



## OPEN ACCESS

## EDITED BY

Yutian Zou,  
Sun Yat-sen University Cancer Center  
(SYSUCC), China

## REVIEWED BY

Yajie Xiao,  
The Chinese University of  
Hong Kong, China  
Yuelin Kong,  
University of California, San Diego,  
United States

## \*CORRESPONDENCE

Zan Fu

✉ fuzan1971@njmu.edu.cn

<sup>†</sup>These authors have contributed equally to  
this work

RECEIVED 08 March 2023

ACCEPTED 25 April 2023

PUBLISHED 05 May 2023

## CITATION

Wang Z, Shao Y, Zhang H, Lu Y, Chen Y,  
Shen H, Huang C, Wu J and Fu Z (2023)  
Machine learning-based glycolysis-  
associated molecular classification  
reveals differences in prognosis, TME,  
and immunotherapy for colorectal  
cancer patients.  
*Front. Immunol.* 14:1181985.  
doi: 10.3389/fimmu.2023.1181985

## COPYRIGHT

© 2023 Wang, Shao, Zhang, Lu, Chen, Shen,  
Huang, Wu and Fu. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Machine learning-based glycolysis-associated molecular classification reveals differences in prognosis, TME, and immunotherapy for colorectal cancer patients

Zhenling Wang<sup>1,2†</sup>, Yu Shao<sup>1,2†</sup>, Hongqiang Zhang<sup>1,2†</sup>,  
Yunfei Lu<sup>1,2</sup>, Yang Chen<sup>1,2</sup>, Hengyang Shen<sup>1,2</sup>,  
Changzhi Huang<sup>1,2</sup>, Jingyu Wu<sup>1,2</sup> and Zan Fu<sup>1,2\*</sup>

<sup>1</sup>Department of General Surgery, The First Affiliated Hospital of Nanjing Medical University, Nanjing, Jiangsu, China, <sup>2</sup>The First School of Clinical Medicine, Nanjing Medical University, Nanjing, Jiangsu, China

**Background:** Aerobic glycolysis is a process that metabolizes glucose under aerobic conditions, finally producing pyruvate, lactic acid, and ATP for tumor cells. Nevertheless, the overall significance of glycolysis-related genes in colorectal cancer and how they affect the immune microenvironment have not been investigated.

**Methods:** By combining the transcriptome and single-cell analysis, we summarize the various expression patterns of glycolysis-related genes in colorectal cancer. Three glycolysis-associated clusters (GAC) were identified with distinct clinical, genomic, and tumor microenvironment (TME). By mapping GAC to single-cell RNA sequencing analysis (scRNA-seq), we next discovered that the immune infiltration profile of GACs was similar to that of bulk RNA sequencing analysis (bulk RNA-seq). In order to determine the kind of GAC for each sample, we developed the GAC predictor using markers of single cells and GACs that were most pertinent to clinical prognostic indications. Additionally, potential drugs for each GAC were discovered using different algorithms.

**Results:** GAC1 was comparable to the immune-desert type, with a low mutation probability and a relatively general prognosis; GAC2 was more likely to be immune-inflamed/excluded, with more immunosuppressive cells and stromal components, which also carried the risk of the poorest prognosis; Similar to the immune-activated type, GAC3 had a high mutation rate, more active immune cells, and excellent therapeutic potential.

**Conclusion:** In conclusion, we combined transcriptome and single-cell data to identify new molecular subtypes using glycolysis-related genes in colorectal cancer based on machine-learning methods, which provided therapeutic direction for colorectal patients.

#### KEYWORDS

glycolysis, colorectal cancer, molecular subtypes, tumor immune infiltration, machine learning, single-cell analysis

## Introduction

Recently, reprogramming of the tumor's energy metabolism has emerged as one of the tumor's hallmarks (1). Since the "Warburg effect" was put forth, researchers have steadily investigated the connection between tumors and glycolysis (2, 3). As is a crucial form of the transformation of energy, glycolysis converts glucose into pyruvate, eventually creating lactic acid and giving tumor cells oxygen-independent energy (4). Tumor cells can still restrict energy metabolism to glycolysis, often known as aerobic glycolysis, even in an aerobic environment (3). By reducing cell inhibition and apoptosis, glycolysis creates the conditions for tumor cell growth (5, 6). Numerous intermediate products of glycolysis serve as starting materials for the synthesis of nucleosides and amino acids, both of which aid in the production of new cells (7). Additionally, certain tumor cells fall into one of two subgroups: aerobic or hypoxic. The utilization of lactate and glucose in these two subgroups' energy metabolisms differs from one another (8). In conclusion, glycolysis is crucial for the energy metabolism of tumors.

Studying colorectal cancer (CRC) is important due to its high incidence and fatality rates (9). Glycolysis is implicated in the development of colorectal cancer, according to increasing amounts of evidence (10). Many glycolysis-related genes, including lactate, GLUT1, pyruvate kinase M2, glyceraldehyde-3-phosphate dehydrogenase, enolase-1, lactate dehydrogenase 5, and hexokinase 2, are currently discovered to be up-regulated in colorectal cancer (11–17). Additionally, there is growing appreciation for the function of the tumor microenvironment (TME). Studies show that different areas of a tumor mass exhibit metabolic heterogeneity (18). CRC cells contain both OXPHOS and glycolysis phenotypes (19). Near the blood capillaries, the majority of CRC cells have greater OXPHOS, whereas tumor cells farthest

from the blood vessels exhibit a glycolysis phenotype (20). This unexpected increase in OXPHOS is described as the "reverse Warburg effect" (10). Meanwhile, the interaction between tumor cells and other cells in the TME has an impact on metabolic remodeling. Lactate is transported from CAF to CRC cells when mono carbohydrate transporters (MCTs) are upregulated (21). A study showed the primary enzyme of glycolysis, pyruvate kinase M (PKM1 and PKM2), which is artificially highly expressed in stromal cells, has the potential to encourage tumor cell proliferation and invasion (22). While naive T cells depend on OXPHOS for energy supply, activated T cells require glycolysis (23). All of the evidence provided shows that glycolysis is essential for controlling the interaction between tumor cells and TME.

Immunotherapy has gained significant traction in recent years for the treatment of advanced solid tumors, including colorectal cancer. Especially, Immune checkpoint inhibitors (ICIs) are beneficial for treating patients with metastatic colorectal cancer who have mismatch-repair-deficient (dMMR) or microsatellite instability-high (MSI-H). Unfortunately, ICIs have not yet been effective in CRC patients with mismatch-repair-proficient (pMMR) or microsatellite-stable (MSS) or low microsatellite instability (MSI-L) (24, 25). Additionally, acquired resistance has emerged as one of the impediments to immunotherapy. Hence identifying new CRC subtypes and screening novel treatment targets are crucial. Targeted therapy for glycolysis has gained popularity due to the role glycolysis plays in CRC (4). For instance, compound 2-deoxyglucose (2-DG), a glycolysis inhibitor, can lessen cell invasion (26). LND, another glycolysis inhibitor, can also make chemotherapy treatments more effective and stop the growth of several tumor cells, including CRC (10, 27). However, due to the different functions of glycolysis in various cell types and the heterogeneity of carbohydrate metabolism in tumor cells, targeted medications for glycolysis are not that effective. Therefore, it is crucial to investigate the subtypes of glycolysis in CRC to uncover potential therapeutics. Currently, despite some studies reporting CRC's glycolytic signature employing several genes and long non-coding RNAs (28, 29), there is no research focusing on the gene-based molecular subtypes of glycolysis in CRC. In this study, we used the unsupervised clustering method to categorize each CRC patient using genes associated with glycolysis. The characteristics of each glycolysis-associated cluster's (GAC) clinical, genomics, TME, and enrichment pathway were then addressed. We also examined the distribution and function of each GAC-like subtype in each cell

**Abbreviations:** GAC, glycolysis-associated clusters; TME, tumor microenvironment; scRNA-seq, single-cell RNA sequencing analysis; bulk RNA-seq, bulk RNA sequencing analysis; CRC, colorectal cancer; GEO, Gene Expression Omnibus; TCGA, The Cancer Genome Atlas; DEGs, differentially expressed genes; GSEA, gene set variation analysis; CMS, consensus molecular subtype; ssGSEA, single-sample gene set enrichment analysis; CSC, cancer stem cell; CNV, copy number variation; TMB, tumor mutation burden; OS, overall survival; DFS, disease-free survival; PCA, principal component analysis; MSI-H, microsatellite instability-high; MSI, microsatellite Instability; AUC, Area Under Curve.

after mapping each GAC type to the single-cell data. Based on the findings above, we integrated the results of single cell and bulk RNA-seq to develop a model that predicts the GAC of each CRC patient and verified it, whose role in treatment was also explored.

## Materials and methods

### Patient population and bulk RNA expression acquisition

Five colorectal cancer data were involved in this study, including TCGA-COAD/READ, GSE39582, GSE38832, GSE17538, and GSE14333. Corresponding clinical features and RNA-seq data were achieved from UCSC Xena (<https://xenabrowser.net/>) and the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/geo/>). We excluded patients with OS and DFS less than 30 to ensure reliability in accordance with previous studies (30–32). Particularly, we employed TCGA data as the training set and GEO data as the verification set. For this study, we obtained the characteristic of 1603 CRC patients, the baseline of which was shown in Table S1. Besides, we downloaded expression profiles and clinical data from the two cohorts (IMVigor210, GSE78220) for seeking the role GAC predictor played in immunotherapy. The batch effect has been removed by the “Combat” algorithm.

### Consensus clustering analysis

198 Glycolysis-related genes were acquired from the MSigDB Team (HALLMARK\_GLYCOLYSIS) (Table S2). To further examine the various expression patterns of glycolysis-related genes in CRC, we performed consensus clustering using the *k*-means method with the “ConsensusClusterPlus” package to classify individuals with CRC (33). We set 80% sampling each time and 1000 iterations to ensure the consistency of the clustering process. The optimal number of the clustering was determined by the consensus heatmap and cumulative distribution function (CDF) curves. Kaplan–Meier curves were drawn to exhibit the prognosis of each glycolysis-associated cluster (GAC) using “survival” and “survminer” packages (34). The same procedure was conducted once more by using DEGs between the GACs, which was to acquire gene clusters.

### Gene set variation analysis and consensus molecular subtype classification

Supported by the “GSVA” package, we utilized the GSVA method to explore the differences in pathways enriched by each cluster. Gene sets (C2.cp.kegg, C2.cp.reactome, C2.go.bp, and hallmark gene sets) were downloaded from the MSigDB Team v2022.1. “limma” package was used to screen significant pathways (adjusted *p*-value < 0.05). The symbol pathways of each cluster were condensed and shown as a heatmap. We performed GSVA in bulk and single-cell level. Besides, we applied the CMScaller package to identify each CRC patient’s CMS attribution.

### Single-sample gene set enrichment analysis

With quantification of pathway enrichment in each patient, ssGSEA enables the comparability of various pathways in each sample. The “GSVA” package was utilized in the procedure. To better reveal the clinical and CMS heterogeneity, we plotted a heatmap to visualize the differences. Related pathways were achieved from the “CMScaller” package. Based on Lee et al.’s study (35), markers of 31 cell types with  $\log_2FC > 0.25$  were delivered to the ssGSEA scoring system. Besides, marker genes in Charoentong’s study were utilized (36).

### Single-cell data preparation

We obtained GSE132465 for single-cell analysis. Corresponding RNA-seq and annotation data were downloaded in the GEO database. Clinical information of 23 CRC patients was acquired from the original article (35). Based on the “Seurat” package, we carried out the preliminary processing of the scRNA-seq data following the criterion of Lee et al.’s study (genes detected in each cell, mitochondrial gene expression). 23 tumor samples were extracted and we further run the “Harmony” function to integrate the scRNA-seq. The top 30 dimensions were selected while processing the t-SNE.

### Identification of GAC at single cell level

We further mapped the consensus clusters (GAC) to the scRNA-seq to investigate their function at cellular level. With the “FindAllMarkers” function, we identified the marker genes of each GAC in bulk-seq data ( $\log_2FC > 0.5$ ) and then submitted them to the “AddModuleScore” function in scRNA-seq data. Each cell was given a score, and the cluster with the highest score was used to determine the cell’s final annotation result. The annotation was shown in Table S3. Combing the original and GAC annotation of each cell, we obtained a new glycolysis-related cell subtype such as “Epithelial-G1”. Then in scRNA data, we collected the marker genes of each cell subtype (Table S4). Additionally, glycolysis-related genes signature was also demonstrated by the “AddModuleScore” function.

### Prognosis analysis based on glycolysis-related cell subtype

With the marker genes of glycolysis-related cell subtype obtained using the “FindAllMarkers” function, we selected those with the top 50 ranks in  $\log_2FC$  (If less than 50, follow the existing options) and applied them to GSVA analyses in bulk-seq data. Based on the GSVA score, the prognosis-related cell subtypes were found using Cox regression analysis, and the most significant cell subtypes were identified by combining the findings from various cohorts using the “rma” function of the “metafor” package.

