



Exploring the glycolytic cross-talk genes between inflammatory bowel disease and colorectal cancer

Chenglin Ye¹ · Yabing Huang¹ · Yuan Gao¹ · Sizhe Zhu² · Jingping Yuan¹

Received: 23 March 2023 / Revised: 3 July 2023 / Accepted: 4 July 2023 / Published online: 10 July 2023
© The Author(s) 2023

Abstract

Patients with inflammatory bowel disease (IBD) have a higher risk of developing colorectal cancer (CRC). Glycolysis is involved in the development of both IBD and CRC. However, the mechanisms and outcomes of glycolysis shared between IBD and CRC remain unclear. This study aimed to explore the glycolytic cross-talk genes between IBD and CRC integrating bioinformatics and machine learning. With WGCNA, LASSO, COX, and SVM-RFE algorithms, P4HA1 and PMM2 were identified as glycolytic cross-talk genes. The independent risk signature of P4HA1 and PMM2 was constructed to predict the overall survival rate of patients with CRC. The risk signature correlated with clinical characteristics, prognosis, tumor microenvironment, immune checkpoint, mutants, cancer stemness, and chemotherapeutic drug sensitivity. CRC patients with high risk have increased microsatellite instability, tumor mutation burden. The nomogram integrating risk score, tumor stage, and age showed high accuracy for predicting overall survival rate. In addition, the diagnostic model for IBD based on P4HA1 and PMM2 showed excellent accuracy. Finally, immunohistochemistry results showed that P4HA1 and PMM2 were significantly upregulated in IBD and CRC. Our study reveals the presence of glycolytic cross-talk genes P4HA1 and PMM2 between IBD and CRC. This may prove to be beneficial in advancing research on the mechanism of development of IBD-associated CRC.

Keywords Inflammatory bowel disease · Colorectal cancer · Glycolysis · Cross-talk genes

Introduction

Inflammatory bowel disease (IBD) is one of the most prevalent diseases of the digestive system and includes two major subtypes, Crohn's disease and ulcerative colitis (Hnatyszyn et al. 2019, Ng et al. 2017). Long-standing intestinal inflammation in patients with IBD disrupts the normal intestinal structure and leads to the dysregulation of the intestinal mucosal immune system (Hnatyszyn, Hryhorowicz, Kaczmarek-Rys, Lis, Slomski, Scott and Plawski 2019).

Colorectal cancer (CRC) is the most common IBD-associated cancer. Patients with IBD have a 1.7-fold increased risk of developing CRC (Nadeem et al. 2020). CRC is also the most commonly diagnosed gastrointestinal cancer and ranks third in cancer-related deaths worldwide (Arnold et al. 2020, Biller and Schrag 2021, Kim et al. 2021). Although the incidence of CRC in China is lower than that in Western countries, China ranks first in the number of new cases of CRC and related deaths in the world owing to its comparatively large population (Yang et al. 2020). The association between IBD and CRC is unequivocal, but the mechanisms involved remain unclear.

Glycolysis is linked to the inflammatory response and immunoregulation. Glycolysis is involved in the development of several inflammation-related diseases, such as IBD (Riffelmacher et al. 2021), Alzheimer's disease (Hipkiss 2019), SARS-CoV-2 infection (Duan et al. 2021), and rheumatoid arthritis (Abboud et al. 2018). Cancer development involves cell metabolic reprogramming, mainly via the Warburg effect, which is also termed aerobic glycolysis (Mao et al. 2022, Warburg 1956).

✉ Sizhe Zhu
zhusizhe@hust.edu.cn

✉ Jingping Yuan
yuanjingping@whu.edu.cn

¹ Department of Pathology, Renmin Hospital of Wuhan University, Wuhan, Hubei, People's Republic of China

² Department of Otolaryngology-Head and Neck Surgery, Tongji Hospital, Tongji Medical College, Huazhong University of Sciences and Technology, Wuhan, Hubei, People's Republic of China

In this work, we aim to investigate the effect of glycolysis on IBD and CRC to reveal the cross-talk genes between IBD and CRC. Transcriptomic data from IBD and CRC patients retrieved from Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) were used to identify the glycolytic cross-talk genes. A network of co-regulated transcription factors of the cross-talk genes was constructed. We explored the relationship between the cross-talk genes and prognosis, tumor microenvironment (TME), mutations, and drug sensitivity in CRC. Moreover, a nomogram was constructed to predict the prognosis of CRC, and a combined diagnostic model of IBD was identified. Finally, the expression of cross-talk genes was validated by immunohistochemistry (IHC).

