective repositories.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 18 November 2025 / Accepted: 10 March 2026
Published online: 14 March 2026

References

1. Force LM, Kocarnik JM, May ML, Bhangdia K, Crist A, Penberthy L, Pritchett N, Acheson A, Deitesfeld L, Aalruz H, Abbas N. The global, regional, and national burden of cancer, 1990–2023, with forecasts to 2050: a systematic analysis for the Global Burden of Disease Study 2023. Lancet. 2025;406(10512):1565–86. https://doi.org/10.1016/s0140-6736(25)01635-6.
2. Kishton RJ, Sukumar M, Restifo NP. Metabolic Regulation of T Cell Longevity and Function in Tumor Immunotherapy. Cell Metab. 2017;26(1):94–109. https://doi.org/10.1016/j.cmet.2017.06.016.
3. Talaiezadeh A, Shahriari A, Tabandeh MR, Fathizadeh P, Mansouri S. Kinetic characterization of lactate dehydrogenase in normal and malignant human breast tissues. Cancer Cell Int. 2015;15:19. https://doi.org/10.1186/s12935-015-0171-7.
4. Pelicano H, Martin DS, Xu RH, Huang P. Glycolysis inhibition for anticancer treatment. Oncogene. 2006;25(34):4633–46. https://doi.org/10.1038/sj.onc.1209597.
5. Wei L, Zhou Y, Qiao C, et al. Oroxylin A inhibits glycolysis-dependent proliferation of human breast cancer via promoting SIRT3-mediated SOD2 transcription and HIF1α destabilization. Cell Death Dis. 2015;6(4):e1714. https://doi.org/10.1038/cddis.2015.86.
6. Yoshitake R, Mori H, Ha D et al. Molecular features of luminal breast cancer defined through spatial and single-cell transcriptomics. Clin Transl Med 14(1):e1548. https://doi.org/10.1002/ctm2.1548
7. Liu YM, Ge JY, Chen YF, et al. Combined Single-Cell and Spatial Transcriptomics Reveal the Metabolic Evolvement of Breast Cancer during Early Dissemination. Adv Sci (Weinh). 2023;10(6):e2205395. https://doi.org/10.1002/advs.202205395.
8. Feng K, Li X, Li X, et al. Integrative analysis of bulk and single-cell transcriptomics identifies factors related to immunosuppressive microenvironment to predict unfavorable prognosis in inflammatory breast cancer. Front Immunol. 2025;16:1727590. https://doi.org/10.3389/fimmu.2025.1727590.
9. Wang X, Wang Z, Liao Q, et al. Spatially resolved atlas of breast cancer uncovers intercellular machinery of venular niche governing lymphocyte extravasation. Nat Commun. 2025;16(1):3348. https://doi.org/10.1038/s41467-025-58511-0.
10. Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. KEGG: biological systems database as a model of the real world. Nucleic Acids Res. 2025;53(D1):D672–7. https://doi.org/10.1093/nar/gkae909.
11. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28(11):1947–51. https://doi.org/10.1002/pro.3715.
12. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27.
13. Slovin S, Carissimo A, Panariello F, et al. Single-Cell RNA Sequencing Analysis: A Step-by-Step Overview. Methods Mol Biol. 2021;2284:343–65. https://doi.org/10.1007/978-1-0716-1307-8_19.
14. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol. 2019;20(2):163–72. https://doi.org/10.1038/s41590-018-0276-y.
15. Liu Y, Li H, Zeng T, et al. Integrated bulk and single-cell transcriptomes reveal pyroptotic signature in prognosis and therapeutic options of hepatocellular carcinoma by combining deep learning. Brief Bioinform. 2023. https://doi.org/10.1093/bib/bbad487.
16. Andreatta M, Carmona SJ, UCell. Robust and scalable single-cell gene signature scoring. Comput Struct Biotechnol J. 2021;19:3796–8. https://doi.org/10.1016/j.csbj.2021.06.043.
17. Bhuva DD, Foroutan M, Xie Y, Lyu R, Cursons J, Davis MJ. Using singscore to predict mutation status in acute myeloid leukemia from transcriptomic signatures. F1000Res. 2019;8:776. https://doi.org/10.12688/f1000research.19236.3.
18. Jin Y, Wang Z, He D, Zhu Y, Chen X, Cao K. Identification of novel subtypes based on ssGSEA in immune-related prognostic signature for tongue squamous cell carcinoma. Cancer Med. 2021;10(23):8693–707. https://doi.org/10.1002/cam4.4341.
19. Mei Y, Li M, Wen J, Kong X, Li J. Single-cell characteristics and malignancy regulation of alpha-fetoprotein-producing gastric cancer. Cancer Med. 2023;12(10):12018–33. https://doi.org/10.1002/cam4.5883.
20. Morabito S, Reese F, Rahimzadeh N, Miyoshi E, Swarup V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. Cell Rep Methods. 2023;26(6):100498. https://doi.org/10.1016/j.crmeth.2023.100498.
21. Li CL, Yao ZY, Qu C, et al. Machine learning model reveals the risk, prognosis, and drug response of histamine-related signatures in pancreatic cancer. Discov Oncol. 2025;16(1):155. https://doi.org/10.1007/s12672-025-01910-y.
22. Fang Z, Tian Y, Sui C, et al. Single-Cell Transcriptomics of Proliferative Phase Endometrium: Systems Analysis of Cell-Cell Communication Network Using CellChat. Front Cell Dev Biol. 2022;10:919731. https://doi.org/10.3389/fcell.2022.919731.
23. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017;14(10):979–82. https://doi.org/10.1038/nmeth.4402.
24. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. Protein Sci. 2018;27(1):233–44. https://doi.org/10.1002/pro.3307.
25. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646–74. https://doi.org/10.1016/j.cell.2011.02.013.
26. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science. 2009;324(5930):1029–33. https://doi.org/10.1126/science.1160809.
27. Liu L, Wang J, Liang L, Chen C, Xie M, Tang W. Targeting PKM2 in cancer therapeutics: mechanistic advances and translational opportunities. Front Immunol. 2025;16:1649488. https://doi.org/10.3389/fimmu.2025.1649488.
28. Wang Y, Zhao H, Zhao P, Wang X. Targeting PKM2 promotes chemosensitivity of breast cancer cells in vitro and in vivo. Cancer Biomark. 2021;32(2):221–30. https://doi.org/10.3233/cbm-210111.
29. Yan L, Wu M, Wang T, et al. Breast Cancer Stem Cells Secrete MIF to Mediate Tumor Metabolic Reprogramming That Drives Immune Evasion. Cancer Res. 2024;84(8):1270–85. https://doi.org/10.1158/0008-5472.Can-23-2390.
30. Wang M, Gao Z, Zhao R, et al. METTL14-Mediated M(6)A Modification of LINC01094 Induces Glucose Metabolic Reprogramming in Breast Cancer by Recruiting the PKM2/JMJD5 Complex. Adv Sci (Weinh). 2025;12(35):e10386. https://doi.org/10.1002/advs.202410386.

