



# OPEN Functional genomics integration of glycolysis-related gene networks reveals prognostic biomarkers and immune microenvironment regulation in breast cancer

Yingjie Niu, Yu Jiang, Zhitang Wang & Debo Chen✉

Breast cancer (BC) is the most common malignancy among women worldwide, with metabolic reprogramming, particularly alterations in glycolytic pathways, playing a crucial role in its development and progression. This study aimed to perform an integrative multi-omics analysis of glycolysis-related genes (GRGs) expression profiles in breast cancer, construct a prognostic prediction model, explore the regulatory mechanisms of glycolytic reprogramming on the tumor immune microenvironment, and identify potential therapeutic targets. We integrated multi-omics data from TCGA and METABRIC databases with single-cell RNA sequencing data from GEO. Key GRGs were identified through differential expression analysis and WGCNA. Patients were stratified via consensus clustering, and a prognostic model was developed using Cox and LASSO regression analyses. Immune microenvironment features were characterized using ESTIMATE and CIBERSORT algorithms. Single-cell sequencing analysis revealed glycolytic profiles across different cell types and intercellular communication networks. Mendelian randomization was performed to establish causal relationships between GRGs and BC, using eQTL data as instrumental variables with multiple estimation methods (IVW, weighted median, MR-Egger) and rigorous validity assessments. Potential therapeutic agents were identified through molecular docking, and key gene expression was validated by PCR. We established a prognostic risk model based on GRGs that demonstrated good predictive value in both TCGA and METABRIC cohorts. Immune microenvironment analysis revealed increased eosinophils and M2 macrophages in the high-risk group, while dendritic cells and effector T cells were enriched in the low-risk group. Single-cell analysis confirmed higher glycolytic activity in myeloid cells and T cells within tumor tissues. Cell communication network analysis demonstrated that myeloid cells primarily interact with other cells through MHC-II, MIF, and SPP1 signaling pathways, whereas T cells mainly communicate via MHC-I, CCL, and CXCL pathways. Mendelian randomization (MR) identified NT5E and NRG1 as protective factors and S100B as a risk factor. Molecular docking revealed trametinib and AZD8055 as potential therapeutic agents. Our study established a prognostic model based on GRGs, revealed causal relationships between NT5E, NRG1, and S100B with BC prognosis, and elucidated the connection between glycolytic activity and immune microenvironment remodeling. These findings expand our understanding of metabolic reprogramming mechanisms in breast cancer and provide a theoretical foundation for precision therapeutic strategies based on the metabolism-immune axis, potentially improving clinical outcomes for BC patients.

**Keywords** Breast cancer (BC), Glycolysis, Immune microenvironment, Single-cell sequencing, Mendelian randomization (MR)

Breast cancer (BC), as the most common malignant tumor among women globally, continues to exhibit an increasing disease burden<sup>1</sup>. According to GLOBOCAN 2022 data, there are approximately 2.3 million new cases worldwide annually (accounting for 11.6% of total global cancer cases), with approximately 666,000 deaths,

Department of Breast Surgery, Quanzhou First Hospital of Fujian Medical University, Quanzhou, Fujian, People's Republic of China. ✉email: debochensr@fjmu.edu.cn

ranking first among cancer-related mortality in women<sup>2</sup>. The current standard treatment regimens for BC include surgical resection combined with adjuvant therapies (such as chemotherapy, radiotherapy, endocrine therapy, molecular targeted therapy, and immunotherapy); however, local recurrence and distant metastasis remain major challenges in clinical management<sup>3</sup>. Despite advances in subtype-specific therapies, such as endocrine therapy for luminal subtypes and anti-HER2 agents for HER2-enriched BC, challenges persist, including considerable toxicity, adverse effects, and suboptimal 5-year survival rates, particularly for aggressive subtypes like TNBC<sup>4</sup>. Therefore, identifying novel prognostic biomarkers and constructing high-precision prognostic prediction models hold significant clinical relevance and translational value for achieving precise stratified management of BC patients, guiding individualized treatment decisions, and improving long-term survival outcomes.

Metabolic reprogramming is one of the hallmark features of tumors, with alterations in glycolytic pathways playing a central role in BC initiation and progression<sup>5</sup>. Unlike normal cells that primarily rely on mitochondrial oxidative phosphorylation, BC cells preferentially generate ATP through glycolysis even under oxygen-sufficient conditions (known as the “Warburg effect”), providing energy support and biosynthetic precursors for tumors<sup>6</sup>. Enhanced glycolytic activity promotes tumor cell proliferation, inhibits apoptosis, and is closely linked to invasion, metastasis, and chemotherapy resistance, with subtype-specific differences in metabolic profiles, such as higher glycolytic dependency in TNBC compared to luminal subtypes<sup>7,8</sup>. Clinical evidence indicates that elevated glycolysis correlates with reduced survival in BC patients, particularly in aggressive subtypes<sup>9</sup>. Additionally, lactate produced through glycolysis facilitates tumor progression by acidifying the tumor microenvironment and suppressing immune cell function<sup>10</sup>. However, the genetic regulatory mechanisms underlying the heterogeneous expression of glycolysis-related genes (GRGs) and their potential clinical application value had not yet been systematically elucidated.

In recent years, the rapid development of multi-omics technologies has provided new avenues for deciphering the regulatory networks of glycolysis gene expression. Expression quantitative trait loci (eQTL) analysis offers an important approach for understanding the relationship between genetic variations and gene expression regulation<sup>11</sup>. Multiple eQTL studies have confirmed that specific genetic variants are significantly associated with the expression levels of key glycolytic enzymes (including HK2, PFKFB3, and LDHA), providing a molecular basis for understanding how genotypes influence the metabolic phenotypes of BC cells<sup>12–14</sup>. Genome-wide association studies (GWAS) have identified multiple genetic susceptibility loci associated with BC metabolic phenotypes<sup>15</sup>. Mendelian randomization (MR) methodology, by utilizing genetic variants as instrumental variables, provides a methodological foundation for inferring causal relationships between glycolysis regulatory genes and BC risk<sup>16</sup>. Simultaneously, single-cell RNA sequencing (scRNA-seq) technology has enabled in-depth analysis of glycolysis gene expression heterogeneity and its interaction patterns with the tumor microenvironment at the cellular resolution level, offering new perspectives for understanding the complexity of the BC metabolic microenvironment<sup>17</sup>.

This study integrates multi-omics data (eQTL, GWAS, scRNA-seq, and bulk RNA-seq) to systematically analyze GRG expression profiles across BC molecular subtypes. Using machine learning algorithms, we constructed a robust prognostic prediction model based on a glycolysis risk score, enabling precise patient stratification. We comprehensively evaluated immune infiltration, drug sensitivity, functional enrichment, and clinical correlations between high- and low-risk groups, with a focus on subtype-specific differences. Furthermore, we innovatively analyzed intercellular communication networks among cells with distinct glycolytic states at single-cell resolution, revealing novel mechanisms of metabolism-immune interactions in the context of BC molecular heterogeneity. This study not only deepens the molecular understanding of metabolic reprogramming in BC but also identifies potential therapeutic targets and provides subtype-specific clinical stratification strategies, laying a theoretical foundation for precision oncology in breast cancer.

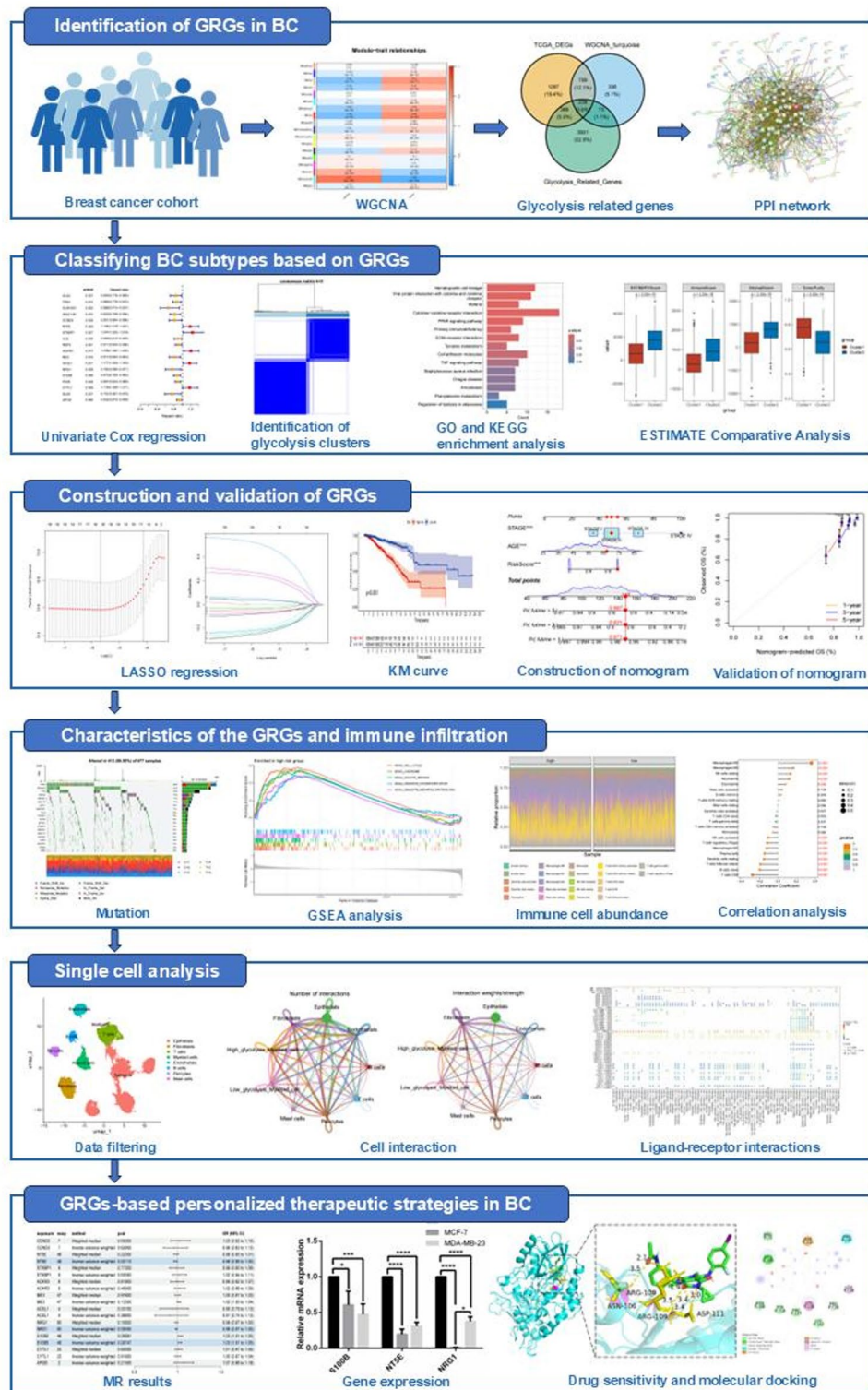
## Methods and materials

### Data acquisition and processing

We initially obtained RNA sequencing, mutation, and clinical data of breast cancer (BC) from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). After excluding male patients, cases with incomplete clinical information, and those with follow-up duration less than 30 days, a total of 1,130 patients were included, comprising 1,017 BC samples and 113 normal tissue samples. Additionally, we utilized the METABRIC database as a validation cohort, which was downloaded from the cBioPortal database (<http://www.cbioportal.org/>). In addition, we further stratified the TCGA and METABRIC cohorts into LumA, LumB, HER2 positive and triple negative breast cancer (TNBC) subtypes according to the clinical HR/HER2 status. Single-cell RNA sequencing (scRNA-seq) data were acquired from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>) under the accession number GSE161529, from which 18 samples were selected, including 11 BC samples and 7 normal samples. Based on previous literature reports<sup>18</sup>, combined with GeneCards (<https://www.genecards.org/>) and gene sets from HALLMARK\_GLYCOLYSIS and REACTO ME\_GLYCOLYSIS in MsigDB, we identified a total of 4,200 glycolysis-related genes (GRGs) (Supplementary Table 1). As all patient data incorporated in this study were obtained from public databases and strictly adhered to relevant usage guidelines, ethical committee approval was not required. Figure 1 presents a flow diagram illustrating the study design (Fig. 1).