Results

Identification of differentially expressed glycolysis-related genes

The dataset containing 638 CRC and 49 normal samples with clinical information was obtained from TCGA-Colon Adenocarcinoma (COAD) Rectum Adenocarcinoma (READ). The IBD datasets containing 70 inflamed IBD tissues and 31 healthy controls were obtained from GSE17928. We downloaded glycolysis-related gene sets from MsigDB and GSEA was performed on the CRC and IBD datasets, respectively. As shown in Fig. 1A–D, HALLMARK_GLYCOLYSIS and REACTOME_GLYCOLYSIS were significantly enriched in CRC and IBD samples. The DEGs of IBD and CRC were identified via the R package limma. The overlap between IBD-DEGs, CRC-DEGs, and GRGs was defined as differential expression glycolysis-related genes (DE-GRGs) for further analysis (Fig. 1E).

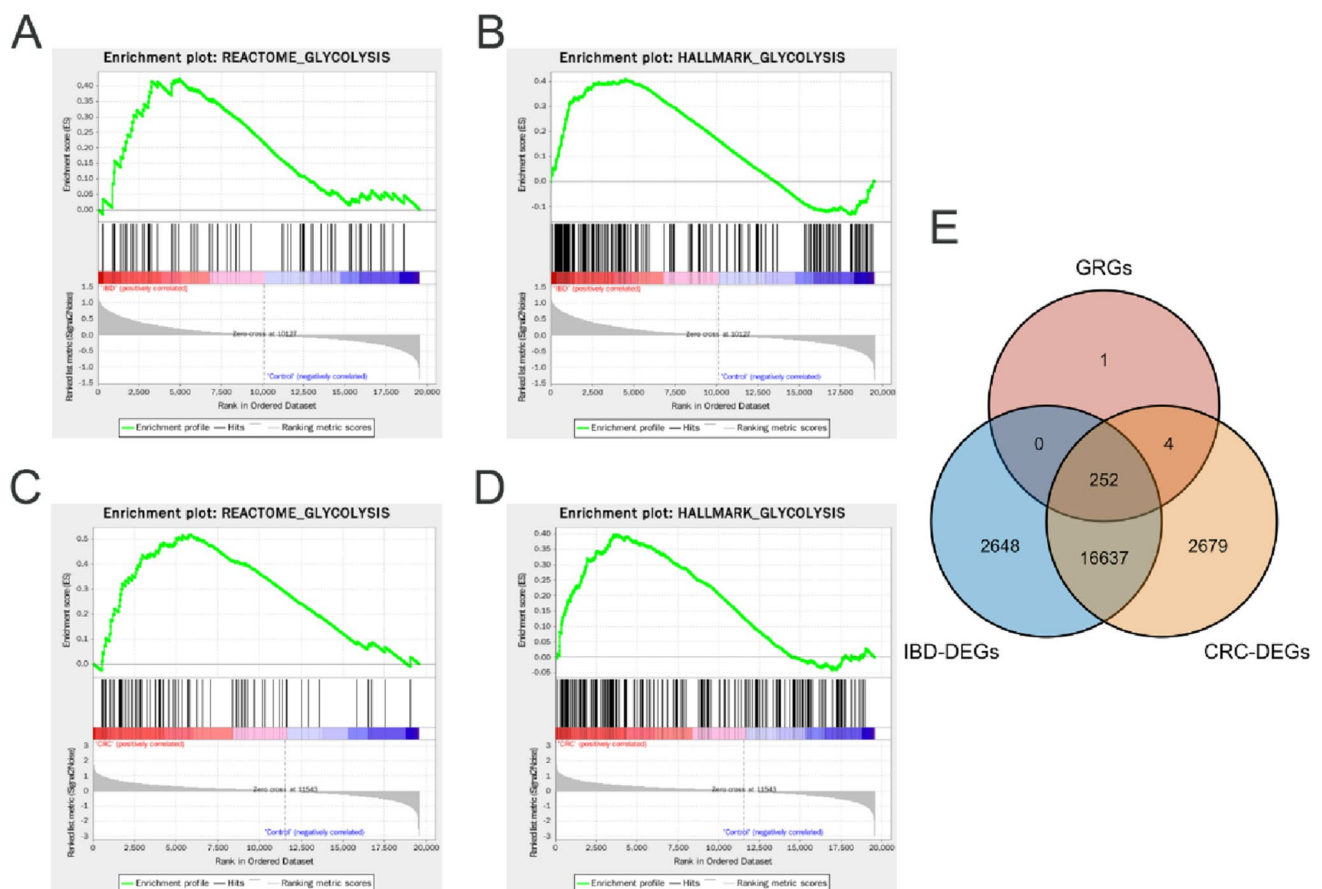


Fig. 1 Glycolysis-related genes in IBD and CRC. **A–B** Enrichment plots of glycolysis in IBD using GSEA. **C–D** Enrichment plots of glycolysis in CRC using GSEA. **E** Intersection of GRGs, IBD-DEGs, and CRC-DEGs

Weighted gene co-expression network analysis

To explore the relative gene modules of IBD, WCGNA was applied in GSE17928 with a soft threshold at 6 (Figure S1A). Sixteen gene modules were identified, the lightgreen module in Fig. 2A–B, and Figure S1B show strong positive correlation with CRC, with module trait correlation = 0.71. The intersection of the lightgreen module and the DE-GRGs obtained included 48 genes (Fig. 2C).

Screening for prognosis-related genes

The obtained 48 genes were used to explore prognostic markers in CRC via LASSO regression model and seven genes (P4HA1, PMM2, ENO1, LDHC, SPAG8, CHPF2, and AGRN) were obtained (Fig. 3A–B). SVM-RFE, a machine

learning algorithm for feature selection, was applied to select prognosis-related genes ranking by parameter “Importance” (Fig. 3C). The intersection genes of LASSO and SVM-RFE were used for next analysis (Fig. 3D). We performed univariate Cox regression analysis. As shown in Fig. 3E, only P4HA1, PMM2, and AGRN exhibited significant correlation with overall survival (OS), and the transcriptomic expression of P4HA1, PMM2, and AGRN was significantly upregulated in tumor tissues (Fig. 3F). Furthermore, we retrieved proteomic data and IHC staining from the CPTAC and HPA databases to investigate proteomic expression levels of P4HA1, PMM2, and AGRN. As shown in Fig. 3G, P4HA1 and PMM2 were highly expressed, whereas the expression level of AGRN was low in CRC tissues. The proteomic expression levels of P4HA1, PMM2, and AGRN were further confirmed via IHC staining from the HPA database