## Pseudotime trajectory analysis and cell-cell communication analysis of glycolysis-related epithelial cell subtypes

By creating the evolutionary trajectory between cells, pseudotime analysis predicts the transformation of cells over time. First, we extracted epithelial cells and performed the same quality control, integration, and dimension reduction process. Then we conducted pseudotime analysis on the processed data with the “monocle3” package. Additionally, cell communication analysis was completed with the “CellChat” package.

## Assessment of TME among GACs using different algorithms

To depict the landscape of tumor immune infiltration with bulk-seq data, we performed three algorithms: ssGSEA, Cibersort, and Estimate. We calculated the ssGSEA score of 31 cell types in Lee et al.’s research (35). Charoentong’s cell types were also described by ssGSEA algorithms. We further calculated the proportion of each immune cell with Cibersort analysis. Estimate algorithm contains three scores: Stromal score, immune score, and combined score. All the above analyses were conducted in TCGA and GEO datasets for the robustness of the results.

## Genome-related analysis and cancer stem cell index

In this part, we focused on the mutation of the tumor, copy number variation (CNV), methylation, microsatellite stability, and cancer stem cell index. The somatic mutation data and CNV files were downloaded from the GDC TCGA database. Supported by the “maftools” package, we drew waterfall plots to demonstrate the somatic mutation of COAD and READ patients in TCGA. TMB was quantified into  $\log(\text{TMB})$ , termed as TMB score. Downloaded by “TCGAbiolinks”, CNV and methylation were exhibited by lollipop and violin plots, respectively. MSI status was plotted using a graph of proportion. Subsequently, Utilizing the data from Progenitor Cell Biology Consortium (PCBC), we trained the model of mRNAsi prediction and further applied it to calculate the CSC index of our samples. The relationship between GAC and CSC index was discussed.

## Construction and validation of GAC predictor using four machine learning methods

The TCGA cohort served as our training set, and the GEO cohorts served as our validation set. To increase the reliability of the results at the macro (Bulk-seq) and micro (scRNA-seq) levels, the genes used to predict GAC were determined to be the intersection of DEGs of GAC and marker genes for glycolysis-related single-cell subtypes that have prognostic implications. Four machine learning

methods were conducted to select important genes, including least absolute shrinkage and selection operator (LASSO) regression, random forest (Boruta), extreme gradient boosting (XGBoost), and support vector machine (SVM) (37–40). The R package we used contained “glmnet”, “randomForest”, “Boruta”, “XGBoost”, “e1071”, and “caret”. We used these important genes to identify the GAC of each patient. The final model, termed as “GAC predictor”, was decided as multinomial logistic regression and constructed by the “multinom” function of the “nnet” package. We estimated which GAC each patient might belong to using the “prediction” function, and we chose the GAC with the highest probability. “pROC” and “caret” packages were used to evaluate the results by calculating AUCs, accuracy, sensitivity, and specificity. We further tested the “GAC predictor” in a similar way on the validation sets. The model was shown as follows:

$$\text{Model}_{\text{GAC3}} = \sum_1^i (\text{EstiGAC3}_i * \text{ExpGene}_i)$$

$$\text{Model}_{\text{GAC2}} = \sum_1^i (\text{EstiGAC2}_i * \text{ExpGene}_i)$$

$$\text{Model}_{\text{GAC1}} = 0 \quad (\text{reference group})$$

$$P_{\text{GAC3}} = \frac{\exp(\text{Model}_{\text{GAC3}})}{[\exp(\text{Model}_{\text{GAC1}}) + \exp(\text{Model}_{\text{GAC2}}) + \exp(\text{Model}_{\text{GAC3}})]}$$

$$P_{\text{GAC2}} = \frac{\exp(\text{Model}_{\text{GAC2}})}{[\exp(\text{Model}_{\text{GAC1}}) + \exp(\text{Model}_{\text{GAC2}}) + \exp(\text{Model}_{\text{GAC3}})]}$$

$$P_{\text{GAC1}} = \frac{\exp(\text{Model}_{\text{GAC1}})}{[\exp(\text{Model}_{\text{GAC1}}) + \exp(\text{Model}_{\text{GAC2}}) + \exp(\text{Model}_{\text{GAC3}})]}$$

EstiGAC3 and EstiGAC2 represented the estimate of multinomial logistic regression. ExpGene represented the expression of the gene. P represented the probability that each patient belongs to each GAC.

## Treatment-related analysis

To excavate the treatment of each GAC, we obtained potential therapeutic compounds related to each GAC based on typical marker genes using Cmap analysis (<https://clue.io/>). For each GAC, the 10 compounds with the lowest enrichment score, which were considered antagonistic drugs, will be presented in the form of a heatmap. Besides, to gauge each GAC’s drug sensitivity, we computed the semi-inhibitory concentration (IC50) values of common medicines using the “pRRophetic” package.

## Statistical analysis

Wilcoxon rank-sum test was applied to show the difference between the two groups. Wilcoxon signed-rank test was used in two

paired groups. Kruskal-Wallis H test was performed to compare three or more groups. Dunn test was used for multiple comparisons. Chi-square test was used in the proportion test. The Pearson test was used to identify correlations of the data. The log-rank test method was employed to analyze survival data. All of the statistical analyses were conducted using R 4.1.3 ( $p < 0.05$ ).

## Results

### Classification of glycolysis-associated molecular clusters in colorectal cancer

The whole flow chart was shown in Figure 1. We collected 198 glycolysis-related genes from the MSigDB Team (HALLMARK\_GLYCOLYSIS) for further study. Employing GO enrichment analysis, we found these genes were characterized by the metabolic process of carbohydrates and derivatives (Figure 2A). The comparison of glycolysis-related genes between normal and tumor

tissues from TCGA COAD/READ was visualized with a PCA map. The two main components indicated that glycolysis-related genes were effective at separating tumors from normal tissue (Figure 2B). To investigate the latent role these genes played in oncogenesis, we used consensus clustering to categorize CRC patients (TCGA and GEO datasets). Three glycolysis-associated clusters were identified, namely GAC, according to Figures 2C and S1. To explore the distribution of each GAC, t-SNE analysis was employed and showed significant GAC heterogeneity (Figure 2D).

### The clinical discrepancy among GACs

Built on GAC clustering, we discussed the corresponding clinical features of each GAC. First, we plotted survival curves of overall survival (OS) and disease-free survival (DFS) for three GACs and found GAC2 was in an unsatisfactory prognosis (TCGA-OS:  $p = 0.05$ ; TCGA-DFS:  $p = 0.03$ ; GEO-OS:  $p < 0.001$ ; GEO-DFS:  $p = 0.011$ , log-rank test) (Figures 2E–H). Moreover, the relationship

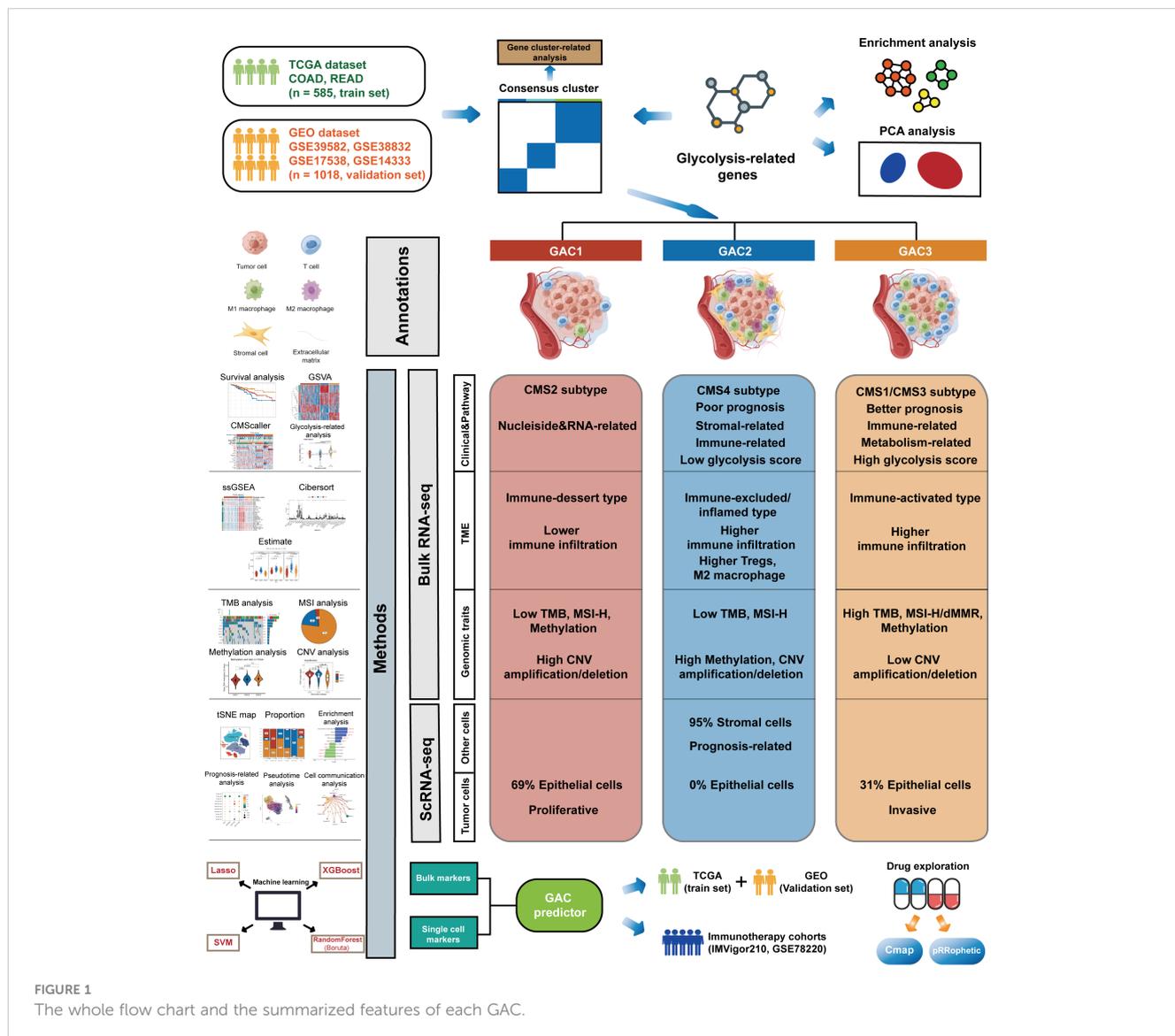
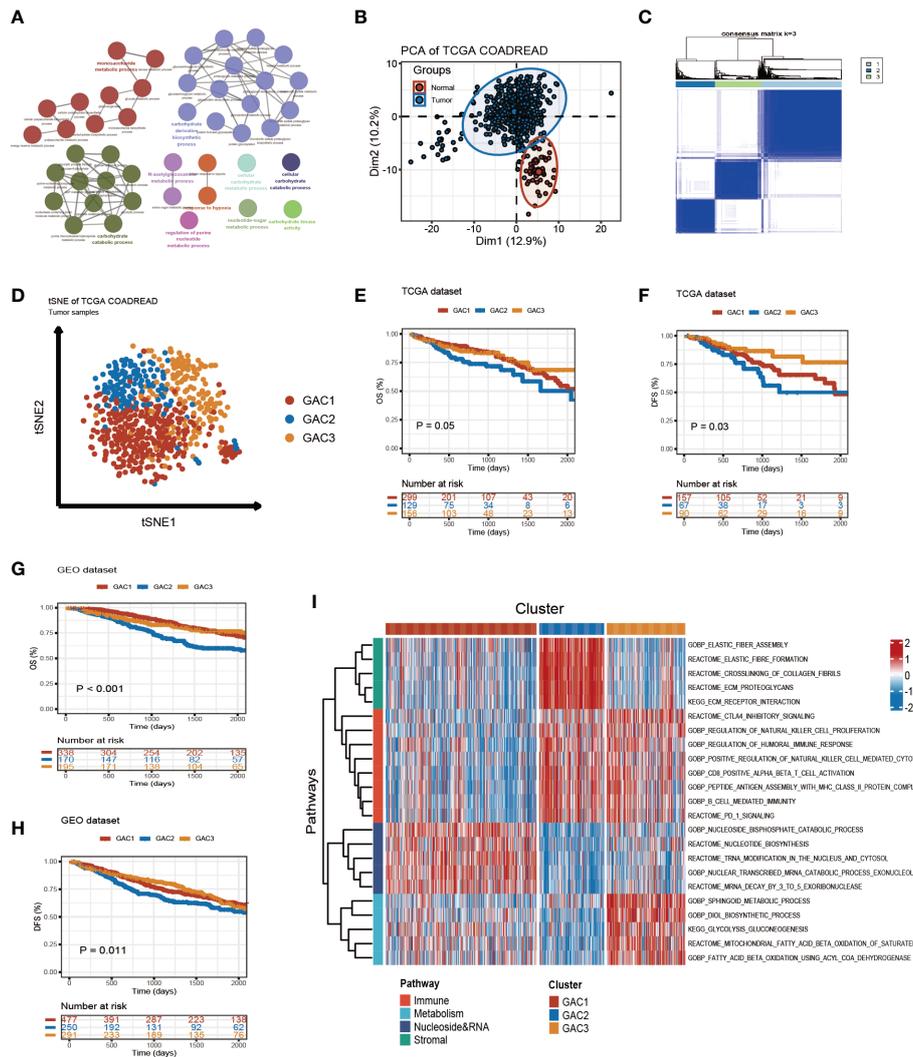


FIGURE 1 The whole flow chart and the summarized features of each GAC.