### Integrated bioinformatics approach for identifying key GRGs in BC

This study employed a multilevel bioinformatics approach to identify critical GRGs in BC. Initially, differentially expressed genes (DEGs) were screened from the TCGA-BRCA dataset using the “limma” package in R (version 4.3.0), with selection criteria established as  $|\log_2FC| > 1$  and  $P < 0.05$ . Subsequently, based on the TCGA-BRCA dataset, we implemented weighted gene co-expression network analysis (WGCNA) to comprehensively explore the molecular characteristics of BC through rigorous data preprocessing and network construction



**Fig. 1.** Graphical abstract for comprehensive characterization of the GRGs in BC. This flowchart outlines the study of GRGs in BC, detailing methods for GRGs identification using WGCNA and PPI networks, BC subtype classification via gene expression, and validation through LASSO regression and nomograms. It also highlights GRGs characteristics related to immune infiltration, explores single-cell interactions, and suggests personalized therapeutic strategies to enhance treatment outcomes for BC patients.

procedures. The specific analytical steps encompassed: (1) evaluation of sample and gene quality using the “goodSamplesGenes” function to effectively filter unstable genes and samples; (2) hierarchical clustering of samples via the “hclust” function to eliminate outliers; (3) calculation of the soft threshold  $\beta$  to construct a scale-free gene network, followed by transformation of the weighted adjacency matrix into a topological overlap matrix (TOM); (4) application of the dynamic tree-cutting algorithm to identify gene modules with consistent expression patterns; and (5) computation of correlations between module eigenvectors and clinical traits using the “cor” and “corPvalueStudent” functions, ultimately selecting key gene modules highly associated with BC biological features. This comprehensive methodology not only enhanced the precision of gene selection but also provided systematic insights into the molecular mechanisms underlying BC. Finally, we performed cross-analysis among DEGs, WGCNA results, and GRG-related gene sets to identify differentially expressed GRGs, which were subsequently visualized using Venn diagrams.

### Consensus clustering based on GRGs

Unsupervised consensus clustering analysis was performed using the “ConsensusClusterPlus” package (version 1.66.0) in R to stratify patients into distinct clusters. The optimal number of clusters was determined through incremental area analysis, conducting consensus clustering within a range of 2 to 9 clusters and quantifying clustering stability via area changes in the cumulative distribution function (CDF) curves. By evaluating the incremental area between consecutive  $k$  values, we observed a significant reduction when  $k=2$ , indicating an optimal balance between data representativeness and computational efficiency. To ensure the stability of the supervised clustering, all procedures were repeated 1,000 times. Subsequently, survival analysis was conducted to compare outcomes between the identified groups. Principal component analysis (PCA) was employed to visualize the clustering results, facilitating interpretation of the findings.

### Functional enrichment analysis

To identify specific biological pathways enriched with GRGs, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and Gene Ontology (GO) functional enrichment analysis on differentially expressed genes using the “clusterProfiler” (version 4.10.1)<sup>19–21</sup>. The “enrichplot” package (version 1.22.0) was utilized to visualize the KEGG and GSEA analysis results, revealing gene functional networks and biological significance from multiple dimensions. Additionally, protein-protein interaction (PPI) networks were constructed using the STRING tool (<https://www.string-db.org/>) to further elucidate molecular interactions among the identified genes.

### Characterization of the immune microenvironment based on GRGs subtypes

The tumor immune microenvironment (TME), as a critical regulatory platform in tumor biology, profoundly influences tumor cell progression dynamics and metastatic potential<sup>22</sup>. To gain deeper insights into the TME of BC, we employed the ESTIMATE algorithm from the “estimate” R package (version 1.0.13) to calculate immune scores, stromal scores, tumor purity, and ESTIMATE scores (the sum of immune and stromal scores) for tumor samples. Additionally, the CIBERSORT algorithm was utilized to quantify the relative abundance of various immune cell infiltrations within the BC TME, with results visualized using the “ggpubr” package (version 0.6.0) in R.

### Construction and validation of GRGs risk signature

We performed univariate Cox regression analysis on 238 intersection genes to identify GRGs with prognostic value. To avoid overfitting, the least absolute shrinkage and selection operator (LASSO) regression was employed to select genes with high prognostic significance. Based on the expression levels of these genes and their corresponding regression coefficients, a glycolysis-related score was calculated using the following formula: glycolysis score = expression level of gene1  $\times$  coefficient of gene1 + expression level of gene2  $\times$  coefficient of gene2 + ... + expression level of gene $n$   $\times$  coefficient of gene $n$ . According to the median value of the GRGs risk score, we stratified both the TCGA validation cohort and METABRIC validation cohort into two groups: glycolysis-high group and glycolysis-low group. Subsequently, Kaplan-Meier analysis was conducted using the R package “survminer” (version 0.5.0) to compare differences in overall survival (OS) between the two groups. Receiver operating characteristic (ROC) curves were used to evaluate the predictive performance of this signature. Additionally, we explored the correlations between GRGs scores and patients’ clinicopathological characteristics (age, TNM classification, and stage). Univariate and multivariate Cox regression analyses were performed to determine whether GRGs served as independent prognostic factors for BC patient survival. To enhance the prognostic accuracy and predictive capability of the model, we integrated clinicopathological factors with the risk score to construct a nomogram for predicting 1-year, 3-year, and 5-year OS rates in BC patients. Finally, the accuracy and sensitivity of the nomogram were assessed through calibration curves, decision curve analysis (DCA), and ROC analysis. R packages “rms” (version 6.8-1.8), “regplot” (version 1.1) and “survival” (version 3.8-3.8) were utilized for constructing the nomogram and its corresponding calibration curves.

### Analysis of tumor mutation burden (TMB) between GRGs risk subgroups

We extracted somatic mutation data of breast cancer patients from TCGA database and conducted comprehensive analysis of the mutation landscape in GRGs using the “Maftools” package (version 2.22.10). Through systematic mutation frequency analysis, mutation type identification, high-frequency mutation gene screening, and waterfall plot visualization, we thoroughly elucidated the genetic variation characteristics across different glycolysis-related gene expression groups. We selected the 20 genes with the most significant differences between high-risk and low-risk groups for copy number variation (CNV) analysis. Additionally, Spearman correlation analysis was employed to examine the relationship between risk scores and tumor mutational burden (TMB).

### Single-cell RNA-seq data processing and analysis

The Seurat package (version 5.2.1) was employed for single-cell RNA sequencing data analysis. To ensure data quality, stringent filtering criteria were applied: retaining only cells with gene expression counts ranging from 200 to 8000 and mitochondrial gene percentage below 10%. Concurrently, only genes detected in at least 5 cells were preserved for subsequent analysis. The quality-controlled data were normalized and standardized using the “NormalizeData” and “ScaleData” functions, while the “FindVariableFeatures” function identified the top 3000 highly variable genes for dimensionality reduction. Principal component analysis (PCA) was conducted via the “RunPCA” function. To mitigate batch effects between samples, the Harmony algorithm was implemented for data integration. Subsequently, Uniform Manifold Approximation and Projection (UMAP) and t-distributed Stochastic Neighbor Embedding (t-SNE) were utilized to visualize the batch-corrected data. Cell clusters were constructed using graph-based clustering methods through Seurat’s “FindNeighbors” and “FindClusters” functions. Cluster-specific marker genes were identified using the “FindAllMarkers” functionality, and cell types were annotated using the CellMarker 2.0 database (<http://117.50.127.228/CellMarker/>). The “AddModuleScore” function from the Seurat package was used to quantify glycolysis-related gene set activity in each cell. Cells were classified into high and low glycolysis score groups based on the median glycolysis score, with gene visualization performed using the “ggplot2” package (version 3.5.1). Intercellular communication networks between high and low glycolysis score groups were analyzed using the “CellChat” package (version 1.6.1) to identify differential signaling pathways and ligand-receptor interactions between cell populations with distinct metabolic states.

### InferCNV analysis

To distinguish malignant tumor cells from non-cancerous cells within breast epithelial populations in tumor tissues, we inferred large-scale chromosomal copy number variations (CNVs) using the inferCNV R package (version 1.18.1). This method exploits differences in gene expression intensities across genomic positions to detect chromosomal alterations characteristic of malignancy. Input data consisted of a raw gene expression count matrix encompassing all breast epithelial cells, a cell annotation file delineating putative malignant (observation) and normal (reference) cells, and a gene order file specifying chromosomal positions (obtained from <https://github.com/broadinstitute/inferCNV>). Normal breast epithelial cells, preliminarily identified via clustering and marker gene expression, were selected as the reference to establish baseline expression profiles. Results were visualized as heatmaps, with genes ordered by genomic location along rows and cells along columns; inferred amplifications and deletions were represented by red and blue gradients, respectively, enabling the robust separation of malignant from non-malignant epithelial cells based on genomic instability.

### Significance of the GRGs in drug sensitivity

The Genomics of Drug Sensitivity in Cancer (GDSC) (<https://www.cancerrxgene.org>) is a public dataset containing information on cancer cell drug sensitivity and molecular markers of drug response<sup>23</sup>. To facilitate personalized therapy, we utilized the “oncoPredict” package (version 1.2) to predict sensitivity to various anticancer drugs in high-risk and low-risk groups. Wilcoxon test was employed to examine differences in drug IC50 values between high-risk and low-risk groups, with  $p < 0.05$  considered statistically significant. Additionally, we investigated the response patterns of key genes to multiple drugs using the Gene Set Cancer Analysis database (GSCA: <http://bioinfo.life.hust.edu.cn/GSCA/#/>).

### Mendelian randomization (MR) analysis

In this study, we employed MR to investigate potential causal relationships between gene expression and BC risk. Expression quantitative trait loci (eQTL) data serving as exposure variables were obtained from the IEU OpenGWAS project (<https://gwas.mrcieu.ac.uk/>), while outcome data were derived from the FinnGen R12 release (<https://www.finnngen.fi/en>) genome-wide association study (GWAS) dataset, comprising 24,270 breast cancer patients and 222,078 healthy controls, all of European ancestry. To ensure analytical validity, we selected single nucleotide polymorphisms (SNPs) significantly associated with target gene expression ( $P < 5 \times 10^{-8}$ ), applied linkage disequilibrium pruning ( $r^2 < 0.1$ , a clumping distance of 5,000 kb) to obtain independent genetic signals, and retained only SNPs with F-statistic  $> 10$  to avoid weak instrument bias. We implemented multiple complementary MR methods for causal effect estimation: Inverse Variance Weighted (IVW) method as our primary analytical approach, supplemented by MR-Egger, Weighted median, Simple mode, and Weighted mode methods to verify the robustness of our findings. Heterogeneity and pleiotropy tests were conducted ( $P > 0.05$  indicating no significant heterogeneity or pleiotropy). All statistical analyses were performed using the “TwoSampleMR” package (version 0.6.14), with statistical significance defined as two-sided  $P < 0.05$ , and causal effect estimates presented as odds ratios (ORs) with 95% confidence intervals (CIs).

### Cell culture

The cell lines used in this study included: the normal breast epithelial cell line MCF-10 A (CVCL\_0598) procured from the Cell Resource Center of Shanghai Institutes for Biological Sciences, and the human breast cancer cell lines MCF-7 (TCH-C247) and MDA-MB-231 (TCH-C453) obtained from Hycyte Biological (China). All cell lines were authenticated by STR analysis and tested negative for mycoplasma contamination. MCF 10 A cells were cultured in a specific epithelial culture medium (CM-0525, Procell Life Science & Technology Co., Ltd., China). MDA-MB-231 cells were cultured in Dulbecco’s modified Eagle’s medium (DMEM, SH30022.01, Cytiva, USA) supplemented with 10% fetal bovine serum (FBS, 900 – 108, Gemini Bio-Products, USA) and 1% penicillin and streptomycin (PS, P1400, Solarbio, China). MCF-7 cells were maintained in minimum essential medium (MEM; SH30024.01, Cytiva, USA) supplemented with 0.01 mg/ml insulin (PB180432, Procell Life Science & Technology Co., Ltd., China), 10% FBS and 1% PS. All cell lines were kept at 37 °C and 5% CO<sub>2</sub> in a humidified atmosphere.

### RNA extraction and quantitative real-time PCR (qRT-PCR)

Total cellular RNAs were isolated from cells using EasyPure<sup>®</sup> RNA Kit (TransGen, ER101-01, China) according to the manufacturer's instructions. The reverse transcription was performed using *EasyScript<sup>®</sup> One-Step gDNA Removal and cDNA Synthesis SuperMix* (TransGen, AE311-02, China). Next, qRT-PCR was performed using a TB Green<sup>®</sup> Premix Ex Taq<sup>™</sup> II (RR820A, Takara, Japan), and on an Applied Biosystems QuantStudio 6 (Thermo, Waltham, MA, United States). Relative quantification was determined using the  $-2^{\Delta\Delta CT}$  method, and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used as an internal control. The cell experiments were carried out using three biological replicates. The primers were synthesized by Sangon Biotech Co., Ltd. (Shanghai, China). The sequences of the PCR primers are listed in Supplementary Table 2.