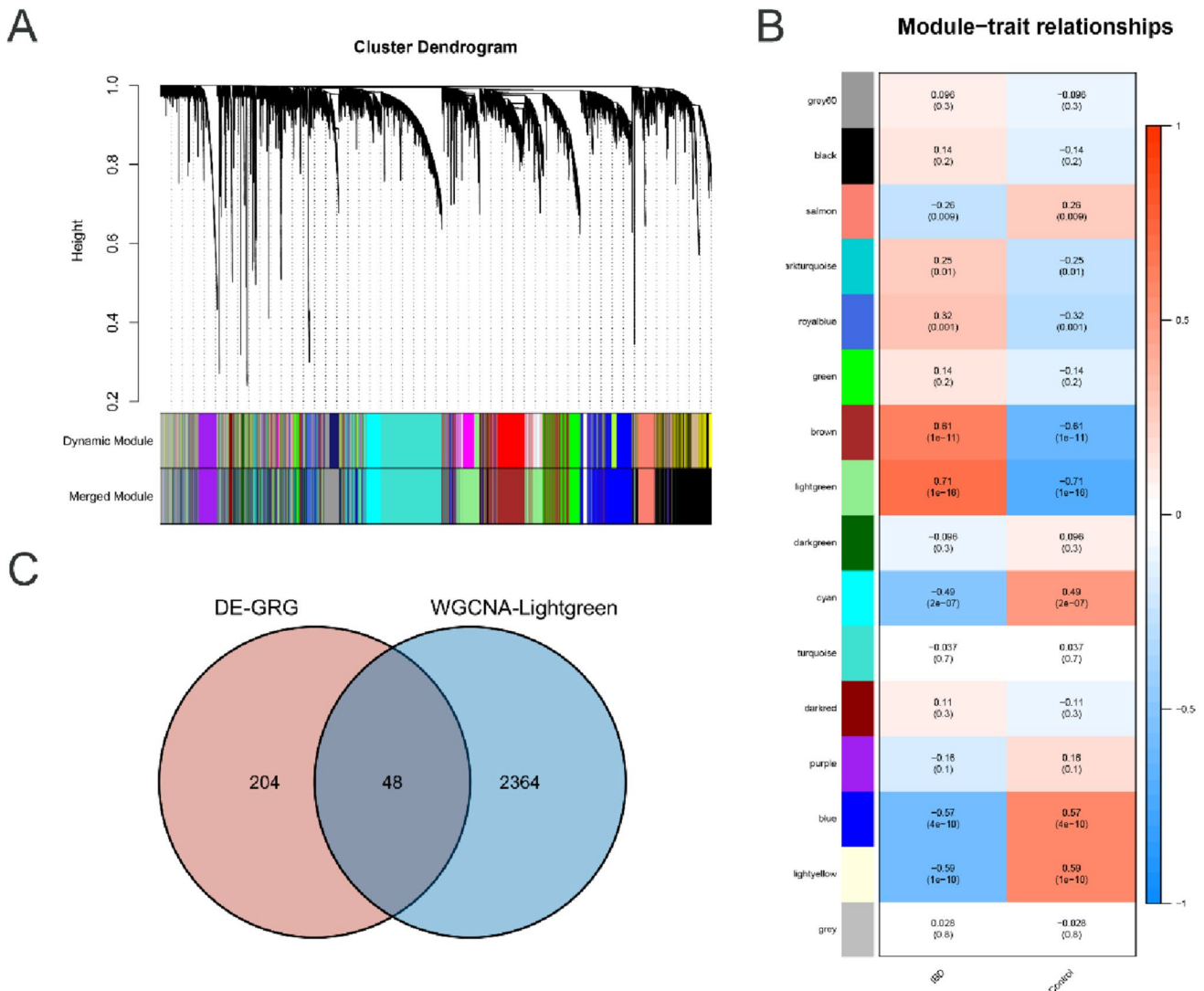


Fig. 2 WGCNA. **A** Resulting gene dendrograms. **B** Module trait relationships in gene subtype A and B, which contained the corresponding correlation and *p* value. **C** Intersection of DE-GRG and lightgreen module