31. Dennison JB, Molina JR, Mitra S, et al. Lactate dehydrogenase B: a metabolic marker of response to neoadjuvant chemotherapy in breast cancer. Clin Cancer Res. 2013;19(13):3703–13. https://doi.org/10.1158/1078-0432.Ccr-13-0623.
32. Zhang H, Ling M, Zhang Y, Fang Q, Wo W, Lv X. OXCT1 promotes triple negative breast cancer immune escape via modulating succinylation modification of PGK1. Commun Biol. 2025;8(1):1033. https://doi.org/10.1038/s42003-025-08433-w.
33. Wang S, Wang Y, Bu Y, et al. ISLR knockdown enhances radiotherapy-induced anti-tumor immunity by disrupting the Treg-mregDC-lymphoid niche in triple-negative breast cancer. Int Immunopharmacol. 2025;161:114988. https://doi.org/10.1016/j.intimp.2025.114988.
34. Luo M, Shen N, Shang L, et al. Simultaneous Targeting of NQO1 and SOD1 Eradicates Breast Cancer Stem Cells via Mitochondrial Futile Redox Cycling. Cancer Res. 2024;84(24):4264–82. https://doi.org/10.1158/0008-5472.Can-24-0800.
35. Xiao H, Long J, Chen X, Tan MD. NUPR1 promotes the proliferation and migration of breast cancer cells by activating TFE3 transcription to induce autophagy. Exp Cell Res. 2022;418(1):113234. https://doi.org/10.1016/j.yexcr.2022.113234.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.