### Molecular docking

To evaluate the potential of key genes in BC treatment, we performed molecular docking between Trametinib and AZD8055 and the key genes. Protein structures of NT5E and S100B were obtained from the UniProt database (<https://www.uniprot.org/>). PyMOL 2.6.0 was successfully employed to eliminate all precursors and water molecules present in the targets. Molecular structures of small molecule drugs were acquired from the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>). Molecular docking was conducted using Auto Dock Tools 1.5.7, and the docking results were ultimately visualized using the PyMOL molecular graphics system.

### Statistical analysis

All data analyses and visualizations were performed using R software (version 4.3.0). For non-normally distributed data with uncertain variance, Wilcoxon rank-sum test was employed to analyze group differences. Cox regression models were utilized for both univariate and multivariate analyses. Survival differences were evaluated using the log-rank test. Bar charts were generated using GraphPad Prism 6.01 software (GraphPad Software, San Diego, CA, USA).

## Results

### Evaluation of key modules in weighted gene co-expression networks (WGCNA) and identification of GRGs in BC

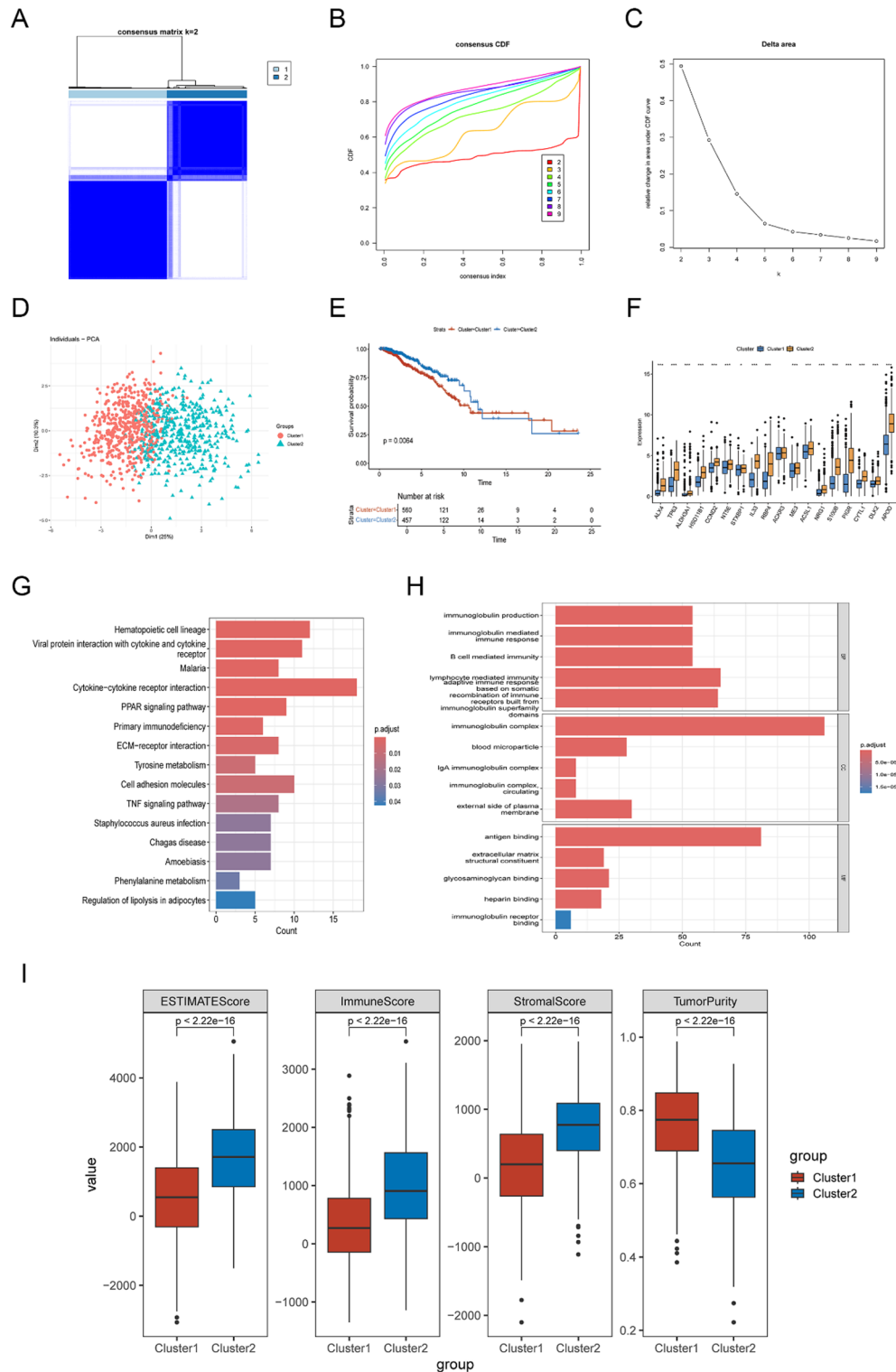
Initially, we identified BC-associated DEGs in the TCGA cohort, determining a total of 2712 DEGs, comprising 989 upregulated and 1723 downregulated genes. The expression profiles of BC-related DEGs are illustrated in a volcano plot, with red and blue representing upregulated and downregulated expression levels, respectively (Fig. 2A). To more comprehensively identify key genes associated with BC phenotype, we conducted WGCNA. An optimal soft threshold (power) of 7 was selected to construct a scale-free topology network (Fig. 2B), ultimately identifying 18 co-expression modules (Fig. 2C). We found that the correlation results between the MEturquoise module and both normal and tumor samples were notably significant ( $cor=0.67$ ,  $p=9e-150$ , Fig. 2D). Furthermore, the scatter plot illustrates a significant correlation ( $cor=0.72$ ,  $p<1e-200$ ) between Gene Significance (GS) for the tumor trait and Module Membership (MM) within the turquoise module (MEturquoise) (Fig. 2E). These results suggest that genes within the MEturquoise module may play crucial roles in BC development and progression. Using a Venn diagram to identify intersection genes among DEGs, MEturquoise module genes, and the glycolysis-related gene set, we discovered a total of 238 overlapping genes (Fig. 2F). The PPI network in Fig. 2G indicates that most proteins encoded by these genes are intricately interconnected in complex patterns.

### Consensus clustering and immune status analysis based on prognosis-related GRGs

To more comprehensively define expression-driven subgroups of prognosis-related GRGs in BC, we performed 1,000 iterations using the "ConsensusClusterPlus" R package (version 1.66.0), with optimal cluster numbers ranging from  $k=2$  to 9. Cumulative distribution curve analysis and the area under this curve indicated that internal clustering consistency reached maximum when  $k$  was set to 2 (Fig. 3A–C). For the consensus matrix heatmaps corresponding to  $k$  values from 3 to 9, please refer to Supplementary Fig. 1. Principal component analysis revealed distinct separation between the two BC subgroups (Fig. 3D). We further compared overall survival between the two clusters, demonstrating that individuals in cluster 2 exhibited superior overall survival compared to those in cluster 1 ( $p=0.0064$ , Fig. 3E). Additionally, we observed significant differences in expression levels of prognosis-related GRGs between the two clusters, with each gene except ACKR3 showing significantly reduced levels in cluster 1 (Fig. 3F). These results suggest that stratification of BC patients based on prognosis-related GRGs is effective. To explore the underlying causes and mechanisms responsible for the significant prognostic differences observed between the two clusters, we conducted KEGG and GO enrichment analyses. KEGG enrichment analysis revealed that prognosis-related GRGs were predominantly enriched in cytokine-cytokine receptor interaction pathways, with numerous enriched genes and significant  $p$ -values (Fig. 3G). Furthermore, substantial enrichment was observed in hematopoietic cell lineage and cell adhesion molecule pathways, suggesting these genes may play crucial roles in immune cell development and function. GO functional enrichment analysis confirmed these findings, with major enrichment in lymphocyte-mediated adaptive immune response, based on somatic recombination of immune receptors built from immunoglobulin superfamily domains, immunoglobulin complex, and antigen binding functions (Fig. 3H). This indicates that prognosis-related GRGs play key roles in adaptive immune responses, particularly humoral immunity. These pathway enrichment results suggest the presence of abnormal immune response activation during BC progression.

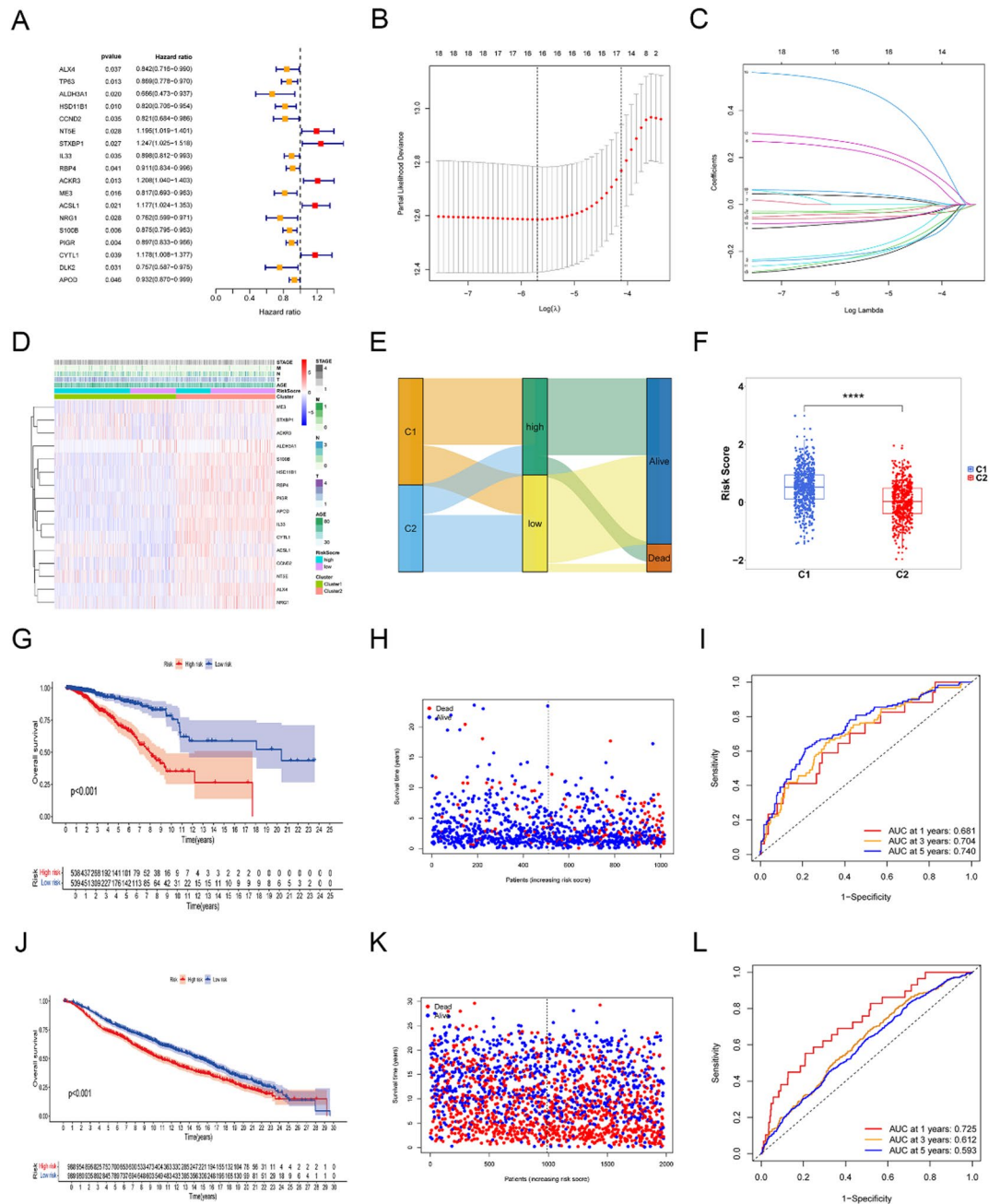
The tumor microenvironment comprises tumor cells, stromal cells, and infiltrating immune cells, which play crucial roles in tumor progression and represent major factors contributing to poor prognosis in cancer patients<sup>22</sup>. To thoroughly explore tumor sample heterogeneity and microenvironmental characteristics, we applied the