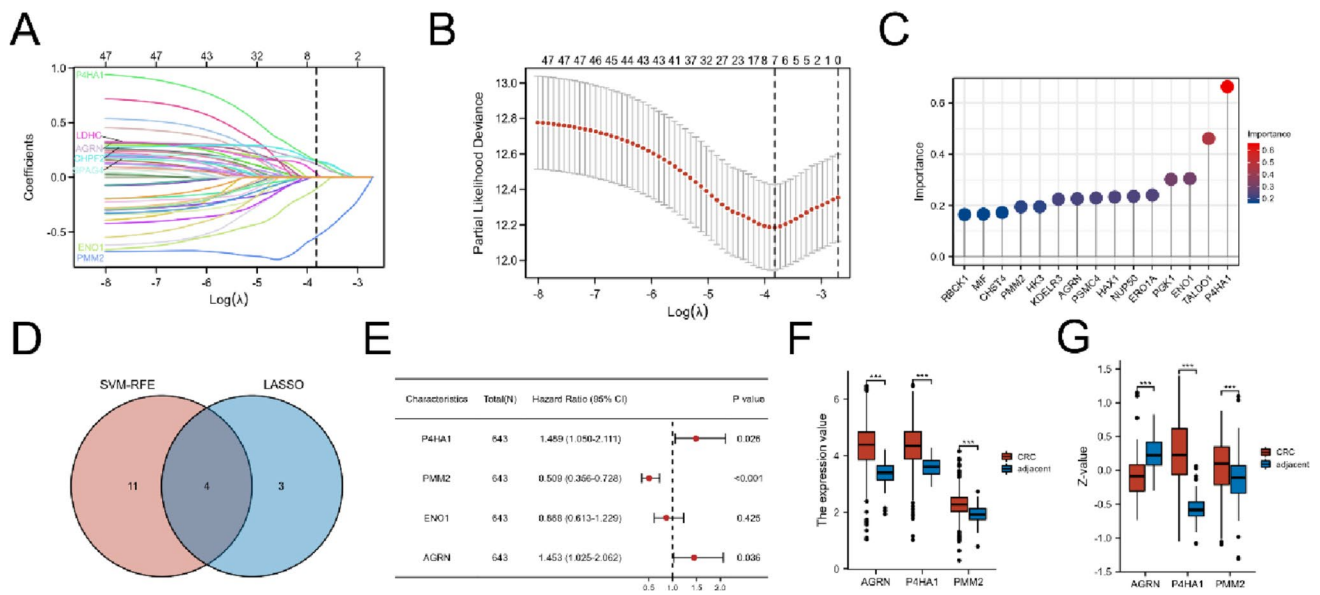


Fig. 3 Screening prognostic genes in CRC. **A** Ten-fold cross-validation for the coefficients of hub genes in the LASSO model. **B** The selection of optimal parameter (lambda) in LASSO model. **C** Lollipop chart of importance in SVM-RFE. **D** Intersection of SVM-

RFE and LASSO. **E** Forest plot of the multivariable Cox model. **F** Transcriptomic expression levels of AGRN, P4HA1, and PMM2. **G** Proteomic expression levels of AGRN, P4HA1, and PMM2. ****p* < 0.001

(Fig. 4A–F). Since proteomic expression levels of P4HA1 and PMM2 were consistent with transcriptomic expression levels, these two genes were identified as glycolytic cross-talk genes between IBD and CRC.

Prediction of co-regulated transcription factors

To explore the cross-talk of regulatory networks between P4HA1 and PMM2 in both IBD and CRC, the hTFtarget