**FIGURE 2** Glycolysis-related molecular subtypes in colorectal cancer. **(A)** Glycolysis-related pathways presented by GO enrichment analyses using metaspape. **(B)** Principal component analysis utilizing glycolysis-related genes to separate tumor from normal tissue. **(C)** Consensus clustering identifying three clusters with different expression pattern of glycolysis-related genes. **(D)** tSNE plot visualizing the three GACs with obvious differentiation. **(E–H)** Survival curves of OS and DFS for the GACs in TCGA and GEO datasets. **(I)** GSEA analyses characterizing the biological process of the three GACs. Red to blue represents the range of enrichment from high to low.

between clinical stage and GAC indicated that the proportion of GAC2 increased with advancing stage and GAC3 on the opposite, which was consistent with the poor prognosis of GAC2 (Figures 3C, S2D). According to Figures 3A, S2A, and Table S1, GAC1 has a larger proportion of Kras mutations, while GAC3 has a higher proportion of Braf mutations and MSI-H status. Besides, GAC2 possessed the youngest age distribution. The clinical characteristics of GACs were described in this section as being heterogeneous: 1. GAC2 had a poor prognosis with advanced stage and younger trend; 2. GAC3 held a favorable prognosis with an early stage and higher mutation probability. 3. GAC1 was in the middle position of various clinical features.

### Biological differences and features of each GAC

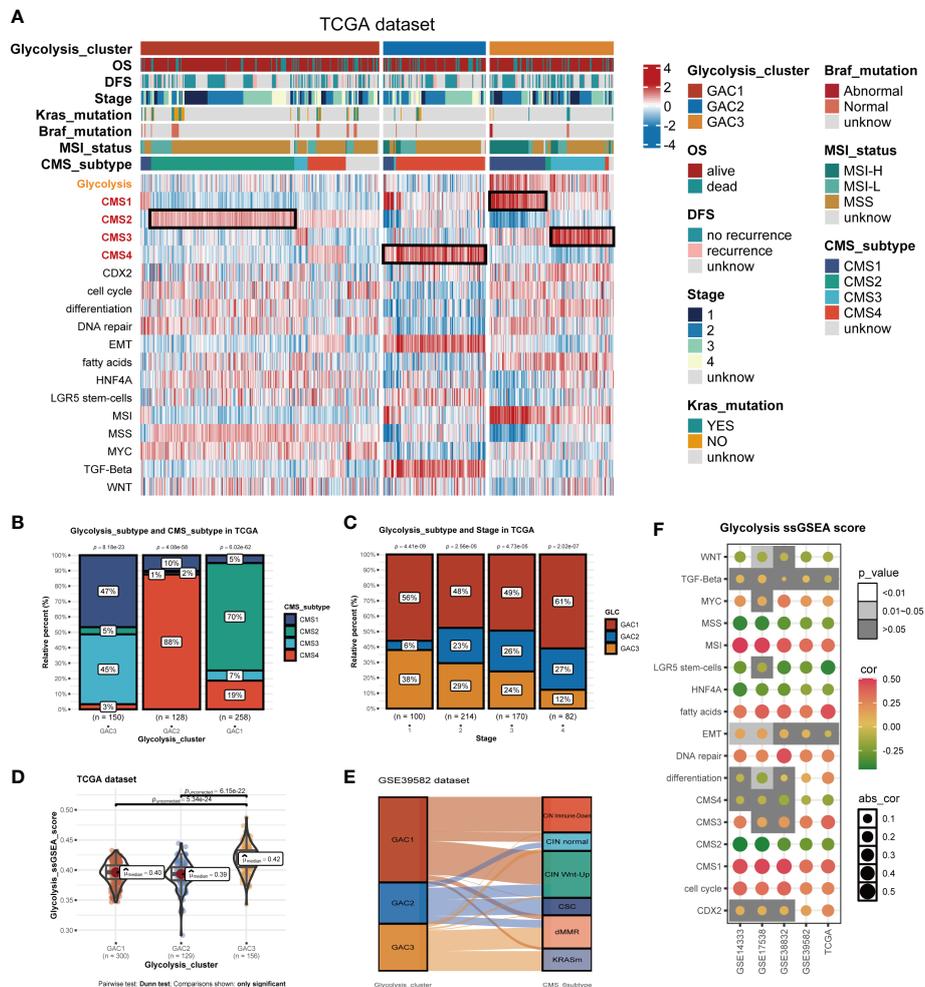
Utilizing “GSEA” algorithm, we elaboratively chose four types of phenotype (immune-related, metabolism-related, nucleoside&RNA-related, and stromal-related) as the exhibition in the form of a heatmap (Figure 2I). The pathways contained “C2.cp.kegg”, “C2.cp.reactome”, and “C2.go.bp”. With the “limma” package, differential pathways were identified. GAC1 was enriched in pathways connected to nucleosides. GAC2 was associated with stromal and immunological functions. Immune and metabolic traits were present in GAC3. Built on this, the relationship between GAC

and CMS was worth discussing. Afterward, we created a heatmap to display the relationship between GAC and CMS (Figure 3A). Surprisingly, we found GAC1 and GAC2 overlap mostly with CMS2 and CMS4, respectively. While in patients with GAC3, CMS1 and CMS3 occupied an equal and largest proportion (Figures 3A, B). We further validated these findings using the classical pathways of the “CMScaller” package. “EMT” and “TGF-Beta” pathways, associated with CMS4, were enriched in GAC2, which was a convincing illustration (Figures 3A, S2A, B). We quantified the glycolysis pathway with ssGSEA and discovered a high expression in GAC3 (Figures 3A, D, S2A, E). Additionally, we found that glycolysis was negatively connected with WNT, MSS, LGR5 stem cells, HNF4A, CMS2, and CMS4, and positively correlated with MSI, fat acid, DNA repair, CMS1, CMS3, and cell cycle (Figure 3F). A displayed alluvial diagram of GSE39582 further demonstrated that GAC1 flows to CIN immune-down and CIN Wnt-up, GAC3 flows to dMMR and KRASm, and GAC2 flows to the remainder (Figure 3E). Conclusively, GAC1 shared the most

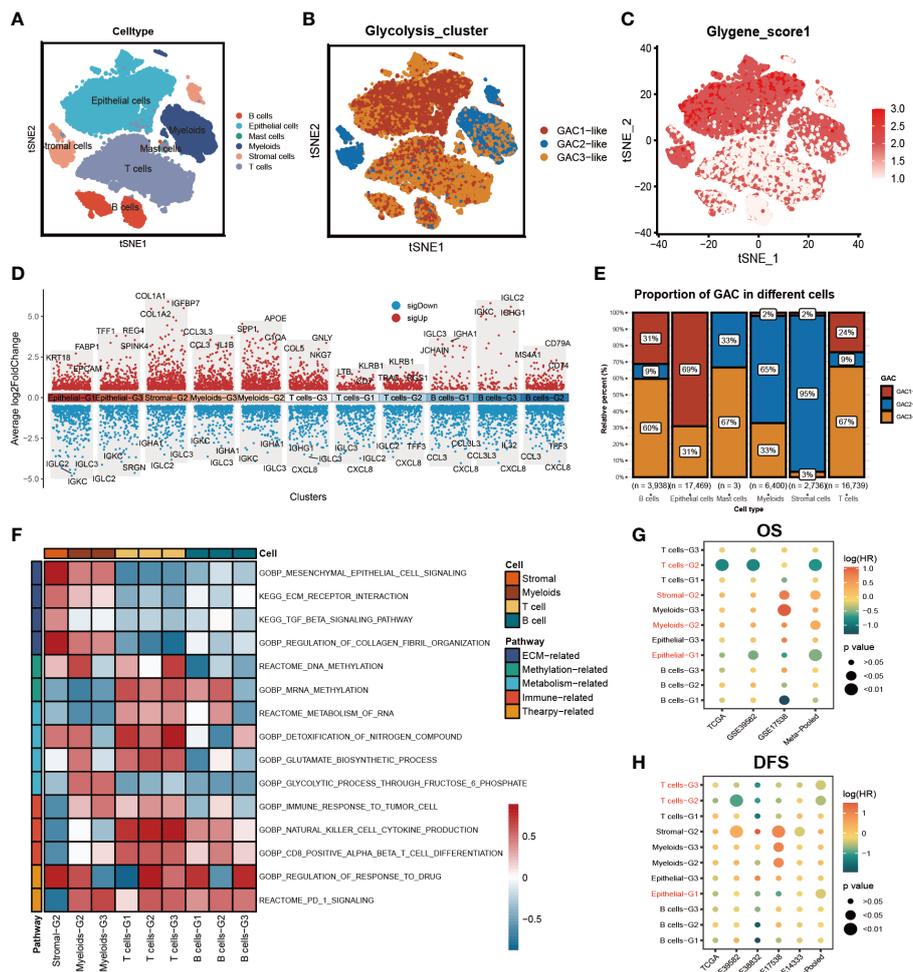
features of CMS2. GAC2 was the most similar to CMS4, and GAC3 share the dual features of CMS1 and CMS3. Same discoveries were found in the GEO dataset (Figure S2).

### Mapping glycolysis-associated clusters onto CRC cells at scRNA-seq level

On the basis of bulk-seq, we attempted to construct this GAC classification on individual cells. Through preprocessing of GSE132465, 23 tumor samples were integrated to perform further analysis. Cell type annotation was based on Lee et al.’s study and visualized by a t-SNE map (Figure 4A). With “AddModuleScore” utilized, we successfully mapped three GAC-like subclusters onto single cells (Figure 4B). Additionally, the glycolysis-related gene signature (glygene) was confirmed in Figure 4C, demonstrating its main presence in myeloid, stromal, and epithelial cells. The proportion of each GAC in each cell was then examined after we



**FIGURE 3** Biological activities and clinical features of the three GACs. (A) The heatmap comprehensively assessed each GAC’s biological and clinical parameters in TCGA dataset. Notable CMS subtypes with high correlations to GACs were represented by the black mark box. (B) Bar charts showing CMSs’ proportion of each GAC with chi-square test. (C) Bar charts showing each stage’s proportion of each GAC with chi-square test. (D) The distribution of glycolysis score acquired from ssGSEA in the three GACs. Comparisons were conducted by Dunn test. (E) Alluvial plot exhibiting the molecular subtype attribution of GACs in GSE39582. (F) The association between glycolysis and various pathways visualized by a bubble chart.



**FIGURE 4** The characteristics of the tumor microenvironment in each GAC at single cell level. (A) The distribution of each cell type visualized by tSNE map. (B) Mapping GACs into each cell and the distribution of GACs at scRNA level. (C) The dispersion of glycogenesis score in the tSNE map. (D) Marker genes of each cell subtype. (E) The bar graph displaying the percentage of each GAC in various cells. (F) GSVA analysis using a heatmap to depict each cell subtype's typical biological processes. (G, H) Meta-analysis determining the cell subtypes related to prognosis (OS, DFS).

added the GAC-like subclusters to each cell type (Figure 4E). Interestingly, we discovered that epithelial cells were filled only with GAC1-like (69%) and GAC3-like (31%). Though T and B cells shared GAC characteristics with epithelial cells, GAC3-like was presented in higher concentrations than GAC1-like in these cells, which makes them distinct from epithelial cells. Stromal cells were almost GAC2-like (95%), consistent with Figures 11, 2B. Myeloids contain 65% GAC2-like and 33% GAC3-like. The marker genes of each cell subtype were depicted by a combined volcano plot (Figure 4D, Table S4). GSVA analysis revealed distinct biological pathways among GAC-like subclusters (Figure 4F). Apparently, pathways of drug response, DNA methylation, and glutamate metabolism were enriched in myeloids-G2, while myeloid-G3 was on the opposite. T cell-G2 has a favorable response to drugs. B cell-G1, 2 had higher mRNA methylation. Importantly, we filtered the most clinically relevant GAC-like subclusters uniting prognosis (OS and DFS), the marker genes of which were responsible for later GAC predictor development. With meta-pool analysis showing the significance of T cells-G2, T cells-G3, Stromal-G2, Myeloids-G2,

and Epithelial-G1, we identified these five GAC-like subclusters as independent prognostic factors (Figures 4G, H).