**Fig. 3.** Classification and immune microenvironment of GRGs in BC. (A) A heatmap demonstrating clustering is provided. (B) A representation of the cumulative distribution curve is shown. (C) The area curve of the CDF Delta is depicted. (D) Graph of PCA analysis of C1 and C2 clusters. (E) Evaluation of overall survival differences between the clusters. (F) Comparison of GRGs expressions between the clusters. (G, H) KEGG and GO analyses for GRGs. (I) The ESTIMATE score, immune score, stromal status and tumor purity were applied to quantify the different immune statuses between the clusters.

the median risk score. A heatmap further illustrated the relationships between clinical characteristics (stage, T, N, M classifications, age), glycolysis score, clustering subtypes, and the 16 model genes (Fig. 4D). Additionally, we constructed a Sankey diagram depicting the associations among GRG clusters, risk scores, and patient survival status (Fig. 4E). Within the GRG clusters, BC samples in cluster C1 exhibited higher risk scores and poorer clinical prognosis (Fig. 4F). We then utilized TCGA data as the training set and METABRIC as the validation

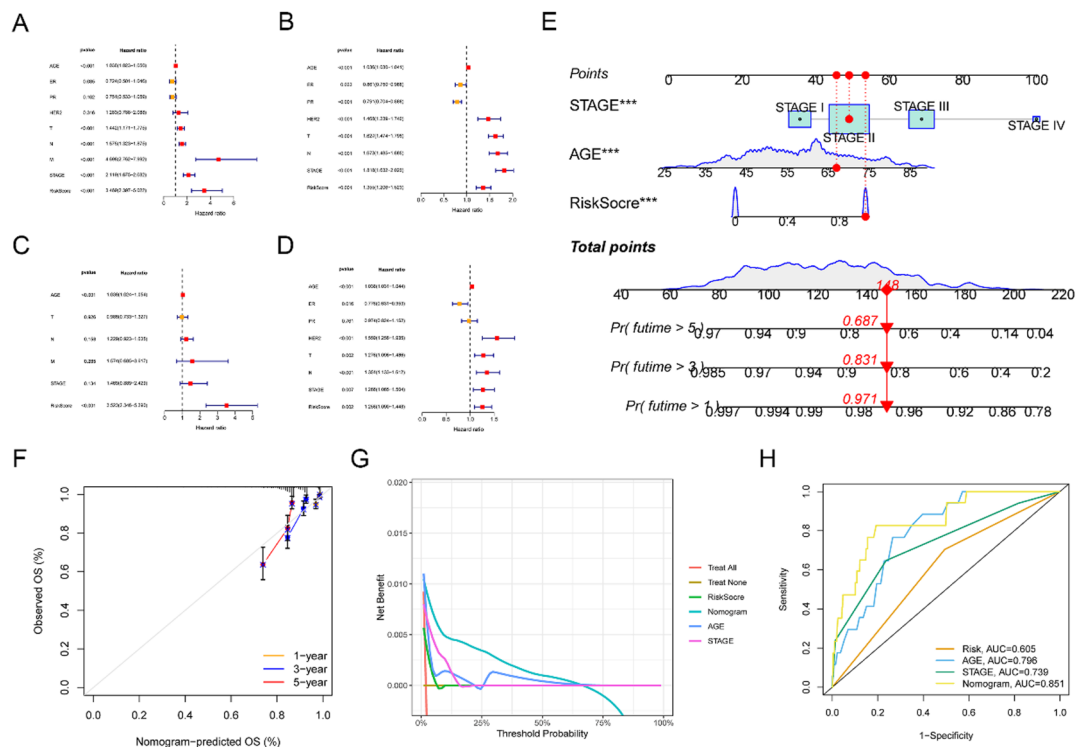


**Fig. 4.** Development and validation of a multi-gene prognostic signature for patient stratification and survival prediction. **(A)** Forest plot of univariate Cox regression analysis identifying prognostic genes. **(B, C)** LASSO regression analysis of selected prognostic genes. **(D)** The distribution of clinical characteristics and the expression of model genes according to the GRGs risk score. **(E)** Sankey diagram showing the relationship between survival status, GRGs clusters, and risk scores. **(F)** Difference-in-difference analysis of cluster risk scores. **(G)** KM curve showing the correlation between Riskscore prediction model and prognosis in the TCGA cohort. **(H)** Survival time status in the TCGA cohort. **(I)** ROC curves for 1-year, 3-year, and 5-year prognoses based on gene prognostic features in the TCGA cohort. **(J)** KM curve showing the correlation between Riskscore prediction model and prognosis in the METABRIC cohort. **(K)** Survival time status in the METABRIC cohort. **(L)** ROC curves for 1-year, 3-year, and 5-year prognoses based on gene prognostic features in the METABRIC cohort. \*\*\*\* $p < 0.0001$ .

set. Kaplan-Meier survival analysis demonstrated that the high-risk group had significantly worse prognosis (Fig. 4G,I). Furthermore, Kaplan-Meier survival analysis revealed significant prognostic differences between BC risk groups stratified by glycolytic scores, with distinct subtype-specific implications. In the Luminal A subtype, a high glycolytic score is associated with markedly reduced survival ( $p < 0.001$ ), reflecting a strong prognostic impact. Similarly, the Luminal B subtype shows a significant survival disadvantage with high glycolytic scores ( $p = 0.005$ ), suggesting a consistent influence across hormone receptor-positive subtypes. In contrast, the HER2-positive subtype exhibits no significant survival difference ( $p = 0.316$ ), indicating that glycolytic activity may have limited prognostic relevance in this group. The basal-like (TNBC) subtype demonstrates a highly significant survival reduction with high glycolytic scores ( $p < 0.001$ ) (Supplementary Fig. 2A–D). Scatter plots of patient survival status also indicated that mortality rates increased with higher risk scores (Fig. 4H,K). Furthermore, we employed ROC curves to evaluate the predictive accuracy of the risk score, with AUC values for 1-year, 3-year, and 5-year risk scores of 0.681, 0.704, 0.740 in the training set and 0.725, 0.612, and 0.593 in the testing set, respectively (Fig. 4L).

### Establishment and evaluation of a nomogram based on clinical characteristics and glycolysis score

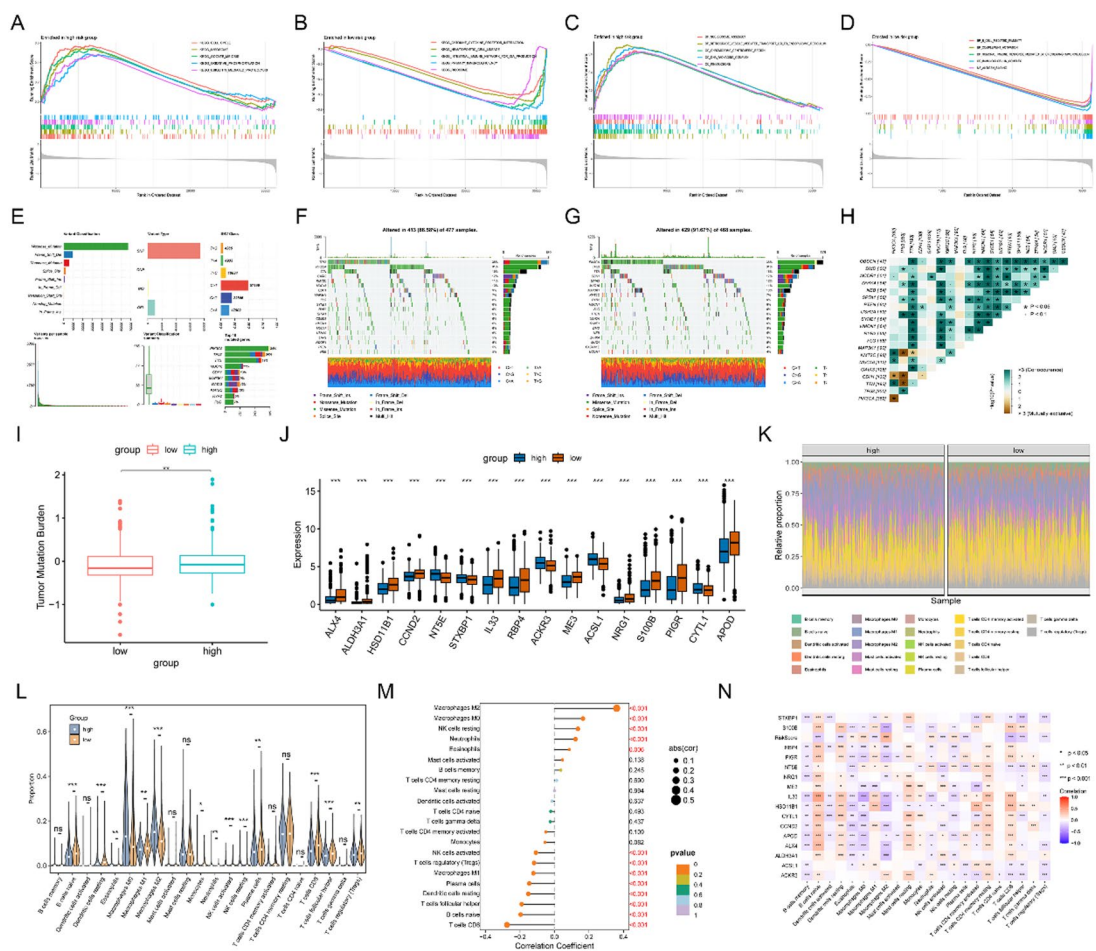
To further validate whether the glycolysis risk score could serve as an independent prognostic factor for BC, we conducted multivariate and univariate Cox analyses, incorporating potential clinical indicators including age, ER, PR, HER2, T, N, M classifications, and STAGE. As shown in Fig. 5A–D, multivariate Cox regression analysis further confirmed that the risk score was an independent risk factor for BC patients. To facilitate personalized assessment for each patient, we constructed a nomogram based on the glycolysis risk score and clinical characteristics (Fig. 5E). Calibration curves demonstrated good concordance between nomogram predictions and actual observations (Fig. 5F). Decision curve analysis (DCA) indicated that the predictive model possessed favorable clinical utility (Fig. 5G). Furthermore, ROC curve analysis revealed that our nomogram performed well in discriminating outcomes, with an AUC value (0.851) significantly higher than the predictive capability of using AGE (AUC=0.796) or STAGE (AUC=0.739) alone. These findings suggest that the GRG-based nomogram provides a reliable and accurate tool for personalized prognostic prediction in BC patients. The temporal consistency and robustness of our nomogram’s predictive performance across different time intervals are further validated in Fig. 5H, demonstrating sustained accuracy in prognostic assessment (Fig. 5H).



**Fig. 5.** Development and validation of the nomogram. (A, B) The univariate and multivariate Cox regression analyses in the TCGA cohort. (C, D) The univariate and multivariate Cox regression analyses in the METABRIC cohort. (E, C) The nomogram by combining glycolysis score with age and stage for predicting the 1-, 3-, and 5-year survival probability of patients with BC. (F) The calibration curves of the nomogram for predicting overall survival (OS) probability for 1-, 3-, and 5-years OS probabilities. (G) Decision curve analysis (DCA) of the nomogram. (H) Receiver operating characteristic (ROC) curves of the nomogram.

## Prediction of biological mechanisms associated with GRGs signature

To explore the molecular mechanisms underlying transcriptomic and genetic differences between high and low risk groups, and to gain deeper insights into the biological basis of poor prognosis in the high-risk group, we conducted GRGs model-related genomic heterogeneity analysis in the TCGA cohort. First, we performed GSEA analysis. In KEGG gene set-based GSEA, the high-risk group showed enrichment in “CELL\_CYCLE,” “LYSOSOME,” “OOCYTE\_MEIOSIS,” “OXIDATIVE\_PHOSPHORYLATION,” and “OOCYTE\_MEIOSIS” pathways (Fig. 6A). In contrast, the low-risk group exhibited enrichment in “CYTOKINE\_CYTOKINE\_RECEPTOR\_INTERACTION,” “RIBOSOME,” “HEMATOPOIETIC\_CELL\_LINEAGE,” “INTESTINAL\_IMMUNE\_NETWORK\_FOR\_IGA\_PRODUCTION,” and “PRIMARY\_IMMUNODEFICIENCY” pathways (Fig. 6B, Supplementary Table 4). Furthermore, in GO gene set-based GSEA analysis, the high-risk group demonstrated enrichment in “BP\_RETROGRADE\_VESICLE\_MEDIATED\_TRANSPORT\_LGI\_TO\_ENDOPLASMIC\_RETICULUM,” “CC\_CHROMOSOME\_CENTROMERIC\_REGION,” “CC\_DNA\_PACKAGING\_COMPLEX,” “BP\_NUCLEOSOME\_ASSEMBLY,” and “CC\_KINETOCHORE” (Fig. 6C). The low-risk group, however, showed enrichment in “CC\_IMMUNOGLOBULIN\_COMPLEX,” “BP\_HUMORAL\_IMMUNE\_RESPONSE\_MEDIATED\_BY\_CIRCULATING\_IMMUNOGLOBULIN,” “BP\_COMPLEMENT\_ACTIVATION,” “MF\_ANTIGEN\_BINDING,” and “BP\_B\_CELL\_MEDIATED\_IMMUNITY” pathways (Fig. 6D, Supplementary Table 5). This suggests that patients with different risk scores may possess distinct immune states. To further explore molecular mechanisms, we conducted analyses based on molecular subtyping.