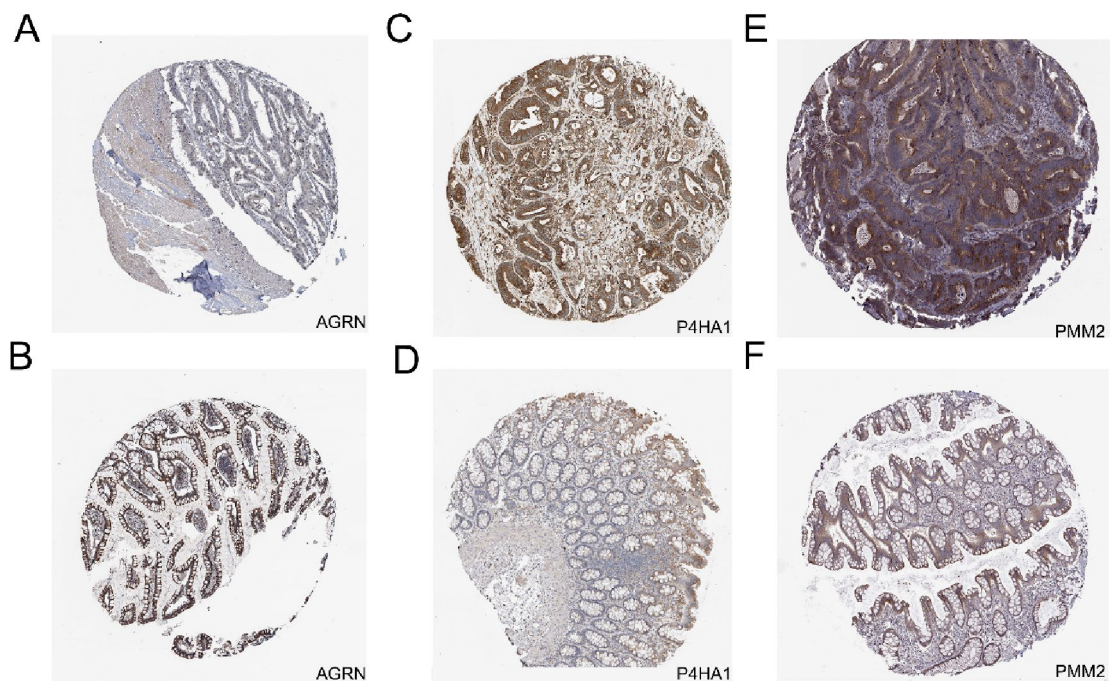


Fig. 4 IHC validation in HPA database. **A–B** IHC of AGRN in normal and CRC, respectively. **C–D** IHC of P4HA1 in normal and CRC, respectively. **E–F** IHC of PMM2 in normal and CRC, respectively

database was used to predict co-regulated TFs of P4HA1 and PMM2. One hundred twelve co-regulated TFs were found, and only four TFs (GRHL3, CEBPB, TCF3, and SUPT5H) were significantly upregulated in both IBD and CRC (Fig. 5A–B). Therefore, GRHL3, CEBPB, TCF3, and SUPT5H were used to construct the co-regulated network of P4HA1 and PMM2, which is presented in Fig. 5C.

Construction of prognostic risk signature

To further explore the role of the two cross-talk genes, a prognostic signature was established as follows: Risk score = (0.3419)*P4HA1 + (-0.7116)*PMM2. Risk scores of each patient in the TCGA-COADREAD and GSE17536 datasets were calculated, and the patients were divided into a high risk group and a low risk group based on the median risk score. As shown in Fig. 6A, high risk patients had relatively low OS than patients with low risk in TCGA-COADREAD ($p = 0.00015$). The risk scores, survival status, and heatmap of P4HA1 and PMM2 are presented in Fig. 6B. The AUC values of TCGA-COADREAD at 1, 3, and 5 years

were 0.649, 0.646, and 0.650, respectively (Fig. 6C). The prognostic signature of the two genes was validated in the GSE17536 dataset. Similarly, in the TCGA-COADREAD dataset, patients with low risk had relatively higher OS than patients in the high risk group (Fig. 6D). The risk scores, survival status, and heatmap of P4HA1 and PMM2 are presented in Fig. 6E. The AUC values of GSE17536 at 1, 3, and 5 years were 0.654, 0.655, and 0.670, respectively (Fig. 6F).

Clinical correlation analysis

We conducted correlation analysis between clinical characteristics and risk scores to explore the impact of risk scores on clinical characteristics. As shown in Fig. 7A, risk scores of patients in T stage 3 and 4 were significantly higher than those in T stage 1 and 2. In addition, we observed significantly higher risk scores for patients in stage II/III/IV relative to patients in stage I (Fig. 7B). Moreover, risk scores in right-side CRC were higher than that in left-side CRC (Fig. 7C). However, risk scores were not correlated with age and gender (Figure S2A–B). Univariate and multivariate