### Analysis of the two GAC-like subclusters in epithelial cells

We subsequently chose epithelial cells to analyze their characteristics. The distribution of GAC-like subclusters and CMS subtypes was presented in the t-SNE map (Figures 5A, B). Discovery was found that the corresponding relationship between CMS and GAC-like subclusters was comparable to the bulk-seq's, demonstrating the relationship's robustness. Through GO enrichment analysis we found that epithelial-G3 was enriched in the feature of metastasis, whereas epithelial-G1 was mostly associated with RNA biological process (Figure 5C). Likewise, by GSVA analysis we displayed 50 hallmark signatures in the two types of epithelial cells, demonstrating GAC3-like's anaerobic traits and GAC1-like's ability to up-regulate oxidative phosphorylation

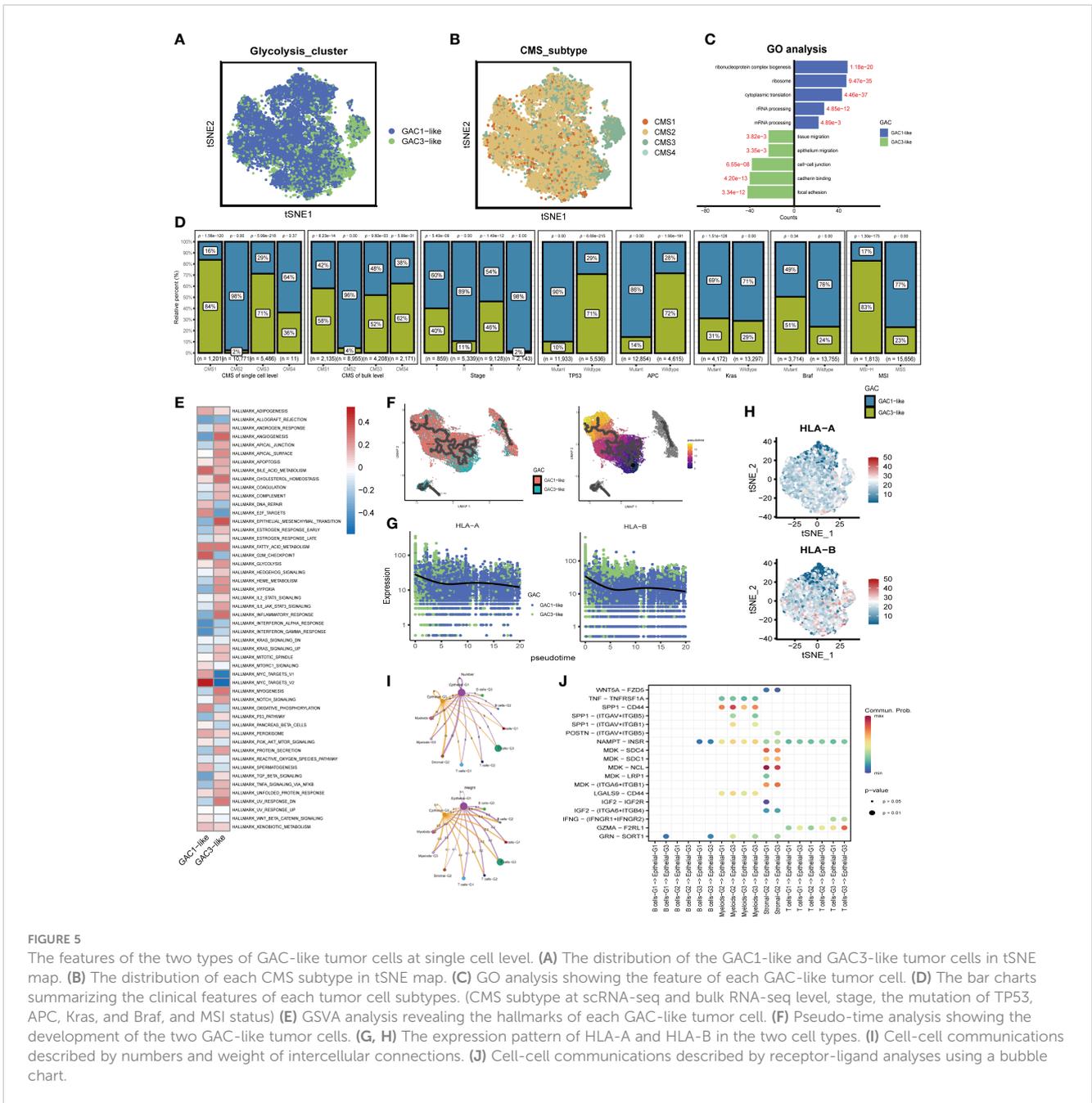
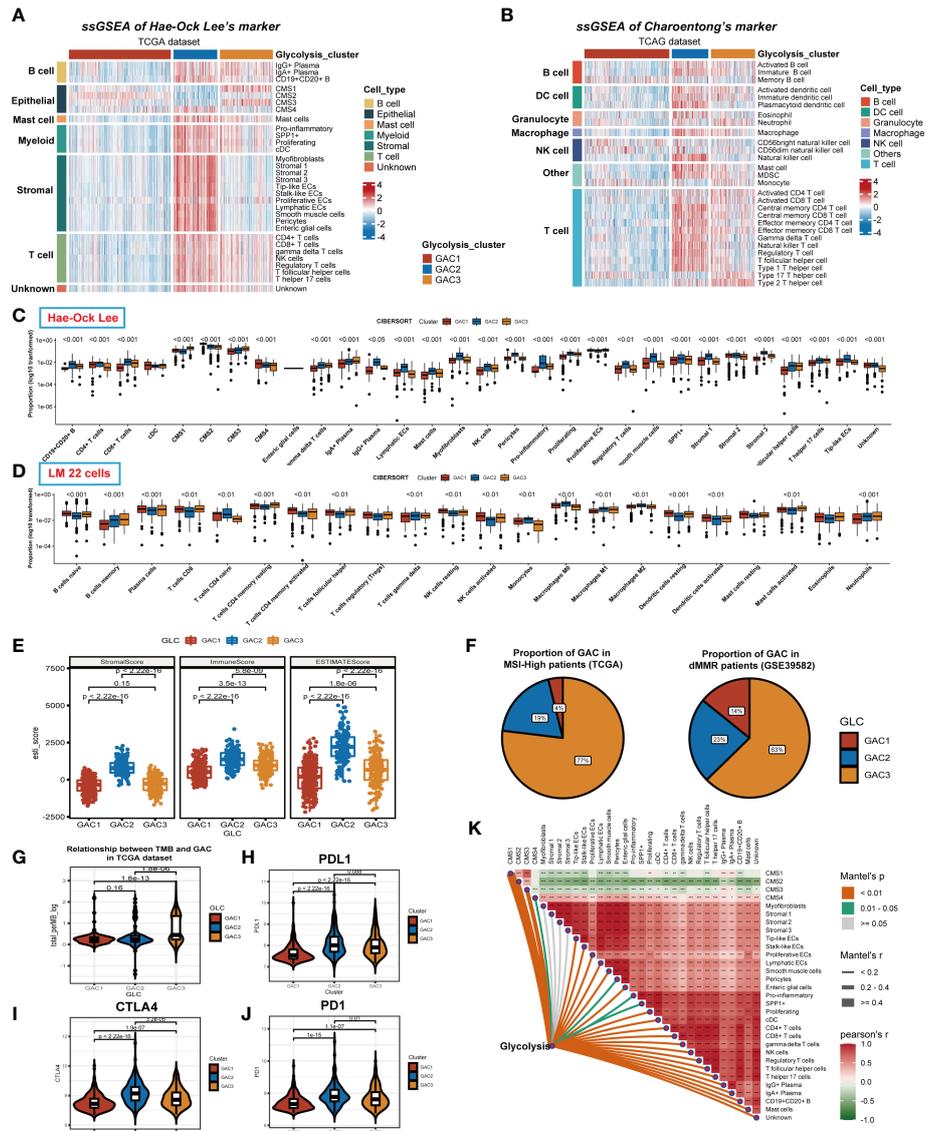


FIGURE 5

The features of the two types of GAC-like tumor cells at single cell level. (A) The distribution of the GAC1-like and GAC3-like tumor cells in tSNE map. (B) The distribution of each CMS subtype in tSNE map. (C) GO analysis showing the feature of each GAC-like tumor cell. (D) The bar charts summarizing the clinical features of each tumor cell subtypes. (CMS subtype at scRNA-seq and bulk RNA-seq level, stage, the mutation of TP53, APC, Kras, and Braf, and MSI status) (E) GSA analysis revealing the hallmarks of each GAC-like tumor cell. (F) Pseudo-time analysis showing the development of the two GAC-like tumor cells. (G, H) The expression pattern of HLA-A and HLA-B in the two cell types. (I) Cell-cell communications described by numbers and weight of intercellular connections. (J) Cell-cell communications described by receptor-ligand analyses using a bubble chart.

(Figure 5E). The results of both enrichment analyses showed that GAC1-like epithelial cells proliferated and that GAC3-like epithelial cells invaded. In addition, the study on the clinical characteristics of the two epithelial subtypes showed varieties. For the CMS subtypes, CMS2 was GAC1-like at both the scRNA and bulk-RNA levels, while a significant fraction of GAC3-like was found in CMS1 and CM3, which was in line with previous results. For stage, patients in the middle and late phases of CRC demonstrated an increased percentage of GAC1-like epithelial cells. For mutation, GAC1-like epithelial cells harbored high levels of TP53 and APC mutations, the rest site not significant. For MSS, most GAC3-like cells have MSI-H features (Figure 5D). According to pseudo-time analysis, epithelial cells will transition from being GAC3-like to being GAC1-like

(Figure 5F). Meanwhile, HLA-A and HLA-B were primarily found in GAC3-like cells, suggesting that these cells had a stronger propensity for immunological responses (Figures 5G, H). Finally, cell communication analysis was used to investigate the impact of non-tumor cells on the tumor epithelium. The frequency of linkages between Stromal-G2 and Myeloid-G2 and epithelial cells was substantially higher, but Epithelial-G3 was more significantly impacted by B cells-G2 (Figure 5I). A bubble plot specifically showed the ligand-receptor network, from which we could tell SPP1-CD44 was noteworthy in myeloids-epithelial interactions. CD44 was considered to promote stemness and invasiveness in colorectal cancer (41). MDK ligands were associated with multiple receptors to regulate the interaction between stromal and epithelial



**FIGURE 6** Immune infiltration analyses of the three GACs at bulk RNA-seq level in TCGA dataset. **(A, B)** ssGSEA analyses based on markers of Hae-ock Lee and Charoentong revealing the expression of TME cells of the GACs. **(C, D)** Cibersort analysis depicting the percentage of each cell in GACs based on markers of Hae-ock Lee and LM 22 cells. The proportion was converted to log10 format for better visual effects. Statistical analyses still used the unconverted original data source. **(E)** Estimate algorithm calculating stromal, immune and overall score of the GACs. **(F)** The pie chart showing the proportion of MSI-H and dMMR in each GAC. **(G)** Violin chart exhibiting the TMB score of each GAC. **(H-J)** The expression of PD1, PDL1, and CTLA4 in each GAC. **(K)** The relationship between TME cells and glycolysis. (\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ ).

cells. GZMA-F2RL1 was identified in the communication of T cells-G3 and Epithelial-G3 (Figure 5J). Conclusively, the proliferative type, epithelial-G1, and the invasive type, epithelial-G3, were both regulated by TME.