**Fig. 6.** Prediction of biological mechanisms associated with GRGs risk. **(A)** Identification of KEGG terms enriched in the high-risk group through GSEA analysis. **(B)** Identification of KEGG terms enriched in the low-risk group through GSEA analysis. **(C)** Identification of GO terms enriched in the high-risk group through GSEA analysis. **(D)** Identification of GO terms enriched in the low-risk group through GSEA analysis. **(E)** Summary of somatic mutations in the TCGA cohort. **(F)** The waterfall plot of the somatic mutation landscape in high-risk patients in the TCGA cohort. **(G)** The waterfall plot of the somatic mutation landscape in low-risk patients in the TCGA cohort. **(H)** Heatmaps showing the association of co-occurrence and exclusive mutation among the top 20 mutated genes. **(I)** Tumor mutation burden (TMB) between different groups. **(J)** Expression differences of 16 GRGs between high and low-risk groups. **(K)** The heatmap displaying immune infiltration of different subgroups. **(L)** Comparison of immune infiltration between high and low groups. **(M)** Correlation coefficients between immune cells and glycolysis risk score. **(N)** Correlation analysis of model genes with immune cells. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

Supplementary Fig. 2E–H sequentially illustrate the glycolysis activity profiles and gene set enrichment analysis (GSEA) for LumA, LumB, HER2, and Basal subtypes, elucidating their distinct molecular mechanisms in glycolysis risk scoring and metabolic-immune regulation. In the LumA subtype, the high-risk group was significantly enriched in “CELL\_CYCLE”-related pathways, including HALLMARK\_E2F\_TARGETS, HALLMARK\_G2M\_CHECKPOINT, HALLMARK\_MYC\_TARGETS\_V1, HALLMARK\_MYC\_TARGETS\_V2, and HALLMARK\_DNA\_REPAIR, suggesting that metabolic reprogramming may support limited proliferative activity through the activation of key cell cycle regulators (e.g., E2F and MYC) and DNA repair mechanisms. In contrast, the high-risk group of the LumB subtype was enriched in HALLMARK\_GLYCOLYSIS, HALLMARK\_INTERFERON\_ALPHA\_RESPONSE, HALLMARK\_INTERFERON\_GAMMA\_RESPONSE, HALLMARK\_MTORC1\_SIGNALING, and HALLMARK\_OXIDATIVE\_PHOSPHORYLATION pathways, indicating pronounced metabolic heterogeneity characterized by enhanced glycolysis, immune-inflammatory responses, and oxidative metabolism. The HER2 subtype exhibited an enrichment pattern similar to LumA, primarily involving “CELL\_CYCLE”-related pathways such as HALLMARK\_E2F\_TARGETS, HALLMARK\_G2M\_CHECKPOINT, HALLMARK\_MYC\_TARGETS\_V1, HALLMARK\_MTORC1\_SIGNALING, and HALLMARK\_PROTEIN\_SECRETION, suggesting that metabolic reprogramming drives rapid proliferation and tumor progression through accelerated cell cycle activity and protein synthesis regulation. Notably, the low-risk group of the Basal subtype displayed an enrichment pattern similar to the high-risk groups of LumA and HER2 subtypes, with significant enrichment in “CELL\_CYCLE”-related pathways, including HALLMARK\_E2F\_TARGETS, HALLMARK\_G2M\_CHECKPOINT, HALLMARK\_MYC\_TARGETS\_V1, HALLMARK\_MTORC1\_SIGNALING, and HALLMARK\_MITOTIC\_SPINDLE. This pattern, contrary to typical high-risk group characteristics, indicates that the Basal subtype may sustain proliferative potential through cell cycle-related pathways even under low glycolysis conditions.

We further conducted gene mutation analysis, which revealed missense mutation as the predominant mutation classification and single nucleotide polymorphism as the primary variant type (Fig. 6E). To investigate genomic mutation differences between GRG subgroups, we delineated mutation profiles between high-risk and low-risk groups. Figure 6F, H display the 20 most common mutations identified in high-risk and low-risk populations, with TP53 and PIK3CA showing the highest mutation frequencies in high-risk and low-risk groups, respectively (Fig. 6F, H). Additionally, tumor mutation burden (TMB) analysis demonstrated elevated TMB in the high-risk group compared to the low-risk group (Fig. 6I). We further examined the expression patterns of the 16 GRGs in high-risk and low-risk groups, finding that except for NT5E, STXBP1, ACKR3, ACSL1, and CYTL1, the remaining GRGs were downregulated in the high-risk population (Fig. 6J).

In BC, the tumor immune microenvironment is influenced by glycolysis scores, affecting tumor proliferation and dissemination<sup>23</sup>. We utilized the “CIBERSORT” algorithm to calculate infiltration levels of 22 immune cell types. Both heatmap and violin plots illustrated the immune infiltration landscape across high and low-risk groups (Fig. 6K–M). The high-risk group exhibited increased numbers of Eosinophils, Macrophage M0, Macrophage M2, Neutrophils, NK cells resting, while the low-risk group displayed higher levels of B cells naïve, Dendritic cells resting, Macrophages M1, Monocytes, NK cells activated, Plasma cells, T cells CD8, T cells follicular helper, T cells regulatory (Tregs) (Fig. 6L, M). To further investigate the immune infiltration profiles across different molecular subtypes of breast cancer, we analyzed the immune infiltration results for LumA, LumB, HER2, and Basal subtypes from left to right, revealing distinct heterogeneous patterns (Supplementary Fig. 2I–L). In the LumA subtype, the high-risk group exhibited significantly elevated proportions of Macrophage M0, Macrophage M2, and resting NK cells, whereas the low-risk group showed higher levels of naïve B cells, resting dendritic cells, plasma cells, CD8 + T cells, and follicular helper T cells. For the LumB subtype, the high-risk group demonstrated notable increases in Macrophage M2 and neutrophils, while the low-risk group was characterized by elevated levels of naïve B cells, CD8 + T cells, and follicular helper T cells. In the HER2 subtype, Macrophage M2 levels were significantly higher in the high-risk group, contrasted by increased proportions of memory B cells, activated memory CD4 + T cells, CD8 + T cells, and regulatory T cells (Tregs) in the low-risk group. In the Basal subtype, the high-risk group displayed significant elevations in Macrophage M0, Macrophage M2, and resting memory CD4 + T cells, whereas the low-risk group showed higher levels of Macrophage M1, activated memory CD4 + T cells, CD8 + T cells, follicular helper T cells, and regulatory T cells (Tregs). These findings highlight subtype-specific immune infiltration patterns that may reflect underlying metabolic and immunological regulatory mechanisms. Furthermore, we demonstrated the correlations between the 16 genes involved in constructing the risk model and immune cells (Fig. 6N).

### Heterogeneity of glycolysis score status in the immune microenvironment at single-cell transcriptome level

We analyzed single-cell RNA transcriptome data from BC patients to explore the relationship between glycolysis score status and immune cells. Using established markers, we annotated eight cell clusters. The bubble plot illustrates the expression levels of cell type-specific marker genes as follows: Epithelials (EPCAM, KRT19, KRT8, KRT17), Fibroblasts (COL1A1, COL3A1), T cells (IL7R, CD3D, CD2), Myeloid cells (CD68, CSF1R, LYZ), Endothelials (VWF, PECAM1, CDH5), B cells (CD79A, MS4A1, CD19), Pericytes (RGS5, MCAM, MYH11), and Mast cells (GATA2, KIT, CPA3) (Fig. 7A, B). We assessed the proportion of each cell subtype in each sample, revealing significant differences between tumor tissue and normal tissue (Fig. 7C). To enhance our understanding, we utilized the “AddModuleScore” function in Seurat to calculate the scores of 16 risk-associated genes across various cell types, revealing significantly elevated glycolysis scores in myeloid cells and T cells within tumor tissue compared to their counterparts in normal tissue (Fig. 7E, F). Breast epithelial cells represent a primary driver of intratumoral heterogeneity and malignancy in breast cancer. To dissect this heterogeneity, we performed dimensionality reduction, clustering, and subpopulation identification on epithelial cells isolated from breast cancer samples, resolving them into 13 distinct subgroups. Given the clinical characteristics of



breast cancer, we conducted a detailed glycolysis score analysis at the single-cell level across distinct molecular subtypes, with cells color-coded by cell type and glycolysis score (UP indicating high and DOWN indicating low) (Supplementary Fig. 2M,N). This analysis, integrated with the provided figure, reveals diverse clustering patterns among the LumA, LumB, HER2, and Basal subtypes (Supplementary Fig. 2M,N). Our results further indicate that the prognostic model does not influence the overall strength and quantity of interactions between myeloid cells and T cells with other cell types but may regulate the immune microenvironment by modulating the biological functions of specific signaling pathways in myeloid cells (e.g., MHC-II, MIF, SPP1) and T cells (e.g., MHC-I, CCL, CXCL), thereby promoting distinct immune responses in breast cancer. This is supported by the elevated glycolysis scores in tumor tissues (Fig. 7E,F) and distinct pathway activities (Fig. 8D–K, Supplementary Fig. 4B–I). Additionally, we further demonstrated the distribution patterns of the 16 glycolysis-related genes used for model construction across different cell types (Supplementary Fig. 3B–Q).

### The correlation of the GRGs with single-cell characteristics

First, we analyzed the quantity and strength of cellular communication between myeloid cells with different GRGs risk scores and other cell types. Both high and low glycolysis myeloid cells established extensive communication connections with various cell types, yet they exhibited similar overall communication strength and quantity (Fig. 8A,B). Similarly, when we applied GRGs risk scoring to T cells, the results paralleled those of myeloid cells (Supplementary Fig. 4A). This suggests that glycolytic metabolic reprogramming may not directly alter the overall communication strength or specific communication preferences of myeloid cells and T cells, but rather might regulate downstream responses through different signal transduction mechanisms. Among these interactions, high and low glycolysis myeloid cells contributed most significantly to incoming interactions, while fibroblasts had the greatest impact on outgoing interactions (Fig. 8C).

We utilized CellChat to explore the communication characteristics between myeloid cells, T cells, and other cell types. Results indicated that in myeloid cells, ligand-receptor-mediated cellular interactions primarily occurred through MHC-II, MIF, and SPP1 signaling pathways (Fig. 8D–I). The quantitative assessment of ligand-receptor interaction strengths reveals distinct communication patterns, with Fig. 8J,K providing detailed analysis of the interaction weights between high and low glycolysis myeloid cells, respectively, highlighting the differential intercellular communication networks under varying glycolytic states (Fig. 8J,K). However, in T cells, ligand-receptor-mediated cellular interactions predominantly existed in MHC-I, CCL, and CXCL signaling pathways (Supplementary Fig. 4B–I).

For more in-depth investigation, we carefully identified key senders, receivers, mediators, and influencers in the cellular signaling networks. Our findings revealed that high glycolysis myeloid cells acted as stronger receivers and influencers in the MHC-II signaling pathway, whereas low glycolysis myeloid cells played prominent roles as senders, mediators, and influencers (Fig. 8G). In MIF signaling, epithelial cells functioned as the primary senders, while low glycolysis myeloid cells served as both receivers and mediators, and together with high glycolysis myeloid cells, played roles as influencers (Fig. 8H). In SPP1 signaling, high glycolysis myeloid cells were the primary senders, fibroblasts and pericytes acted as receivers, and both high and low glycolysis myeloid cells mainly functioned as mediators and influencers (Fig. 8I).