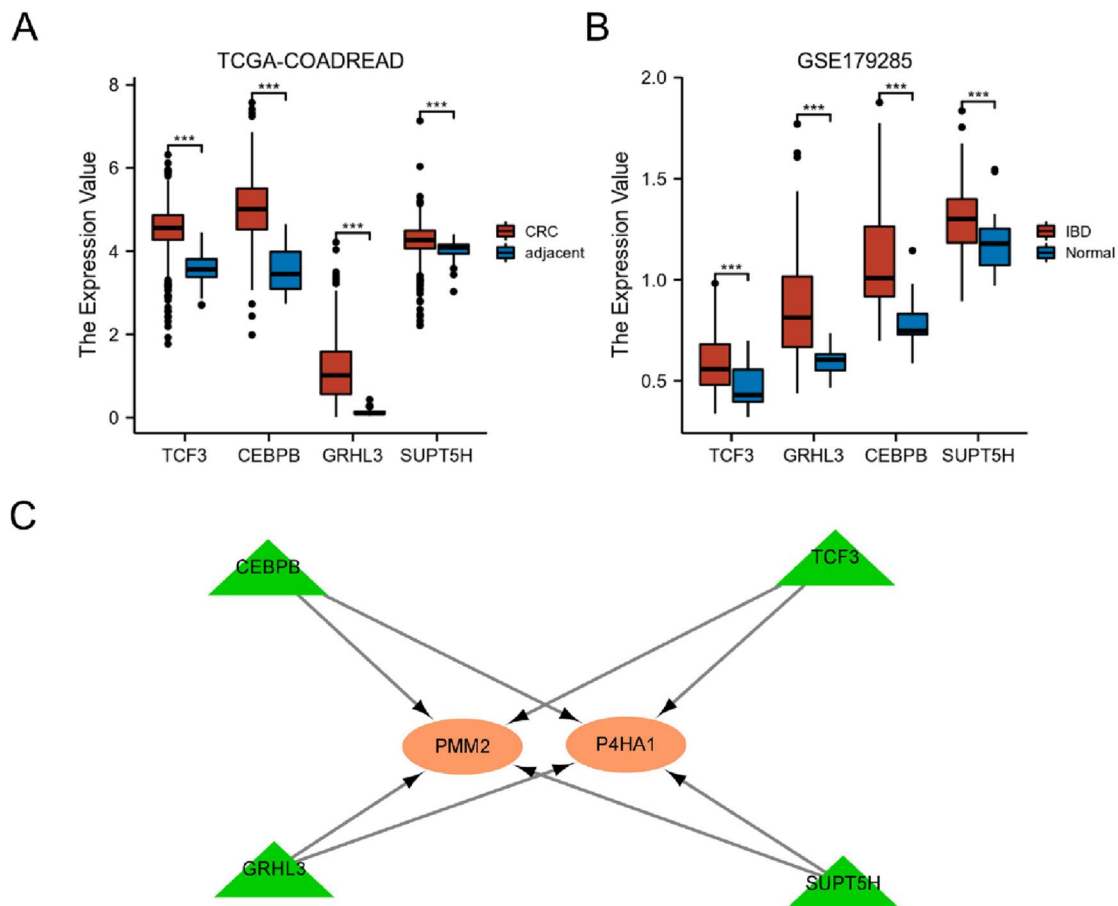


Fig. 5 Co-related TF networks in CRC and IBD. **A** The expression of TCF3, CEBPB, GRHL3, and SUPT5H in CRC. **B** The expression of TCF3, CEBPB, GRHL3, and SUPT5H in IBD. **C** Co-related TF networks of P4HA1 and PMM2. *** $p < 0.001$