### Correlation between GAC and tumor microenvironment

We previously discussed the heterogeneity of GAC in TME at single-cell level, and in this part, we would go back to the bulk level. First, we collected two groups of TME cell signatures (Lee's and Charoentong's) for ssGSEA analysis. Same as the scRNA level, we

found stromal cells mainly converged in GAC2 in bulk data. B cells, Myeloid cells, and T cells were aggregated in GAC2 and 3. The expression of epithelial cells were in line with the results in scRNA. Remarkably, GAC2 possessed more regulatory T cells (Tregs) (Figures 6A, S3A). In Charoentong's signature, macrophage, T, and B cell expression trends were similar to Lee's observed findings (Figures 6B, S3B, S4A, B). The Cibersort algorithm calculated the expression proportion of each cell. In Lee's cell signatures, the fraction of proliferative ECs was the highest in TME, proliferating macrophage the second, among which GAC2 occupies the lowest and highest, respectively (Figures 6C, S3C). In LM cell signatures, the same phenomenon was seen in macrophages, while T cells were the second proportion of all cells and the lowest in GAC2. It was

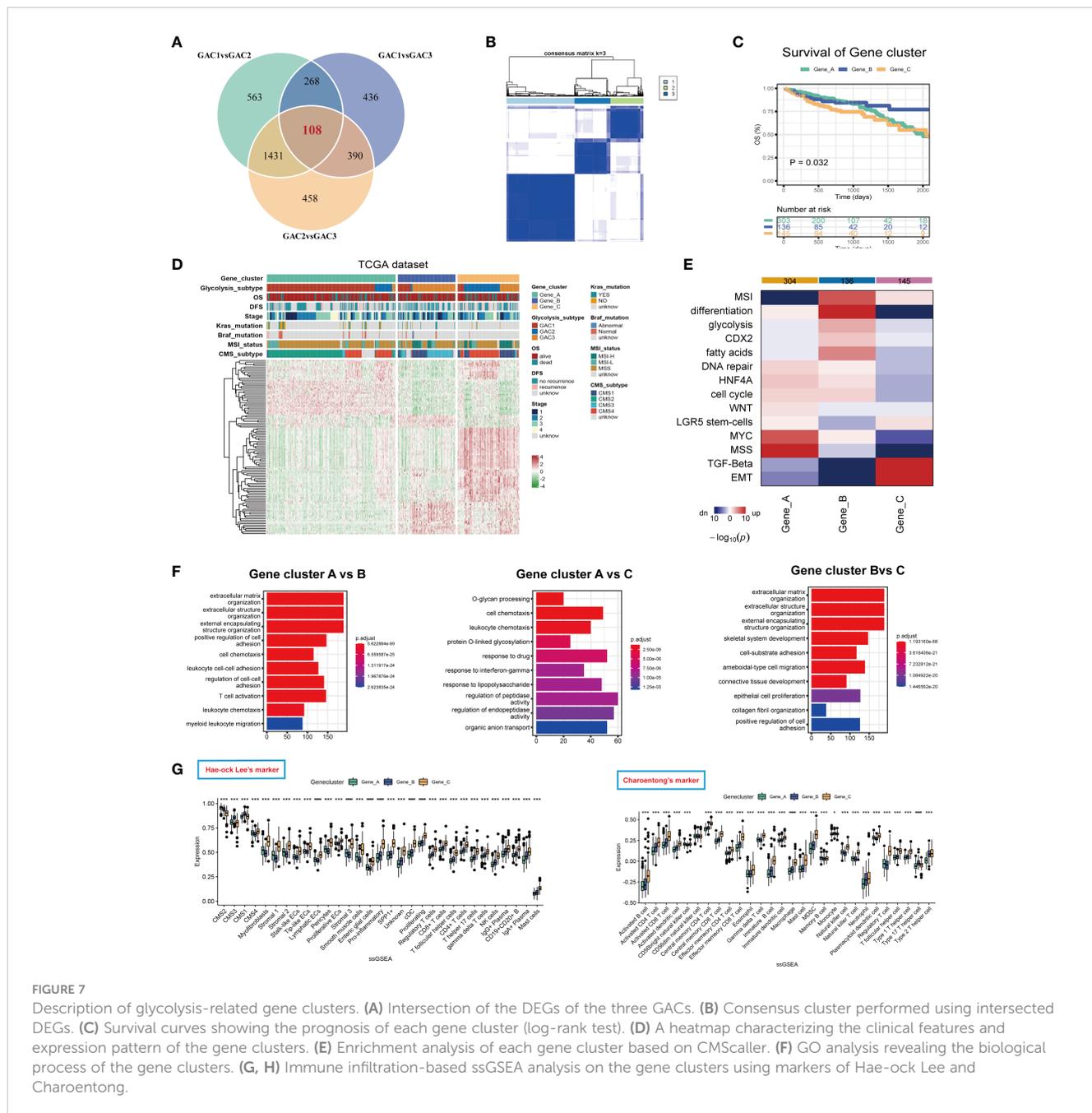


FIGURE 7

Description of glycolysis-related gene clusters. (A) Intersection of the DEGs of the three GACs. (B) Consensus cluster performed using intersected DEGs. (C) Survival curves showing the prognosis of each gene cluster (log-rank test). (D) A heatmap characterizing the clinical features and expression pattern of the gene clusters. (E) Enrichment analysis of each gene cluster based on CMScaller. (F) GO analysis revealing the biological process of the gene clusters. (G, H) Immune infiltration-based ssGSEA analysis on the gene clusters using markers of Hae-ock Lee and Charoentong.

noteworthy that the proportion of M2 macrophage in GAC2 was greater (Figures 6D, S3D). According to the Estimate methodology, GAC2 and GAC3 had higher stromal and immune scores than GAC1 (Figures 6E, S3E). To predict the responses of different GACs to immunotherapy, we collected MSS, dMMR, TMB, and PD-1 information. The outcome was significant: GAC3 had more percentage of MSI-High and dMMR (Figure 6F), and achieved a higher TMB score (Figure 6G). While GAC2 had relatively higher expression of PD-1, PDL-1, and CTLA-4 (Figure 6H–J, S3F, H). The correlations of Lee’s cells were presented in Figures 6K and S3I, in which we showed the relationship between glycolysis and each cell. We concluded the same results in the GEO dataset (Figure S3). GAC1 had the least immune cell infiltration, GAC2 possessed a higher amount of

stromal and immune cells but also a large proportion of tumor-promoting cells (Tregs, M2 macrophage), and GAC3 had a higher immune cell infiltration. Summarily, GAC1 was comparable to the immune-desert type, GAC2 to an immune-inflamed/excluded type, and GAC3 to an immune-activated type.

### Identification and description of gene clusters

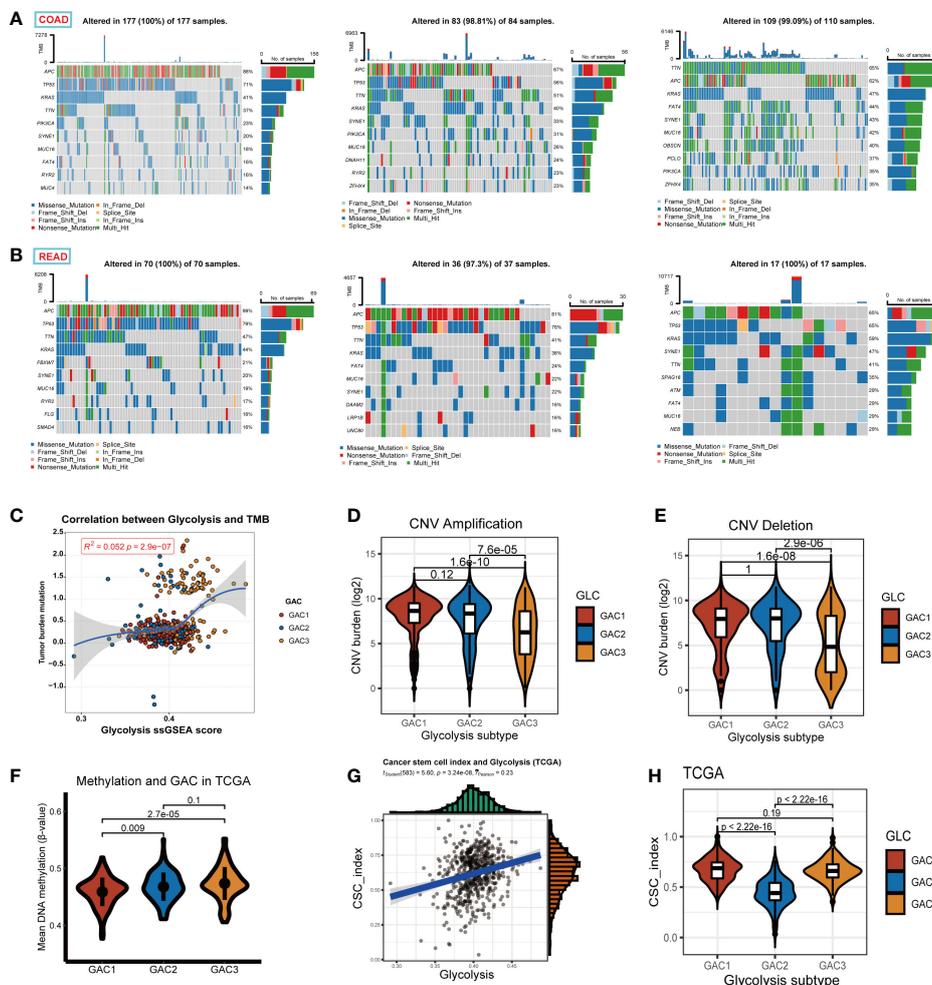
We identified gene clusters to investigate the intrinsic causes for the existence of DEGs between GACs (Table S6). To obtain the final

DEGs, the intersection of the DEGs between each pair of GACs was employed (Figure 7A, Table S5). Then Three clusters, referred to as gene clusters, were obtained after 108 genes were submitted to consensus clustering (Figure 7B). In terms of clinical characteristics, we discovered that gene cluster B had a significant prevalence of GAC2 and a bad prognosis (Figure 7C). These 108 genes' clustering expression patterns and additional clinical traits were displayed in Figure 7D, showing discrepancies among gene clusters. In terms of enrichment analysis, gene cluster A has upregulated MSS and MYC. In gene cluster B, differentiation, MSI, glycolysis, and fatty acids were highly enriched. TGF-Beta and EMT were activated in gene cluster C (Figure 7E). According to GO analysis, gene clusters B and C were linked to extracellular matrix synthesis and epithelial cell proliferation, whereas gene cluster A was linked to T cell activation and O-glycan process (Figure 7F). We discovered that the gene cluster C group has strong expression in the majority of TME cells using two ssGSEA signatures. Immune cells from gene cluster B were relatively activated (Figures 7G, H). All considered, the

clinical, biochemical, and immunological features of GACs were supported by gene clustering.

### Exploration of TMB, CNV, methylation, and CSC index between GACs

This section focused on the investigation of tumor stemness and genomics. We displayed the first 10 mutant genes of each GAC in COAD and READ using waterfall plots. In COAD, the mutation of GAC1 and GAC2 were similar, while in GAC3 we did not find a high mutation rate of TP53. In READ, we detected a high mutation rate of FATA4, especially in GAC2. APC, TP53, TTN, and KRAS were the genes with the top mutation frequency (Figures 8A, B). Besides, we noticed that the TMB score and glycolysis score were favorably connected and GAC3 has high levels of TMB and glycolysis attributes (Figure 8C). Meanwhile, GAC3 possessed a lower copy number variation, which indicated an opposite of pro-



**FIGURE 8** Exploration of the properties of each GAC at genomic level. (A, B) Waterfall Plots revealing the genes with top 10 mutation frequency in each GAC of COAD and READ. (C) The relationship among glycolysis score, tumor mutation burden score, and GACs. (D, E) Copy number variation (amplification, deletion, log2 transformed) of each GAC. (F) The relationship between methylation and GACs in TCGA dataset. Methylation was calculated in the form of the mean value of B-value. (G) The relationship between glycolysis score and CSC index ( $r = 0.23$ ,  $p < 0.001$ ). (H) Violin chart representing the association between GACs and CSC-index.

tumor effect (Figures 8D, E) (42). The overall methylation of GAC1 was lower than the other two (Figure 8F). Through CNV calculation, we found CSC index was in a positive correlation with the glycolysis score, with GAC2 occupying the lowest CSC index (Figures 8G, H). Taken together, we recognized GAC1 as a low methylation type, GAC2 as a low CSC type, and GAC3 as a low CNV and high TMB type.

### Machine learning-based construction of GAC predictor and description of GAC Predictor-related genes (GP genes)

In this part, we constructed a prediction model for the GAC status of each patient using four machine-learning techniques. We referred to TCGA as the training set and GEO as the validation set.