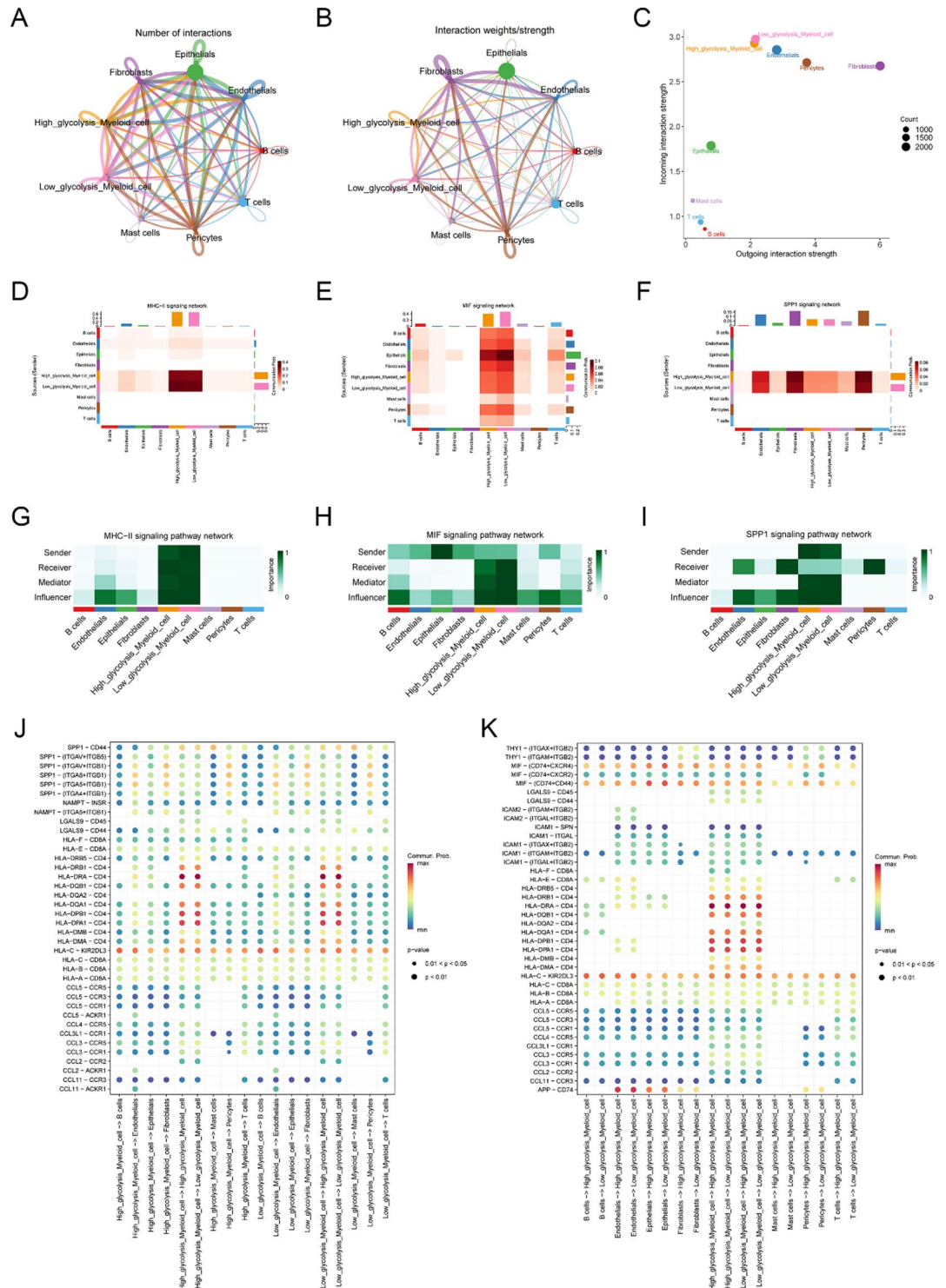
Additionally, we found that high and low glycolysis T cells primarily served as influencers in the MHC-I signaling pathway (Supplementary Fig. 4F). High glycolysis T cells played more extensive roles in the CCL signaling pathway, where myeloid cells acted as senders, high glycolysis T cells functioned as receivers and mediators, and endothelial cells, high glycolysis T cells, low glycolysis T cells, and myeloid cells all could serve as influencers (Supplementary Fig. 4G). However, low glycolysis T cells were more active than high glycolysis T cells in the CXCL signaling pathway, functioning as senders, receivers, mediators, and influencers (Supplementary Fig. 4H). This indicates that glycolytic metabolic reprogramming may fine-tune the functional roles of immune cells in the tumor microenvironment by regulating specific signaling pathways, providing new insights into the metabolic regulatory mechanisms of immune responses.

### Identification of hub genes in BC by MR and search for potential therapeutic agents

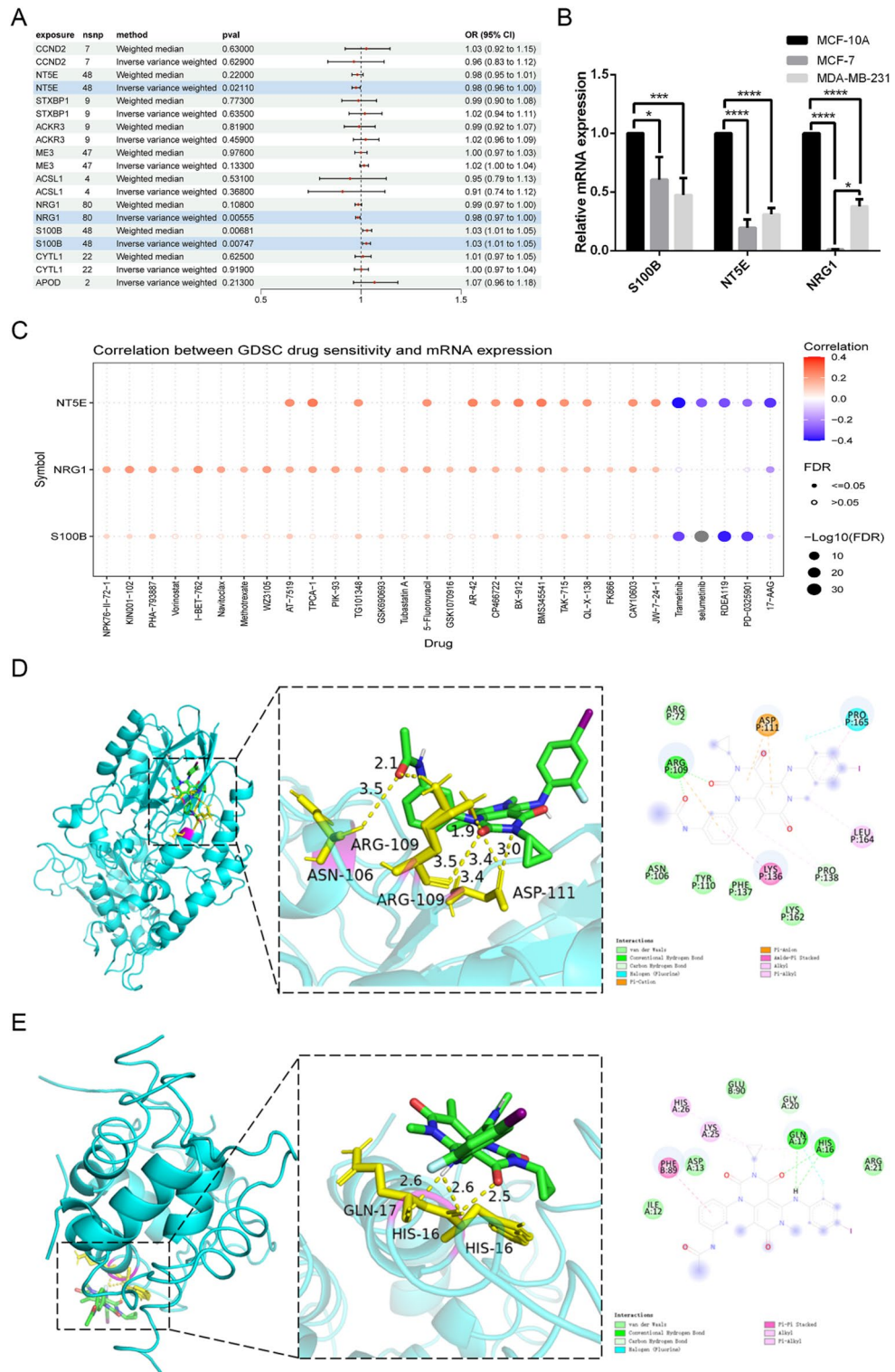
MR analysis of the 16 glycolysis-related prognostic genes in our risk model revealed that NT5E (OR = 0.98, 95% CI = 0.96–1.00,  $P = 0.02110$ ) and NRG1 (OR = 0.98, 95% CI = 0.97–1.00,  $P = 0.00555$ ) were identified as protective factors with decreased expression in breast cancer (BC) tissues, while S100B (OR = 1.03, 95% CI = 1.01–1.05,  $P = 0.00747$ ) was determined to be a risk factor with increased expression in BC tissues (Fig. 9A). We further evaluated the expression of these three MR-screened glycolysis-related genes in cell lines, including one normal cell line (MCF10A) and two BC cell lines (MCF7 and MDA-MB-231). Results demonstrated that NT5E, NRG1, and S100B were significantly downregulated in BC cell lines (Fig. 9B).

To explore the value of GRGs in personalized and precision therapy for BC, we assessed the maximum inhibitory concentration (IC<sub>50</sub>) of various drugs from the GDSC database between the two risk groups. Results indicated that, compared to the low-risk group, the high-risk group exhibited lower sensitivity to most drugs but higher sensitivity to AZD8055, suggesting potential benefit from AZD8055 treatment in the high-risk group (Supplementary Fig. 5). Additionally, through Gene Set Cancer Analysis (GSCA) comprehensive online analysis, we revealed correlations between NT5E, NRG1, and S100B with various drugs. Results showed significant positive correlations between NT5E and S100B with Trametinib, with correlation coefficients (cor) of  $-0.39188$  and  $-0.32788$ , respectively (Fig. 9C, Supplementary Table 6).

Consequently, we conducted molecular docking to investigate protein-ligand binding patterns between NT5E and S100B with Trametinib. Protein structures were downloaded from the UniProt database, and molecular docking was performed using AutoDock. Binding stability assessment was based on binding energy, where values less than  $-5$  kcal/mol were considered indicative of significant binding interactions, and values less than  $-7$  kcal/mol indicated strong binding interactions<sup>24</sup>. In this study, molecular docking models of NT5E



**Fig. 8.** The correlation of GRGs with single-cell characteristics in myeloid cells. **(A)** Communication interactions network plot for all cell types. **(B)** Communication interaction weights network plot for all cell types. **(C)** Identification of the signals that contribute the most to the efferent and afferent signals between cell types. **(D)** Cell-cell communication interaction in MHC-II signaling pathway. **(E)** Cell-cell communication interaction in MIF signaling pathway. **(F)** Cell-cell communication interaction in SPP1 signaling pathway. **(G)** The role of high and low glycolysis myeloid cells in the MHC-II signaling pathway. **(H)** The role of high and low glycolysis myeloid cells in the MIF signaling pathway. **(I)** The role of high and low glycolysis myeloid cells in the SPP1 signaling pathway. **(J, K)** The receptor-ligand communication weights in high and low glycolysis myeloid cells.



**Fig. 9.** Identification of hub genes in BC by MR and search for potential therapeutic agents. **(A)** Forest plot for MR results, with the x-axis representing Odds Ratios (OR) and the vertical line representing the null effect line (OR = 1). Red dots represent the OR values for each analysis method, with the horizontal line indicating the 95% confidence interval. **(B)** The relative expression levels of NT5E, S100B and NRG1 genes in normal breast epithelial cell line (MCF10A), breast cancer cell line (MCF7 and MDA - MB - 231). **(C)** Analysis of drug susceptibility of NT5E, S100B and NRG1 performed online by GSCA. **(D)** Molecular docking between Trametinib and NT5E. **(E)** Molecular docking between Trametinib and S100B.

and S100B with Trametinib are shown in Fig. 9D,E, with docking scores of  $-7.76$  kcal/mol and  $-7.27$  kcal/mol, respectively, substantially exceeding the threshold for strong binding interactions. To facilitate observation of intermolecular interactions, we further illustrated the types of interactions and their distances between the drug molecule Trametinib and NT5E and S100B proteins in two-dimensional diagrams (Fig. 9D,E). This finding suggests that Trametinib may have potential therapeutic applications in BC treatment.

Furthermore, we analyzed the binding of NT5E and S100B with AZD8055. Results indicated that AZD8055 could form stable hydrogen bonds with both proteins, with binding energies of  $-6.27$  kcal/mol and  $-6.22$  kcal/mol, respectively (Supplementary Fig. 6), demonstrating high stability of interactions between AZD8055 and proteins encoded by GRG-based genes. Collectively, these findings suggest that Trametinib and AZD8055 could be considered as GRG-related alternative therapeutic options.

## Discussion

Breast cancer (BC), as the most common malignancy among women globally, remains a leading cause of cancer-related mortality despite advances in diagnostic and therapeutic techniques<sup>25</sup>. Due to its significant molecular heterogeneity, existing biomarkers have limited ability to predict recurrence risk and treatment response, making it challenging to guide individualized treatment decisions<sup>26</sup>. Metabolic reprogramming, particularly enhanced glycolytic activity, has been identified as one of the key characteristics of BC. Research indicates that glycolytic key enzymes (such as HK2, PFK, and PKM2) are significantly upregulated in BC tissue, not only promoting tumor proliferation and invasion but also regulating the tumor microenvironment and immune evasion processes through intermediate metabolites<sup>27</sup>. However, the potential of glycolysis-related genes (GRGs) as prognostic markers for BC has not been fully explored. This study integrates multi-omics data to construct a prognostic model based on 16 glycolysis-related genes (GRGs), demonstrating strong predictive performance across TCGA and METABRIC cohorts. The model effectively stratifies patients into high- and low-risk groups, revealing subtype-specific differences in survival outcomes, gene set enrichment analysis (GSEA), and immune infiltration patterns across HR/HER2 molecular subtypes (Luminal A, Luminal B, HER2-positive, and Basal). Through Mendelian randomization (MR), we established causal links between NT5E, NRG1, and S100B and BC risk, while molecular docking identified trametinib and AZD8055 as potential therapeutic agents. These findings deepen our understanding of glycolytic reprogramming and its interplay with the immune microenvironment, providing a foundation for precision medicine in BC.

Our prognostic model, constructed using 16 GRGs, exhibited strong predictive accuracy in both TCGA and METABRIC cohorts, with a nomogram integrating clinical features achieving an AUC of 0.851, surpassing traditional indicators such as age (AUC = 0.796) or stage (AUC = 0.739). Kaplan-Meier survival analysis revealed significant prognostic differences across HR/HER2 molecular subtypes (Supplementary Fig. 2A-D). In Luminal A and Luminal B subtypes, high glycolytic scores were associated with significantly reduced survival ( $p < 0.001$  and  $p = 0.005$ , respectively), underscoring the prognostic impact of glycolysis in hormone receptor-positive BC. These findings align with prior studies linking elevated glycolysis to aggressive tumor behavior and chemoresistance in luminal subtypes<sup>8</sup>. In contrast, the HER2-positive subtype showed no significant survival difference ( $p = 0.316$ ), suggesting that glycolytic activity may have limited prognostic relevance in this group, likely due to the dominance of HER2-driven signaling pathways. The Basal (triple-negative) subtype exhibited a pronounced survival reduction with high glycolytic scores ( $p < 0.001$ ), reflecting its aggressive nature and reliance on metabolic reprogramming for rapid proliferation. These subtype-specific survival patterns highlight the necessity of tailored prognostic models that account for molecular heterogeneity, enhancing the precision of risk stratification and treatment planning.

Gene set enrichment analysis (GSEA) elucidated distinct molecular mechanisms underlying glycolytic reprogramming across HR/HER2 subtypes (Supplementary Fig. 2E-H). We observed that in ER-positive/HER2-negative breast cancer with high glycolysis scores, multiple pro-cancerous Hallmark gene sets (including oxidative phosphorylation, mTORC1 signaling, and others) as well as all four cell proliferation-related gene sets (HALLMARK\_E2F\_TARGETS, HALLMARK\_G2M\_CHECKPOINT, HALLMARK\_MYC\_TARGETS\_V1, and HALLMARK\_MYC\_TARGETS\_V2) were significantly enriched. This finding indicates that glycolysis in ER-positive/HER2-negative breast cancer supports tumor aggressiveness by enhancing cell proliferation and pro-cancerous signaling pathways. In contrast, in the low-glycolysis group of triple-negative breast cancer (TNBC), cell proliferation-related gene sets were also significantly enriched, suggesting that TNBC maintains proliferative potential through alternative cell cycle regulation mechanisms even under low glycolysis conditions. This result is consistent with the findings of Oshi et al.<sup>28</sup>, further underscoring the unique metabolic adaptability of TNBC and providing critical insights for the development of subtype-specific therapeutic strategies.

Immune infiltration analysis reveals that glycolytic activity significantly modulates the breast cancer (BC) immune microenvironment, with distinct patterns observed across HR/HER2 molecular subtypes (Supplementary Fig. 2I-L). In the high-risk group, characterized by elevated glycolytic scores, there is a pronounced enrichment of M2 macrophages, which promote immunosuppression and tumor progression through STAT3 and HIF-1 $\alpha$  signaling pathways<sup>29,30</sup>. This immunosuppressive milieu is particularly evident in Luminal A and Basal subtypes, where high glycolytic activity correlates with increased M2 macrophage infiltration, fostering tumor immune evasion and aggressive disease behavior<sup>31</sup>. Conversely, the low-risk group, marked by reduced glycolytic activity, is enriched with M1 macrophages, CD8 + T cells, and follicular helper T cells, supporting robust anti-tumor immunity<sup>32</sup>. This is especially prominent in Luminal A and Basal subtypes, where low glycolytic scores are associated with enhanced effector T cell function, aligning with Li et al.'s<sup>31</sup> findings that reduced glycolysis enhances T cell-mediated anti-tumor responses and immunotherapy efficacy. Notably, regulatory T cells (Tregs) in the low-risk group exhibit diminished immunosuppressive effects in a less glycolytic microenvironment, as reported by Hashemi et al.<sup>33</sup>, particularly in Luminal B and HER2-positive subtypes, where Tregs show reduced activity in low-risk settings. Subtype-specific immune infiltration patterns further highlight the interplay between

glycolysis and immune regulation. In Luminal A, the high-risk group shows elevated levels of Macrophage M0, M2 macrophages, and resting NK cells, contributing to an immunosuppressive TME that supports tumor progression. In contrast, the low-risk group is characterized by higher proportions of naïve B cells, CD8 + T cells, and follicular helper T cells, fostering an immune-active environment conducive to better prognosis<sup>32</sup>. Luminal B exhibits a similar trend, with the high-risk group enriched in M2 macrophages and neutrophils, promoting immunosuppression, while the low-risk group displays increased CD8 + T cells and follicular helper T cells, indicative of a balanced immune response<sup>4</sup>. In the HER2-positive subtype, the high-risk group is dominated by M2 macrophages, whereas the low-risk group shows elevated memory B cells, CD8 + T cells, and Tregs, suggesting a complex immune dynamic influenced by HER2 signaling<sup>34</sup>. The Basal subtype presents a distinct profile, with the high-risk group enriched in Macrophage M0, M2 macrophages, and resting memory CD4 + T cells, while the low-risk group is characterized by higher levels of M1 macrophages, CD8 + T cells, and Tregs, reflecting a metabolically driven immune landscape<sup>31</sup>. These subtype-specific patterns underscore the role of glycolysis in shaping immune responses, with high glycolytic activity driving immunosuppression via M2 macrophage polarization and low glycolytic activity promoting anti-tumor immunity through enhanced effector immune cell infiltration. Single-cell RNA sequencing further corroborates these findings, demonstrating elevated glycolytic scores in myeloid and T cells within tumor tissues compared to normal tissues. Subtype-specific signaling pathways, such as MHC-II, MIF, and SPP1 in myeloid cells and MHC-I, CCL, and CXCL in T cells, modulate immune interactions and contribute to distinct immune microenvironments (Fig. 8D–K, Supplementary Figs. 4B–I). For instance, SPP1 signaling in myeloid cells, highly active in Basal subtype high-risk groups, promotes immunosuppression via STAT3 activation, as noted by Behera et al.<sup>35</sup>. Conversely, CXCL-mediated T cell interactions in low-risk Basal and Luminal A subtypes enhance anti-tumor immunity, consistent with Wang et al.'s observations<sup>36</sup>. These findings emphasize that lower glycolytic activity fosters an immune-active TME, correlating with improved clinical outcomes, while high glycolytic activity drives immunosuppression, particularly in aggressive subtypes like Basal BC. The differential immune infiltration profiles across HR/HER2 subtypes highlight the need for subtype-specific immunotherapeutic strategies, with low-risk patients potentially benefiting from immune checkpoint inhibitors and high-risk patients requiring combined metabolic and immune-targeted therapies to overcome immunosuppression<sup>37</sup>.

Mendelian randomization analysis identified NT5E and NRG1 as protective factors (OR = 0.98,  $P = 0.02110$  and  $P = 0.00555$ , respectively) and S100B as a risk factor (OR = 1.03,  $P = 0.00747$ ) for BC, validated by qRT-PCR in MCF7 and MDA-MB-231 cell lines compared to MCF10A. NT5E's dual role, with low expression in luminal subtypes due to estrogen-mediated suppression, suggests subtype-specific functions<sup>38,39</sup>. NRG1's protective effects, linked to reduced glycolysis dependence via AMPK, highlight its potential as a therapeutic target<sup>40</sup>. S100B's oncogenic role, promoting glycolysis and tumor progression via RAGE signaling, aligns with its association with poor prognosis in triple-negative BC<sup>41,42</sup>. Molecular docking identified trametinib and AZD8055 as potential therapeutics, with strong binding affinities to NT5E and S100B (−7.76 and −7.27 kcal/mol for trametinib; −6.27 and −6.22 kcal/mol for AZD8055). Trametinib, a MAPK inhibitor, may counteract S100B-driven metabolic reprogramming, while AZD8055, an mTOR inhibitor, could modulate NT5E and NRG1 functions via the PI3K/AKT/mTOR pathway<sup>43–47</sup>. These findings suggest that targeting glycolysis-related pathways in a subtype-specific manner could enhance therapeutic efficacy, particularly in combination with immune checkpoint inhibitors for high-risk patients<sup>48</sup>.

The subtype-specific survival, GSEA, and immune infiltration patterns underscore the clinical utility of our GRG-based prognostic model. Low-risk patients, particularly in Luminal A and Basal subtypes, may benefit from immune checkpoint inhibitor monotherapy due to their immune-active microenvironment. High-risk patients, especially in Luminal B and Basal subtypes, may require combined metabolic and immunotherapies to overcome immunosuppression driven by M2 macrophages and high glycolytic activity. The identification of trametinib and AZD8055 as potential therapeutics targeting NT5E and S100B provides a foundation for precision therapies, particularly for aggressive subtypes like Basal BC. In conclusion, this study confirms the crucial role of GRGs in shaping the BC immune microenvironment, providing a theoretical foundation for precision therapeutic strategies based on the metabolism-immune axis.

Despite constructing an effective prognostic prediction model through multi-omics data integration and machine learning algorithms, and exploring potential therapeutic targets, our study has several limitations. First, although we validated the model's predictive capability in two independent cohorts, it lacks confirmation from large-scale prospective clinical trials, which may limit its generalizability. Second, while we established causal relationships between NT5E, NRG1, and S100B and breast cancer risk, and predicted potential drug targets through molecular docking analysis, these findings require further verification through *in vivo* and *in vitro* experiments. Third, our PCR analysis was limited to a restricted set of cell lines, failing to comprehensively reflect the expression profile of S100B across different molecular subtypes. The dynamic changes in the glycolysis regulatory network and its differential roles across various molecular subtypes also require further exploration, necessitating more comprehensive research to reveal its complexity.

## Conclusions

In summary, this study constructed a BC prognostic prediction model based on GRGs by integrating multi-omics data, revealed the causal relationships between NT5E, NRG1, and S100B and BC prognosis, and elucidated the close connection between glycolytic activity and immune microenvironment remodeling. These findings expand our understanding of metabolic reprogramming mechanisms in BC and provide a theoretical foundation for precision stratification management and individualized treatment strategies based on the metabolism-immune axis, potentially improving the clinical prognosis of BC patients.