The whole process was shown in Figure 9A. To ensure the reliability of the results, we used clinically significant genes at both bulk and single-cell levels for subsequent analyses. Thus, the intersection of GAC DEGs and marker genes of significant cell subtypes (Figures 4G, H) was obtained. To retain the genes that have the most influence on the prediction results, we used Lasso, SVM, Randomforest (Boruta), and XGBoost for screening (Figure S5; Table S7). The radar graphic demonstrated that the four methods' AUC values in the test set and training set were high. We again intersected the genes filtered by the four machine-learning methods to acquire a final gene set, named GP genes. Subsequently, GP genes were used for multinomial logistic regression to construct the final GAC predictor (Table S8). With the GAC predictor applied for verification, in both training and validation sets we gained high levels of AUC and accuracy values (training set: AUC = 0.9984, accuracy = 0.9658; validation set: AUC = 0.9380, accuracy = 0.8114)

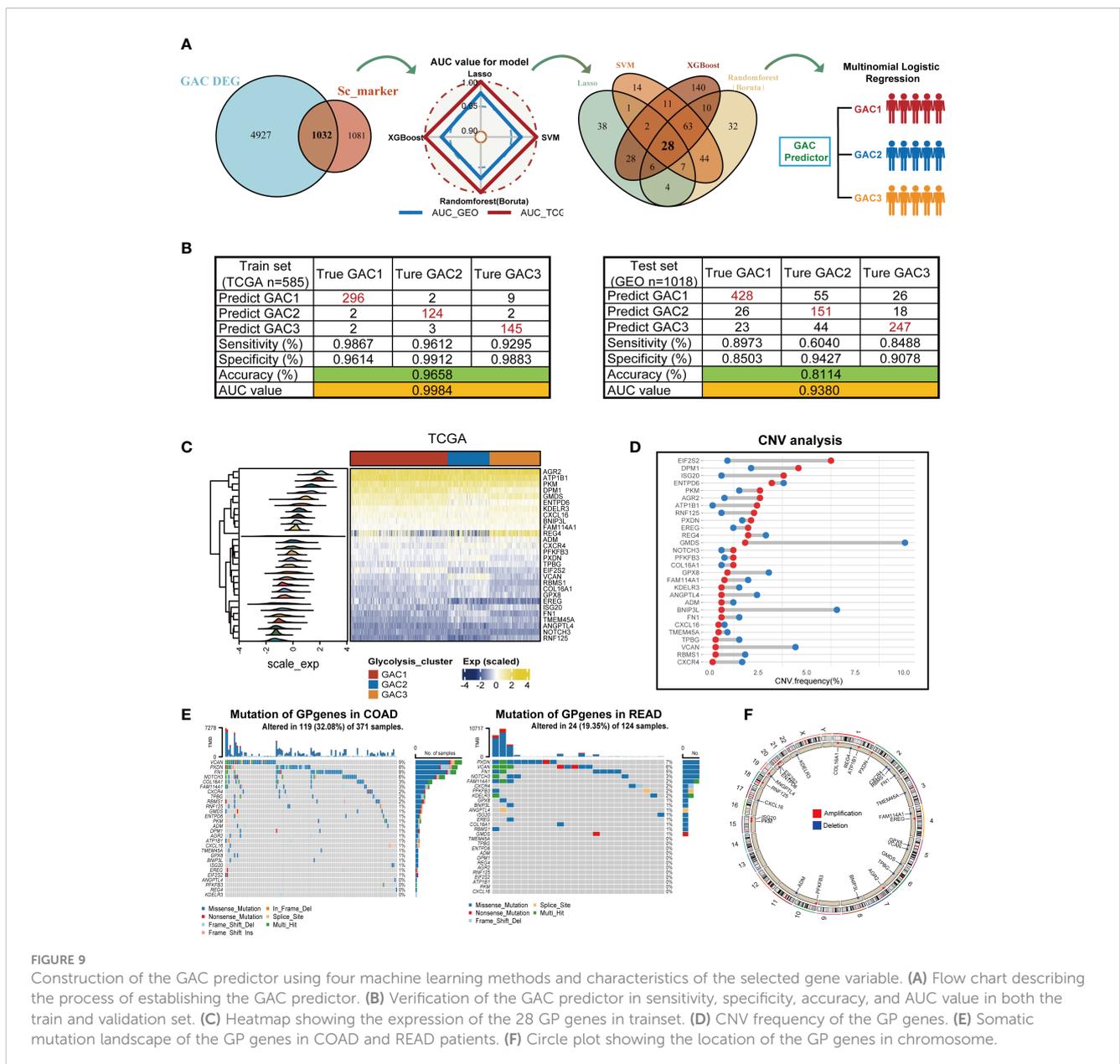


FIGURE 9

Construction of the GAC predictor using four machine learning methods and characteristics of the selected gene variable. (A) Flow chart describing the process of establishing the GAC predictor. (B) Verification of the GAC predictor in sensitivity, specificity, accuracy, and AUC value in both the train and validation set. (C) Heatmap showing the expression of the 28 GP genes in trainset. (D) CNV frequency of the GP genes. (E) Somatic mutation landscape of the GP genes in COAD and READ patients. (F) Circle plot showing the location of the GP genes in chromosome.

(Figure 9B). We further conducted several analyses for GP genes. The heatmap showed the expression of GP genes in order (Figure 9C). Genomics-related research on GP genes, including TMB and CNV analyses, was performed. We discovered EIF2S2, DPM1, and ISG20 as CNV amplification, GMD5, BNIP3L, and VCAN as CNV deletion (Figure 9D). All GP genes were presented in the chromosome loop diagram (Figure 9F). By TMB analysis, we identified VCAN, PXDN, FN1, and NOTCH3 as genes with high mutation frequency in both COAD and READ (Figure 9E).

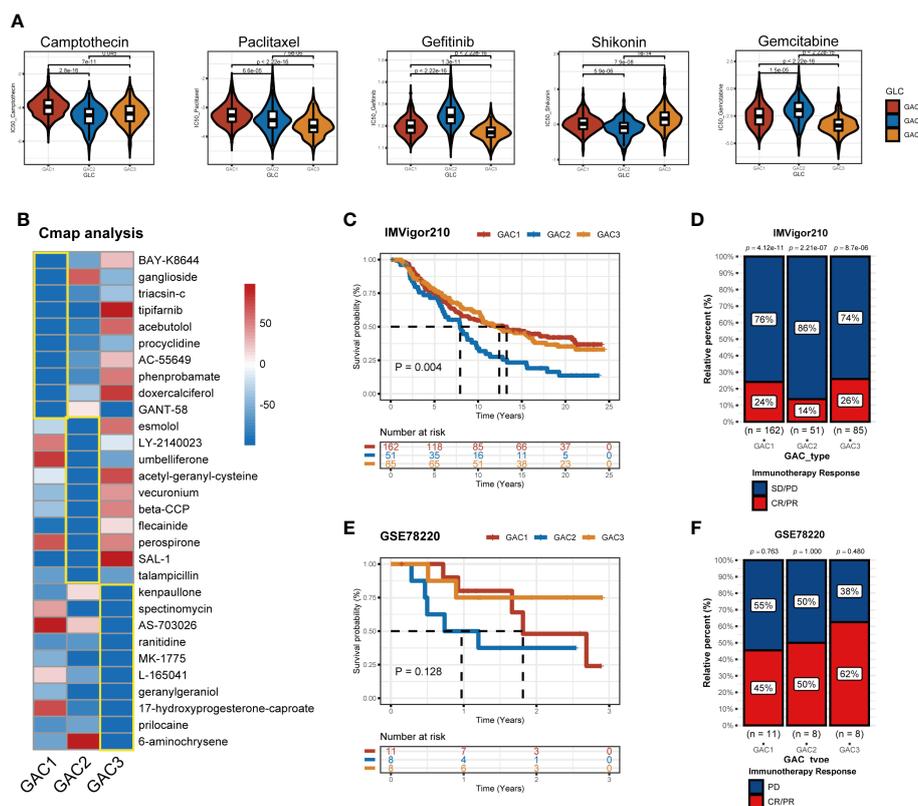
### Treatment strategies for each GAC and immunotherapy cohorts for GAC predictor validation

Chemotherapy and targeted medications increase the survival rate of colorectal cancer with advanced stages. However, some patients are still achieving little benefits or developing treatment resistance (43). Hence it is necessary to formulate a personalized drug treatment plan for each CRC patient. Based on the significant clinical and biological differences among the three GAC, we assumed that each GAC had its own appropriate treatment and then investigated the sensitivity and antagonistic effects of typical chemotherapeutic medicines as well as possible small molecule medications in three GACs. The “pRRophetic” package and

Cmap analysis were used to implement the treatment prediction for each GAC. Drug sensitivity, in the form of IC50, was produced as Figure 10A showed. For GAC1, camptothecin and paclitaxel were more sensitive. Gefitinib and gemcitabine had higher sensitivity to GAC2. Shikonin had more therapeutic potential for GAC3. Additionally, we conducted Cmap analysis to acquire potential compounds resisting each GAC. Stronger compounds’ efficacy resulted from lower scores. 30 compounds were exhibited in Figure 10B to provide treatment strategies. To test the role of GAC clustering in the immunotherapy cohorts, we obtained two datasets for verification (IMVigor210 and GSE78220). With the GAC predictor applied In IMVigor210, we found GAC2 had a poor prognosis (p = 0.004, log-rank test), along with immunotherapy response (Figure 10C, D). In GSE78220, the same conclusion was achieved that GAC2 was with unfavorable OS rate (p = 0.128, log-rank test). Even though GSE78220’s p-value is not significant, each GAC’s survival pattern is clearly visible (Figure 10E, F). Overall, we explored GAC’s potential compounds and validated the GAC predictor in the field of immunotherapy.

### Discussion

As research into cancer has progressed in depth, it has also been discovered that cancer possesses aberrant metabolic characteristics.



**FIGURE 10** Application of GACs in treatment area. (A) IC50 of several common chemotherapeutic drugs in each GAC. (B) Cmap analysis revealing potential small compounds for the treatment of each GAC. (C, D) Validation of IMVigor210 immunotherapy cohort in prognosis of different GACs. (E, F) Validation of GSE78220 immunotherapy cohort in prognosis of different GACs.

The evidence of the Warburg effect demonstrates that aerobic glycolysis has a significant part to play in tumor growth. An illustration of this is hexokinase 2 (HK2), which is prevalently expressed in tumors and is able to gain access to ATP in the inner mitochondrial membrane through voltage-dependent anion channels, thus enhancing glucose metabolism and preventing cell death (44). Glycolysis with its high rate of output not only produces energy rapidly but also furnishes tumor cells with the necessary components for the assembly of numerous biological macromolecules (7). Lipids, nucleotides, and amino acids, which are converted from glycolysis, provide the raw materials for cell growth and division (45, 46). Furthermore, glycolysis is associated with numerous carcinogenic signaling pathways and genetic mutations. Abnormal activity of the Wnt signaling may interfere with the activity of Pyruvate Dehydrogenase Kinase 1 (PDK1), thus preventing the connection between glycolysis and the TCA cycle, consequently hampering OXPHOS (47). HK2, a key enzyme of glycolysis, can be induced by p53. In the absence of glucose, mutant p53 can reduce AMPK signaling, resulting in an increase in aerobic glycolysis (48). Additionally, the focus is gradually being directed to the metabolic remodeling of TME. Pyruvate is converted by glycolysis to lactate, which makes the tumor microenvironment acidic. The immunological effectiveness of NK cells is reduced by this acidic environment (49). During activation, CD8T cells, NK cells, and M1 cells all display significant glycolytic characteristics, whereas Treg cells and M2 cells favor the usage of OXPHOS (50, 51). According to these findings, tumors may have aberrant energy consumption habits that help them thrive ecologically and impair immunotherapy (52). As an important promoter of tumor progression, glucose metabolism has been extensively exploited to overcome the bottleneck of immunotherapy (53). Glycolysis-targeted drugs in combination with immunotherapies have shown to be highly effective in CRC treatment, according to growing preclinical evidence (54–56). All these demonstrate how important glycolysis is for the study and treatment of colorectal cancer.

In this study, we utilized unsupervised clustering to identify three clusters of CRC patients, termed as GACs, based on the expression of genes related to glycolysis. The three GACs had entirely different clinics, biological processes, immune infiltration, and genomic traits when viewed from bulk-seq perspective. Bioinformatics methods were used to assign a distinct GAC composition to each cell group from a single-cell view. It was discovered that each cell type with a distinct GAC type possessed its unique traits. Finally, we built the GAC predictor by integrating bulk and single-cell RNA-seq data to identify the GAC attribution of each CRC patient.

Initially, when analyzing bulk RNA-seq data, we discovered that genes related to glycolysis were capable of distinguishing between tumor and normal samples. We then divided these genes into three glycolysis-associated clusters *via* unsupervised clustering algorithm in CRC patients. Different characteristics of various GACs were revealed by multiple enrichment analyses (GSVA, ssGSEA). From the perspective of biological pathways, GAC1 tended to be active in cell cycle and replication, and its relatively elevated WNT and MYC pathways were consistent with the CMS2 subtype, as evidenced by

its high proportion of the CMS2 subtype. GAC2 had the greatest CMS4 ratio, a tendency to be more stromal-type, and was heavily enriched in EMT and TGF-Beta pathways. With greater MSI and fatty acid metabolic pathways and high proportion levels of CMS1 and 3 at the same time, GAC3 has a tendency to be immunologically and metabolically enriched (Figures 2I, 3A). Each GAC's enrichment pathway exhibits a high degree of concordance with the enrichment pathway of the GAC-associated CMS subtype described in the previous article (57). From a clinical perspective, the prognosis of GAC2 was worse while that of GAC3 was better, both of which were consistent with the corresponding CMS subtypes. From the perspective of the TME, we used algorithms like ssGSEA, cibersort, and ESTIMATE to investigate the variations in the immune microenvironment of GACs. Similar to the immune desert type, GAC1 is obviously devoid of immune cell infiltration. In contrast, GAC2 and GAC3 exhibit significant immune infiltration properties. GAC2 demonstrated a condition of stromal infiltration, and more tumor-promoting immune cells (Treg, M2 macrophage) were present, indicating that it might be the immune-inflamed/excluded type. A Study revealed that M2 macrophages can promote the invasion and metastasis of CRC cells by releasing exosomes that carry mi-RNA (58), This supported the opinion in the study that GAC2, which was infiltrated by M2 macrophages, enriched in metastasis-related pathways, such as EMT and TGF-Beta. GAC3, however, was more like the immune-activated type. Comparable to Emilie Picard's article (59), this result mapped the immune microenvironment of each GAC. From a genomics perspective, APC mutations were more prevalent in GAC1, whereas TTN mutations were more prevalent in GAC3 (Figures 8A, B). GAC1 resembled a classical type more due to the accumulation of the APC/KRAS/TP53 mutation (60). High TTN mutation rate was related to high TMB status (61), which GAC3 both possessed (Figure 8C). In terms of prospects for immunotherapy, GAC3 demonstrated a more favorable reaction than GAC1 and 2. Clinical trials have demonstrated that immunotherapy is effective for MSI-H/dMMR CRC patients (62, 63). Particularly, in GAC3 there existed simultaneous MSI-H/dMMR status and better prognosis, the correlation between which has been verified in many clinical studies and meta-analyses (64, 65), that is, MSI status has a better prognosis than MSS status in stage II/III colorectal cancer, attributed to the lower probability of recurrence of MSI status (66). When compared to GAC1, GAC2 and 3 expressed more PD-1, PDL-1, and CTLA-4, while GAC3 had more MSI-H, and dMMR phenotypes concurrently (Figures 6F, G), indicating that GAC3 had immunotherapeutic potential. Meanwhile, our results supported the presence of elevated immune checkpoint molecule expression in CMS1 and 4 (57, 67, 68). As expected, GAC3 showed a favorable prognosis in the immunotherapy cohort of bladder and melanoma cancer (Figure 10). From the perspective of therapeutic drugs, shikonin had a high sensitivity to GAC3 and a study has demonstrated that it inhibits the proliferation of colorectal tumor cells when used as a treatment (69). Meanwhile, gemcitabine was more effective against GAC2, and the fact that it can cause apoptosis in oxaliplatin-resistant cells further supports its potential to act as an inhibitor of colorectal cancer (70).