## Data availability

The datasets of this article were generated from the TCGA database (<https://portal.gdc.cancer.gov/>) (<https://portal.gdc.cancer.gov/>), GEO database (<https://www.ncbi.nlm.nih.gov/geo/>) (<https://www.ncbi.nlm.nih.gov/geo/>) and METABRIC database, which was downloaded from the cBioPortal database (<http://www.cbioportal.org/>) (<http://www.cbioportal.org/>)

Received: 2 June 2025; Accepted: 17 November 2025

Published online: 06 March 2026

## References

- Cai, Y., Dai, F., Ye, Y. & Qian, J. The global burden of breast cancer among women of reproductive age: a comprehensive analysis. *Sci. Rep.* **15**, 9347. <https://doi.org/10.1038/s41598-025-93883-9> (2025).
- Bray, F. et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **74**, 229–263. <https://doi.org/10.3322/caac.21834> (2024).
- Xiong, X. et al. Breast cancer: pathogenesis and treatments. *Signal. Transduct. Target. Ther.* <https://doi.org/10.1038/s41392-024-02108-4> (2025).
- Carvalho, E., Canberk, S., Schmitt, F. & Vale, N. Molecular subtypes and mechanisms of breast cancer: precision medicine approaches for targeted therapies. *Cancers (Basel)* <https://doi.org/10.3390/cancers17071102> (2025).
- Vander Heiden, M. G. & DeBerardinis, R. J. Understanding the intersections between metabolism and cancer biology. *Cell* **168**, 657–669. <https://doi.org/10.1016/j.cell.2016.12.039> (2017).
- Mitaishvili, E. et al. The molecular mechanisms behind advanced breast cancer metabolism: Warburg Effect, OXPHOS, and calcium. *Front. Biosci. (Landmark Ed.)* **29**, 99. <https://doi.org/10.31083/j.fb12903099> (2024).
- Yue, S. W. et al. m6A-regulated tumor glycolysis: new advances in epigenetics and metabolism. *Mol. Cancer* **22**, 137. <https://doi.org/10.1186/s12943-023-01841-8> (2023).
- Liu, C., Jin, Y. & Fan, Z. The mechanism of Warburg Effect-Induced chemoresistance in cancer. *Front. Oncol.* **11**, 698023. <https://doi.org/10.3389/fonc.2021.698023> (2021).
- Zou, J., Gu, Y., Zhu, Q., Li, X. & Qin, L. Identifying glycolysis-related lncRNAs for predicting prognosis in breast cancer patients. *Cancer Biomark.* **34**, 393–401. <https://doi.org/10.3233/CBM-210446> (2022).
- Reinfeld, B. I., Rathmell, W. K., Kim, T. K. & Rathmell, J. C. The therapeutic implications of immunosuppressive tumor aerobic glycolysis. *Cell. Mol. Immunol.* **19**, 46–58. <https://doi.org/10.1038/s41423-021-00727-3> (2022).
- Zhang, J. & Zhao, H. eQTL studies: from bulk tissues to single cells. *J. Genet. Genomics* **50**, 925–933. <https://doi.org/10.1016/j.jgg.2023.05.003> (2023).
- Gao, L. et al. Identifying noncoding risk variants using disease-relevant gene regulatory networks. *Nat. Commun.* **9**, 702. <https://doi.org/10.1038/s41467-018-03133-y> (2018).
- Fadista, J. et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. U S A* **111**, 13924–13929. <https://doi.org/10.1073/pnas.1402665111> (2014).
- Rasche, L. et al. Low expression of hexokinase-2 is associated with false-negative FDG-positron emission tomography in multiple myeloma. *Blood* **130**, 30–34. <https://doi.org/10.1182/blood-2017-03-774422> (2017).
- Feng, Y. et al. Causal effects of genetically determined metabolites on cancers included lung, breast, ovarian cancer, and glioma: a Mendelian randomization study. *Transl. Lung Cancer Res.* **11**, 1302–1314. <https://doi.org/10.21037/tlcr-22-34> (2022).
- Huang, R. et al. Genetically evaluating the causal role of peripheral immune cells in colorectal cancer: a two-sample Mendelian randomization study. *BMC Cancer* **24**, 753. <https://doi.org/10.1186/s12885-024-12515-z> (2024).
- Zhou, Z. et al. Single cell atlas reveals multilayered metabolic heterogeneity across tumour types. *EBioMedicine* **109**, 105389. <https://doi.org/10.1016/j.ebiom.2024.105389> (2024).
- Wu, X. et al. Identification and validation of glycolysis-related diagnostic signatures in diabetic nephropathy: a study based on integrative machine learning and single-cell sequence. *Front. Immunol.* **15**, 1427626. <https://doi.org/10.3389/fimmu.2024.1427626> (2024).
- 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.* **53**, D672–D677. <https://doi.org/10.1093/nar/gkae909> (2025).
- Kanehisa, M. Toward Understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).
- Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- Tang, Z. et al. Integrated analysis of multiple programmed cell death-related prognostic genes and functional validation of apoptosis-related genes in osteosarcoma. *Int. J. Biol. Macromol.* **307**, 142113. <https://doi.org/10.1016/j.ijbiomac.2025.142113> (2025).
- Gao, M. et al. Machine learning-based prognostic model of lactylation-related genes for predicting prognosis and immune infiltration in patients with lung adenocarcinoma. *Cancer Cell. Int.* **24**, 400. <https://doi.org/10.1186/s12935-024-03592-y> (2024).
- Li, X. et al. Structure of POU2AF1 Recombinant protein and it affects the progression and treatment of liver cancer based on WGCNA and molecular Docking analysis. *Int. J. Biol. Macromol.* **278**, 134629. <https://doi.org/10.1016/j.ijbiomac.2024.134629> (2024).
- Liao, L. Inequality in breast cancer: global statistics from 2022 to 2050. *Breast* **79**, 103851. <https://doi.org/10.1016/j.breast.2024.103851> (2025).
- Colomer, R. et al. Biomarkers in breast cancer 2024: an updated consensus statement by the Spanish society of medical oncology and the Spanish society of pathology. *Clin. Transl. Oncol.* **26**, 2935–2951. <https://doi.org/10.1007/s12094-024-03541-1> (2024).
- Cordani, M. et al. The role of Glycolysis in tumorigenesis: from biological aspects to therapeutic opportunities. *Neoplasia* **58**, 101076. <https://doi.org/10.1016/j.neo.2024.101076> (2024).
- Oshi, M. et al. Accelerated Glycolysis in tumor microenvironment is associated with worse survival in triple-negative but not consistently with ER+/HER2- breast cancer. *Am. J. Cancer Res.* **13**, 3041–3054 (2023).
- Wang, Z. H., Peng, W. B., Zhang, P., Yang, X. P. & Zhou, Q. Lactate in the tumour microenvironment: from immune modulation to therapy. *EBioMedicine* **73**, 103627. <https://doi.org/10.1016/j.ebiom.2021.103627> (2021).
- N, M. B. et al. Aerobic Glycolysis is a metabolic requirement to maintain the M2-like polarization of tumor-associated macrophages. *Biochim. Biophys. Acta Mol. Cell. Res.* **1867**, 118604. <https://doi.org/10.1016/j.bbamcr.2019.118604> (2020).
- Li, W. et al. Cell metabolism-based optimization strategy of CAR-T cell function in cancer therapy. *Front. Immunol.* **14**, 1186383. <https://doi.org/10.3389/fimmu.2023.1186383> (2023).
- Cascone, T. et al. Increased Tumor Glycolysis Characterizes Immune Resistance to Adoptive T Cell Therapy. *Cell Metab* **27**, 977–987.e4. <https://doi.org/10.1016/j.cmet.2018.02.024> (2018).
- Hashemi, V. et al. Regulatory T cells in breast cancer as a potent anti-cancer therapeutic target. *Int. Immunopharmacol.* **78**, 106087. <https://doi.org/10.1016/j.intimp.2019.106087> (2020).

34. Imani, S. et al. Reprogramming the breast tumor immune microenvironment: cold-to-hot transition for enhanced immunotherapy. *J. Exp. Clin. Cancer Res.* **44**, 131. <https://doi.org/10.1186/s13046-025-03394-8> (2025).
35. Behera, R., Kumar, V., Lohite, K., Karnik, S. & Kundu, G. C. Activation of JAK2/STAT3 signaling by osteopontin promotes tumor growth in human breast cancer cells. *Carcinogenesis* **31**, 192–200. <https://doi.org/10.1093/carcin/bgp289> (2010).
36. Wang, Z. et al. The CXCL family contributes to immunosuppressive microenvironment in gliomas and assists in gliomas chemotherapy. *Front. Immunol.* **12**, 731751. <https://doi.org/10.3389/fimmu.2021.731751> (2021).
37. Li, Y. et al. Tumor necrosis factor alpha-induced protein 8-like-2 controls microglia phenotype via metabolic reprogramming in BV2 microglial cells and responses to neuropathic pain. *Int. J. Biochem. Cell. Biol.* **169**, 106541. <https://doi.org/10.1016/j.biocel.2024.106541> (2024).
38. Zhi, X. et al. Potential prognostic biomarker CD73 regulates epidermal growth factor receptor expression in human breast cancer. *IUBMB Life.* **64**, 911–920. <https://doi.org/10.1002/iub.1086> (2012).
39. Spychala, J. et al. Role of Estrogen receptor in the regulation of ecto-5'-nucleotidase and adenosine in breast cancer. *Clin. Cancer Res.* **10**, 708–717. <https://doi.org/10.1158/1078-0432.ccr-0811-03> (2004).
40. Pentassuglia, L. & Sawyer, D. B. The role of Neuregulin-1beta/ErbB signaling in the heart. *Exp. Cell. Res.* **315**, 627–637. <https://doi.org/10.1016/j.yexcr.2008.08.015> (2009).
41. Liu, Y. et al. Highly heterogeneous-related genes of triple-negative breast cancer: potential diagnostic and prognostic biomarkers. *BMC Cancer.* **21**, 644. <https://doi.org/10.1186/s12885-021-08318-1> (2021).
42. Cancemi, P. et al. A multiomics analysis of S100 protein family in breast cancer. *Oncotarget* **9**, 29064–29081. <https://doi.org/10.18632/oncotarget.25561> (2018).
43. Salama, I., Malone, P. S., Mihaimeed, F. & Jones, J. L. A review of the S100 proteins in cancer. *Eur. J. Surg. Oncol.* **34**, 357–364. <https://doi.org/10.1016/j.ejso.2007.04.009> (2008).
44. Cheng, Y. & Tian, H. Current development status of MEK inhibitors. *Molecules* <https://doi.org/10.3390/molecules22101551> (2017).
45. Chresta, C. M. et al. AZD8055 is a potent, selective, and orally bioavailable ATP-competitive mammalian target of Rapamycin kinase inhibitor with in vitro and in vivo antitumor activity. *Cancer Res.* **70**, 288–298. <https://doi.org/10.1158/0008-5472.CAN-09-1751> (2010).
46. Serra, V. et al. PI3K Inhibition results in enhanced HER signaling and acquired ERK dependency in HER2-overexpressing breast cancer. *Oncogene* **30**, 2547–2557. <https://doi.org/10.1038/nc.2010.626> (2011).
47. Wang, L. et al. Ecto-5'-nucleotidase promotes invasion, migration and adhesion of human breast cancer cells. *J. Cancer Res. Clin. Oncol.* **134**, 365–372. <https://doi.org/10.1007/s00432-007-0292-z> (2008).
48. Bahl, N. et al. MCT4 blockade increases the efficacy of immune checkpoint blockade. *J. Immunother Cancer* <https://doi.org/10.1136/jitc-2023-007349> (2023).

## Acknowledgements

We would like to express our sincere gratitude to the funding agencies that supported this research. We are deeply indebted to all colleagues and collaborators whose expertise and assistance were instrumental in the successful completion of this study. Special thanks to the Medical Research Center of Quanzhou Medical College for providing and maintaining the essential experimental facilities and technical support. We also appreciate the constructive feedback from anonymous reviewers that helped improve the quality of this manuscript.

## Author contributions

YN and DC conceived the study, designed the research framework, and supervised the project. YJ contributed to experimental design, performed data analysis, and drafted the manuscript. YN and YJ conducted the laboratory experiments and collected the data. TW contributed to data interpretation, statistical analysis, and manuscript revision. DC secured funding and provided administrative support. All authors participated in critical discussions of the results, contributed to manuscript editing, and approved the final version for publication.

## Funding

This work was funded by Science and Technology Project of Quanzhou (Grant number, 2024QZC004YR); Startup Fund for scientific research, Fujian Medical University (2022QH1259); Nantong Municipal Science and Technology Project (JC2023033); Nantong Youth Medical Expert Project; Nantong Natural Science Foundation and scientific research special project of Nantong Health Commission (No. MS2022059).

## Declarations

### Competing interests

The authors declare no competing interests.

### Consent for publication

All patients enrolled in the study signed the consent for publication.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-29391-7>.

**Correspondence** and requests for materials should be addressed to D.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**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/>.

© The Author(s) 2025