When analyzing scRNA-seq data, we uncovered that glycolysis predominantly occurred in tumor cells, stromal cells, and myeloid cells (Figures 4A, C), which aligned with the findings of prior research (71, 72). Interestingly, our investigations indicated that stromal cells mainly belong to the GAC2 subtype, while tumor cells were predominantly GAC1 and GAC3 (Figures 4B, E), affirming our previous bulk RNA-seq study's determination that GAC2 was of stromal-type origin. The stroma, where cancer-associated fibroblasts interact with cancer cells to promote invasion, metastasis, EMT, and drug resistance, is crucial in the development of cancer (73). Since there are only two GAC-like subtypes of tumor cells, we investigated this more thoroughly and used enrichment analysis (GO and GSEA) to discover that GAC1-like tumor cells had a higher propensity for division and replication while GAC3-like tumor cells were more likely to invade and adhere. Pseudo-time analysis showed a possible transition of tumors from GAC3-like to GAC1-like, which was worth exploring. Besides, cell communication analysis found significant interactions between myeloid and stromal cells on tumor cells, with stronger signaling from receptor-ligands of SPPI-CD44, MDK-NCL, and MDK-SDC4. Glioma, pancreatic cancer, and intrahepatic cholangiocarcinoma have been reported to possess the SPPI-CD44 axis that is immunosuppressive and pro-tumor (74–76). Although it was discovered that midkine (MDK) was increased in colorectal cancer (77) and that nucleolin (NCL) was linked to DNA and RNA metabolism and proliferation (78), the precise relationship between the two has not yet undergone experimental verification and merits more investigation. Syndecan-4 (SDC4) is connected to a poor prognosis and the invasion of colorectal cancer cells. All these scRNA-related results enhanced the conclusion derived from bulk RNA-seq data and excavated the character of tumor cells of different GAC-like subtypes.

Meanwhile, our study has some limitations. First, the data were based on retrospective studies obtained from public databases, and there was a lack of validation of a prospective cohort originating from our research center. Second, only bioinformatics algorithms were used to anticipate rather than experimentally validate the biological properties of tumor cells with different GAC-like subtypes. Additionally, the sample size used in this study was insufficient, therefore the GAC predictors' sensitivity and accuracy needed to be enhanced by adding samples and validation.

## Conclusions

In summary, we identified three glycolysis-related molecular subtypes of CRC, whose characteristics of the clinic, genomics, and immune infiltration were discussed based on the integration of bulk and single-cell RNA-seq data. We simultaneously created predictive models to forecast the GAC subtype in each sample and offered medication suggestions for treatment.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

## Author contributions

ZW, YS and HZ contributed to the study equally. ZF, ZW and YS conceived and designed the manuscript. ZW, HZ and YL wrote the manuscript. ZW, CH, JW, YL, YC, HZ and YS analyzed the data and drew the figures. ZF and HS helped with the manuscript and data review. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Natural Science Foundation of China (81470881 to ZF, 82172956 to ZF) and Jiangsu Commission of Health (LGY2017031 and BRA2015473 to ZF).

## Acknowledgment

We greatly appreciate the analytical data provided by the TCGA and GEO databases. We also thank Figdraw for its support for the drawing section in Figure 1.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2023.1181985/full#supplementary-material>

### SUPPLEMENTARY FIGURE 1

Consensus clustering of the glycolysis-related genes. (A) The heatmap of consensus clustering with  $k$  from 2 to 8. (B) The CDF threshold curve. (C) The delta area.

### SUPPLEMENTARY FIGURE 2

The validation of clinical and biological pathway characteristics in GEO datasets. (A) The heatmap comprehensively assessed each GAC's biological and clinical parameters in GEO datasets. (B) The boxplots visualizing each GAC's pathway enrichment scores (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). (C) The relationship between GACs and CMS type in GEO datasets. (D) The

relationship between GACs and stage in GEO datasets. (E) The distribution of glycolysis score in each GAC.

#### SUPPLEMENTARY FIGURE 3

The validation of TME features of each GAC in GEO datasets. (A, B) ssGSEA analyses based on markers of Hae-ock Lee and Charoentong revealing the expression of TME cells of the GACs in GEO datasets. (C, D) Cibersort analysis depicting the percentage of each cell in GACs based on markers of Hae-ock Lee and LM 22 cells in GEO datasets (\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ). (E) Estimate algorithm calculating stromal, immune and overall score of the GACs in GEO datasets. (F–H) The expression of PD1, PDL1, and CTLA4 in each GAC in GEO datasets. (I) The relationship between TME cells and glycolysis score.

## References

- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013
- Warburg O. On the origin of cancer cells. *Science* (1956) 123(3191):309–14. doi: 10.1126/science.123.3191.309
- Weinhouse S. On respiratory impairment in cancer cells. *Science* (1956) 124(3215):267–9. doi: 10.1126/science.124.3215.267
- Ganapathy-Kanniappan S, Geschwind JF. Tumor glycolysis as a target for cancer therapy: progress and prospects. *Mol Cancer* (2013) 12:152. doi: 10.1186/1476-4598-12-152
- DeBerardinis RJ, Lum JJ, Hatzivassiliou G, Thompson CB. The biology of cancer: metabolic reprogramming fuels cell growth and proliferation. *Cell Metab* (2008) 7(1):11–20. doi: 10.1016/j.cmet.2007.10.002
- Jones RG, Thompson CB. Tumor suppressors and cell metabolism: a recipe for cancer growth. *Genes Dev* (2009) 23(5):537–48. doi: 10.1101/gad.1756509
- Vander Heiden MG, Cantley LC, Thompson CB. Understanding the warburg effect: the metabolic requirements of cell proliferation. *Science* (2009) 324(5930):1029–33. doi: 10.1126/science.1160809
- Feron O. Pyruvate into lactate and back: from the warburg effect to symbiotic energy fuel exchange in cancer cells. *Radiother Oncol* (2009) 92(3):329–33. doi: 10.1016/j.radonc.2009.06.025
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68(6):394–424. doi: 10.3322/caac.21492
- Nenkov M, Ma Y, Gassler N, Chen Y. Metabolic reprogramming of colorectal cancer cells and the microenvironment: implication for therapy. *Int J Mol Sci* (2021) 22(12):6262. doi: 10.3390/ijms22126262
- Abdullah M, Rani AA, Simadibrata M, Fauzi A, Syam AF. The value of fecal tumor M2 pyruvate kinase as a diagnostic tool for colorectal cancer screening. *Acta Med Indones* (2012) 44(2):94–9.
- Altenberg B, Greulich KO. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* (2004) 84(6):1014–20. doi: 10.1016/j.ygeno.2004.08.010
- Chan EC, Koh PK, Mal M, Cheah PY, Eu KW, Backshall A, et al. Metabolic profiling of human colorectal cancer using high-resolution magic angle spinning nuclear magnetic resonance (HR-MAS NMR) spectroscopy and gas chromatography mass spectrometry (GC/MS). *J Proteome Res* (2009) 8(1):352–61. doi: 10.1021/pr8006232
- Haber RS, Rathan A, Weiser KR, Pritsker A, Itzkowitz SH, Bodian C, et al. GLUT1 glucose transporter expression in colorectal carcinoma: a marker for poor prognosis. *Cancer* (1998) 83(1):34–40. doi: 10.1002/(SICI)1097-0142(19980701)83:1<34::AID-CNCR5>3.0.CO;2-E
- Izuishi K, Yamamoto Y, Sano T, Takebayashi R, Nishiyama Y, Mori H, et al. Molecular mechanism underlying the detection of colorectal cancer by 18F-2-fluoro-2-deoxy-D-glucose positron emission tomography. *J Gastrointest Surg* (2012) 16(2):394–400. doi: 10.1007/s11605-011-1727-z
- Jimenez B, Mirnezami R, Kinross J, Cloarec O, Keun HC, Holmes E, et al. 1H HR-MAS NMR spectroscopy of tumor-induced local metabolic “field-effects” enables colorectal cancer staging and prognostication. *J Proteome Res* (2013) 12(2):959–68. doi: 10.1021/pr3010106
- Koukourakis MI, Giatromanolaki A, Simopoulos C, Polychronidis A, Sivridis E. Lactate dehydrogenase 5 (LDH5) relates to up-regulated hypoxia inducible factor pathway and metastasis in colorectal cancer. *Clin Exp Metastasis* (2005) 22(1):25–30. doi: 10.1007/s10585-005-2343-7
- Lee M, Yoon JH. Metabolic interplay between glycolysis and mitochondrial oxidation: the reverse warburg effect and its therapeutic implication. *World J Biol Chem* (2015) 6(3):148–61. doi: 10.4331/wjbc.v6.i3.148
- Chekulayev V, Mado K, Shevchuk I, Koit A, Kaldma A, Klepinin A, et al. Metabolic remodeling in human colorectal cancer and surrounding tissues: alterations in regulation of mitochondrial respiration and metabolic fluxes. *Biochem Biophys Res* (2015) 4:111–25. doi: 10.1016/j.bbrep.2015.08.020
- Li F, Simon MC. Cancer cells don't live alone: metabolic communication within tumor microenvironments. *Dev Cell* (2020) 54(2):183–95. doi: 10.1016/j.devcel.2020.06.018
- Migneco G, Whitaker-Menezes D, Chiavarina B, Castello-Cros R, Pavlides S, Pestell RG, et al. Glycolytic cancer associated fibroblasts promote breast cancer tumor growth, without a measurable increase in angiogenesis: evidence for stromal-epithelial metabolic coupling. *Cell Cycle* (2010) 9(12):2412–22. doi: 10.4161/cc.9.12.11989
- Chiavarina B, Whitaker-Menezes D, Martinez-Outschoorn UE, Witkiewicz AK, Birbe R, Howell A, et al. Pyruvate kinase expression (PKM1 and PKM2) in cancer-associated fibroblasts drives stromal nutrient production and tumor growth. *Cancer Biol Ther* (2011) 12(12):1101–13. doi: 10.4161/cbt.12.12.18703
- Yin Z, Bai L, Li W, Zeng T, Tian H, Cui J. Targeting T cell metabolism in the tumor microenvironment: an anti-cancer therapeutic strategy. *J Exp Clin Cancer Res* (2019) 38(1):403. doi: 10.1186/s13046-019-1409-3
- Fan A, Wang B, Wang X, Nie Y, Fan D, Zhao X, et al. Immunotherapy in colorectal cancer: current achievements and future perspective. *Int J Biol Sci* (2021) 17(14):3837–49. doi: 10.7150/ijbs.64077
- Ganesh K, Stadler ZK, Cercck A, Mendelsohn RB, Shia J, Segal NH, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* (2019) 16(6):361–75. doi: 10.1038/s41575-019-0126-x
- Park GB, Chung YH, Kim D. 2-Deoxy-D-glucose suppresses the migration and reverses the drug resistance of colon cancer cells through ADAM expression regulation. *Anticancer Drugs* (2017) 28(4):410–20. doi: 10.1097/CAD.0000000000000472
- Cheng G, Zhang Q, Pan J, Lee Y, Ouari O, Hardy M, et al. Targeting lonidamine to mitochondria mitigates lung tumorigenesis and brain metastasis. *Nat Commun* (2019) 10(1):2205. doi: 10.1038/s41467-019-10042-1
- Zhong X, He X, Wang Y, Hu Z, Huang H, Zhao S, et al. Construction of a prognostic glycolysis-related lncRNA signature for patients with colorectal cancer. *Cancer Med* (2022) 12(1):930–48. doi: 10.1002/cam4.4851
- Zhu J, Wang S, Bai H, Wang K, Hao J, Zhang J, et al. Identification of five glycolysis-related gene signature and risk score model for colorectal cancer. *Front Oncol* (2021) 11:588811. doi: 10.3389/fonc.2021.588811
- Cai WY, Dong ZN, Fu XT, Lin LY, Wang L, Ye GD, et al. Identification of a tumor microenvironment-relevant gene set-based prognostic signature and related therapy targets in gastric cancer. *Theranostics* (2020) 10(19):8633–47. doi: 10.7150/tno.47938
- Tang Y, Guo C, Yang Z, Wang Y, Zhang Y, Wang D. Identification of a tumor immunological phenotype-related gene signature for predicting prognosis, immunotherapy efficacy, and drug candidates in hepatocellular carcinoma. *Front Immunol* (2022) 13:862527. doi: 10.3389/fimmu.2022.862527
- Wu Z, Tan J, Zhuang Y, Zhong M, Xiong Y, Ma J, et al. Identification of crucial genes of pyrimidine metabolism as biomarkers for gastric cancer prognosis. *Cancer Cell Int* (2021) 21(1):668. doi: 10.1186/s12935-021-02385-x
- Seiler M, Huang CC, Szalma S, Bhanot G. ConsensusCluster: a software tool for unsupervised cluster discovery in numerical data. *OMICS* (2010) 14(1):109–13. doi: 10.1089/omi.2009.0083
- Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* (2010) 143(3):331–6. doi: 10.1016/j.otohns.2010.05.007
- Lee HO, Hong Y, Etioglu HE, Cho YB, Pomella V, Van den Bosch B, et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat Genet* (2020) 52(6):594–603. doi: 10.1038/s41588-020-0636-z
- Charoentong P, Finotello F, Angelova M, Mayer C, Efreanova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype

- relationships and predictors of response to checkpoint blockade. *Cell Rep* (2017) 18(1):248–62. doi: 10.1016/j.celrep.2016.12.019
37. Kim S. Margin-maximised redundancy-minimised SVM-RFE for diagnostic classification of mammograms. *Int J Data Min Bioinform* (2014) 10(4):374–90. doi: 10.1504/IJDMB.2014.064889
38. Yperman J, Becker T, Valkenburg D, Popescu V, Hellings N, Wijmeersch BV, et al. Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. *BMC Neurol* (2020) 20(1):105. doi: 10.1186/s12883-020-01672-w
39. Li W, Yin Y, Quan X, Zhang H. Gene expression value prediction based on XGBoost algorithm. *Front Genet* (2019) 10:1077. doi: 10.3389/fgene.2019.01077
40. McEligot AJ, Poyner V, Sharma R, Panangadan A. Logistic LASSO regression for dietary intakes and breast cancer. *Nutrients* (2020) 12(9):2652. doi: 10.3390/nu12092652
41. Fujiwara-Tani R, Sasaki T, Ohmori H, Luo Y, Goto K, Nishiguchi Y, et al. Concurrent expression of CD47 and CD44 in colorectal cancer promotes malignancy. *Pathobiology* (2019) 86(4):182–9. doi: 10.1159/000496027
42. Ried T, Meijer GA, Harrison DJ, Grech G, Franch-Exposito S, Briffa R, et al. The landscape of genomic copy number alterations in colorectal cancer and their consequences on gene expression levels and disease outcome. *Mol Aspects Med* (2019) 69:48–61. doi: 10.1016/j.mam.2019.07.007
43. Modest DP, Pant S, Sartore-Bianchi A. Treatment sequencing in metastatic colorectal cancer. *Eur J Cancer* (2019) 109:70–83. doi: 10.1016/j.ejca.2018.12.019
44. Nakashima RA, Mangan PS, Colombini M, Pedersen PL. Hexokinase receptor complex in hepatoma mitochondria: evidence from N,N'-dicyclohexylcarbodiimide-labeling studies for the involvement of the pore-forming protein VDAC. *Biochemistry* (1986) 25(5):1015–21. doi: 10.1021/bi00353a010
45. Counihan JL, Grossman EA, Nomura DK. Cancer metabolism: current understanding and therapies. *Chem Rev* (2018) 118(14):6893–923. doi: 10.1021/acs.chemrev.7b00775
46. Kroemer G, Pouyssegur J. Tumor cell metabolism: cancer's achilles' heel. *Cancer Cell* (2008) 13(6):472–82. doi: 10.1016/j.ccr.2008.05.005
47. El-Sahli S, Xie Y, Wang L, Liu S. Wnt signaling in cancer metabolism and immunity. *Cancers (Basel)* (2019) 11(7):904. doi: 10.3390/cancers11070904
48. Zhou G, Wang J, Zhao M, Xie TX, Tanaka N, Sano D, et al. Gain-of-function mutant p53 promotes cell growth and cancer cell metabolism via inhibition of AMPK activation. *Mol Cell* (2014) 54(6):960–74. doi: 10.1016/j.molcel.2014.04.024
49. Terren I, Orrantia A, Vitale J, Zenarruabeitia O, Borrego F. NK cell metabolism and tumor microenvironment. *Front Immunol* (2019) 10:2278. doi: 10.3389/fimmu.2019.02278
50. Bantug GR, Galluzzi L, Kroemer G, Hess C. The spectrum of T cell metabolism in health and disease. *Nat Rev Immunol* (2018) 18(1):19–34. doi: 10.1038/nri.2017.99
51. O'Neill LA, Kishton RJ, Rathmell J. A guide to immunometabolism for immunologists. *Nat Rev Immunol* (2016) 16(9):553–65. doi: 10.1038/nri.2016.70
52. DePeaux K, Delgoffe GM. Metabolic barriers to cancer immunotherapy. *Nat Rev Immunol* (2021) 21(12):785–97. doi: 10.1038/s41577-021-00541-y
53. Bader JE, Voss K, Rathmell JC. Targeting metabolism to improve the tumor microenvironment for cancer immunotherapy. *Mol Cell* (2020) 78(6):1019–33. doi: 10.1016/j.molcel.2020.05.034
54. Huang T, Feng Q, Wang Z, Li W, Sun Z, Wilhelm J, et al. Tumor-targeted inhibition of monocarboxylate transporter 1 improves T-cell immunotherapy of solid tumors. *Adv Healthc Mater* (2021) 10(4):e2000549. doi: 10.1002/adhm.202000549
55. Lei J, Yang Y, Lu Z, Pan H, Fang J, Jing B, et al. Taming metabolic competition via glycolysis inhibition for safe and potent tumor immunotherapy. *Biochem Pharmacol* (2022) 202:115153. doi: 10.1016/j.bcp.2022.115153
56. Yu YR, Imrichova H, Wang H, Chao T, Xiao Z, Gao M, et al. Disturbed mitochondrial dynamics in CD8(+) TILs reinforce T cell exhaustion. *Nat Immunol* (2020) 21(12):1540–51. doi: 10.1038/s41590-020-0793-3
57. Guinney J, Dienstmann R, Wang X, de Reynies A, Schlicker A, Soneson C, et al. The consensus molecular subtypes of colorectal cancer. *Nat Med* (2015) 21(11):1350–6. doi: 10.1038/nm.3967
58. Lan J, Sun L, Xu F, Liu L, Hu F, Song D, et al. M2 macrophage-derived exosomes promote cell migration and invasion in colon cancer. *Cancer Res* (2019) 79(1):146–58. doi: 10.1158/0008-5472.CAN-18-0014
59. Picard E, Verschoor CP, Ma GW, Pawelec G. Relationships between immune landscapes, genetic subtypes and responses to immunotherapy in colorectal cancer. *Front Immunol* (2020) 11:369. doi: 10.3389/fimmu.2020.00369
60. Li J, Ma X, Chakravarti D, Shalpour S, DePinho RA. Genetic and biological hallmarks of colorectal cancer. *Genes Dev* (2021) 35(11-12):787–820. doi: 10.1101/gad.348226.120
61. Oh JH, Jang SJ, Kim J, Sohn I, Lee JY, Cho EJ, et al. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *NPJ Genom Med* (2020) 5:33. doi: 10.1038/s41525-019-0107-6
62. Asaoka Y, Ijichi H, Koike K. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med* (2015) 373(20):1979. doi: 10.1056/NEJMc1510353
63. Overman MJ, Lonardi S, Wong KYM, Lenz HJ, Gelsomino F, Aglietta M, et al. Durable clinical benefit with nivolumab plus ipilimumab in DNA mismatch repair-Deficient/Microsatellite instability-high metastatic colorectal cancer. *J Clin Oncol* (2018) 36(8):773–9. doi: 10.1200/JCO.2017.76.9901
64. Cohen R, Taieb J, Fiskum J, Yothers G, Goldberg R, Yoshino T, et al. Microsatellite instability in patients with stage III colon cancer receiving fluoropyrimidine with or without oxaliplatin: an ACCENT pooled analysis of 12 adjuvant trials. *J Clin Oncol* (2021) 39(6):642–51. doi: 10.1200/JCO.20.01600
65. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. *J Clin Oncol* (2005) 23(3):609–18. doi: 10.1200/JCO.2005.01.086
66. Hutchins G, Southward K, Handley K, Magill L, Beaumont C, Stahlschmidt J, et al. Value of mismatch repair, KRAS, and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. *J Clin Oncol* (2011) 29(10):1261–70. doi: 10.1200/JCO.2010.30.1366
67. Becht E, de Reynies A, Giraldo NA, Pilati C, Buttard B, Lacroix L, et al. Immune and stromal classification of colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clin Cancer Res* (2016) 22(16):4057–66. doi: 10.1158/1078-0432.CCR-15-2879
68. Karpinski P, Rossowska J, Sasiadek MM. Immunological landscape of consensus clusters in colorectal cancer. *Oncotarget* (2017) 8(62):105299–311. doi: 10.18632/oncotarget.22169
69. Shi W, Men L, Pi X, Jiang T, Peng D, Huo S, et al. Shikonin suppresses colon cancer cell growth and exerts synergistic effects by regulating ADAM17 and the IL-6/STAT3 signaling pathway. *Int J Oncol* (2021) 59(6):99. doi: 10.3892/ijo.2021.5279
70. Chocry M, Leloup L, Parat F, Messe M, Pagano A, Kovacic H. Gemcitabine: an alternative treatment for oxaliplatin-resistant colorectal cancer. *Cancers (Basel)* (2022) 14(23):5894. doi: 10.3390/cancers14235894
71. Van den Bossche J, O'Neill LA, Menon D. Macrophage immunometabolism: where are we (Going)? *Trends Immunol* (2017) 38(6):395–406. doi: 10.1016/j.it.2017.03.001
72. Zhang D, Wang Y, Shi Z, Liu J, Sun P, Hou X, et al. Metabolic reprogramming of cancer-associated fibroblasts by IDH3alpha downregulation. *Cell Rep* (2015) 10(8):1335–48. doi: 10.1016/j.celrep.2015.02.006
73. Asif PJ, Longobardi C, Hahne M, Medema JP. The role of cancer-associated fibroblasts in cancer invasion and metastasis. *Cancers (Basel)* (2021) 13(18):4720. doi: 10.3390/cancers13184720
74. Cheng M, Liang G, Yin Z, Lin X, Sun Q, Liu Y. Immunosuppressive role of SPP1-CD44 in the tumor microenvironment of intrahepatic cholangiocarcinoma assessed by single-cell RNA sequencing. *J Cancer Res Clin Oncol* (2022). doi: 10.1007/s00432-022-04498-w
75. He C, Sheng L, Pan D, Jiang S, Ding L, Ma X, et al. Single-cell transcriptomic analysis revealed a critical role of SPP1/CD44-mediated crosstalk between macrophages and cancer cells in glioma. *Front Cell Dev Biol* (2021) 9:779319. doi: 10.3389/fcell.2021.779319
76. Nallasamy P, Nimmakayala RK, Karmakar S, Leon F, Seshacharyulu P, Lakshmanan I, et al. Pancreatic tumor microenvironment factor promotes cancer stemness via SPP1-CD44 axis. *Gastroenterology* (2021) 161(6):1998–2013.e7. doi: 10.1053/j.gastro.2021.08.023
77. Krzystek-Korpacka M, Gorska S, Diakowska D, Kapturkiewicz B, Podkowik M, Gamian A, et al. Midkine is up-regulated in both cancerous and inflamed bowel, reflecting lymph node metastasis in colorectal cancer and clinical activity of ulcerative colitis. *Cytokine* (2017) 89:68–75. doi: 10.1016/j.cyto.2016.09.020
78. Jia W, Yao Z, Zhao J, Guan Q, Gao L. New perspectives of physiological and pathological functions of nucleolin (NCL). *Life Sci* (2017) 186:1–10. doi: 10.1016/j.lfs.2017.07.025