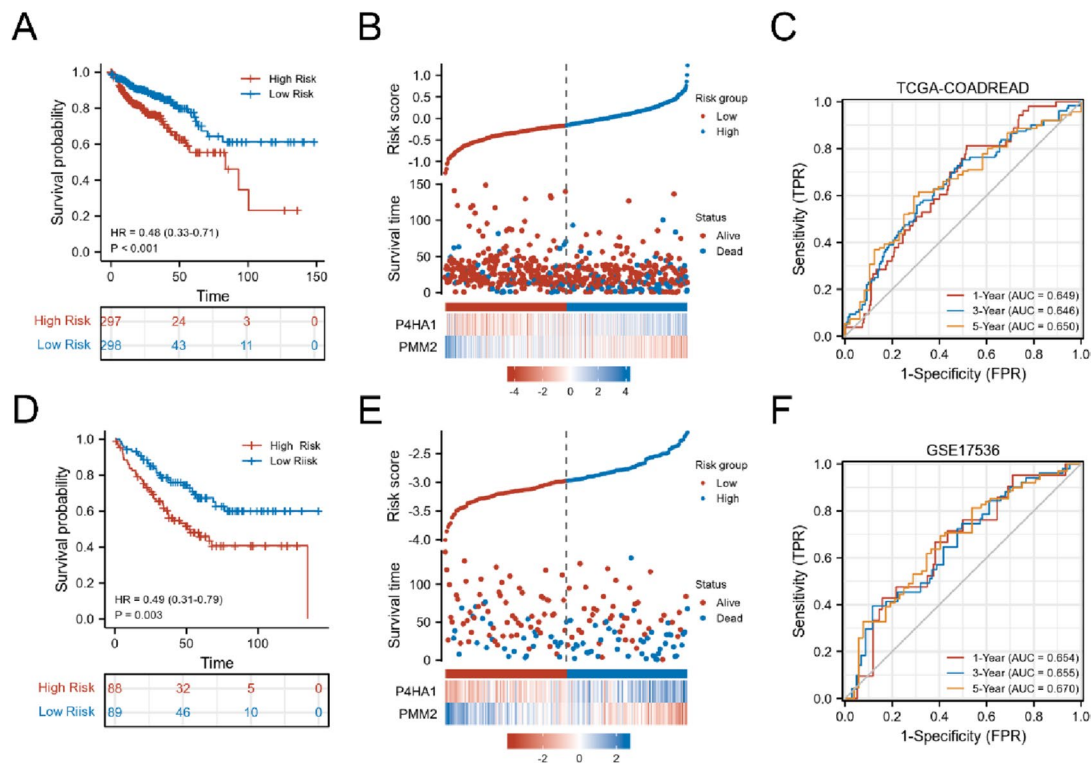


Fig. 6 KM survival analysis, risk score assessment, and time-dependent ROC curves in the TCGA and GEO datasets. **A–C** TCGA-COADREAD. **D–F** GSE17536

Cox regression analyses were performed to estimate prognostic potential of the signatures of the two genes with several clinic pathological characteristics, such as age, pathological stage, T stage, M stage, N stage, and gender in CRC. As shown in Fig. 7D, the risk scores significantly correlated with OS (HR = 3.045, 95% CI = 1.836–5.049, $p < 0.001$) in univariate Cox regression. After correcting for other confounding factors, the risk score (HR = 2.716, 95% CI = 1.549–4.764, $p < 0.001$), age ≥ 65 (HR = 2.752, 95% CI = 1.769–4.281, $p = 0.008$), and M stage (HR = 2.320, 95% CI = 1.438–3.741, $p < 0.001$) still proved to be independent predictors for OS in the multivariate Cox regression analysis (Fig. 7E). These results demonstrate that risk scores could be a potential prognostic indicator for predicting OS in CRC.

Characteristics of tumor microenvironment and immune checkpoints

ssGSEA was used to estimate immune cell infiltration in CRC patients from TCGA-COADREAD and GSE17536 datasets. As shown in Fig. 8A and Figure S3A, higher infiltration levels were observed in the high risk group relative to those in the low risk group in both TCGA-COADREAD and GSE17536. To explore tumor microenvironment (TME) of CRC, ESTIMATE was applied to calculate the immune score, stromal score, and ESTIMATE score in

TCGA-COADREAD and GSE17536. The immune score, stromal score, and ESTIMATE scores were significantly higher in the high risk group than in the low risk group (Fig. 8B–D and Figure S3B–D). Considering the vital role of immune checkpoint molecules (PD-1, PD-L1, and CTLA-4) in the tumor immune microenvironment, we analyzed the expression of PD-1, PD-L1, and CTLA-4 in the two groups. As shown in Fig. 8E–G, the expression of all immune checkpoint molecules was significantly elevated in the high risk group.

Correlation of gene signature with microsatellite instability and stemness

A recent study reported that patients with high microsatellite instability (MSI-H) were more sensitive to immunotherapy (Ganesh et al. 2019). The MSI and microsatellite stable (MSS) status of CRC patients were estimated in the TCGA-COADREAD dataset. As shown in Fig. 9A–B, high risk patients were significantly associated with MSI-H, whereas low risk patients were associated with MSS. Furthermore, we evaluated cancer stemness by mRNasi. Correlation analysis revealed that risk scores were negatively associated with mRNasi, indicating that CRC cells with a lower risk score had more distinct stem cell properties and a lower degree of cell differentiation (Fig. 9C–D).

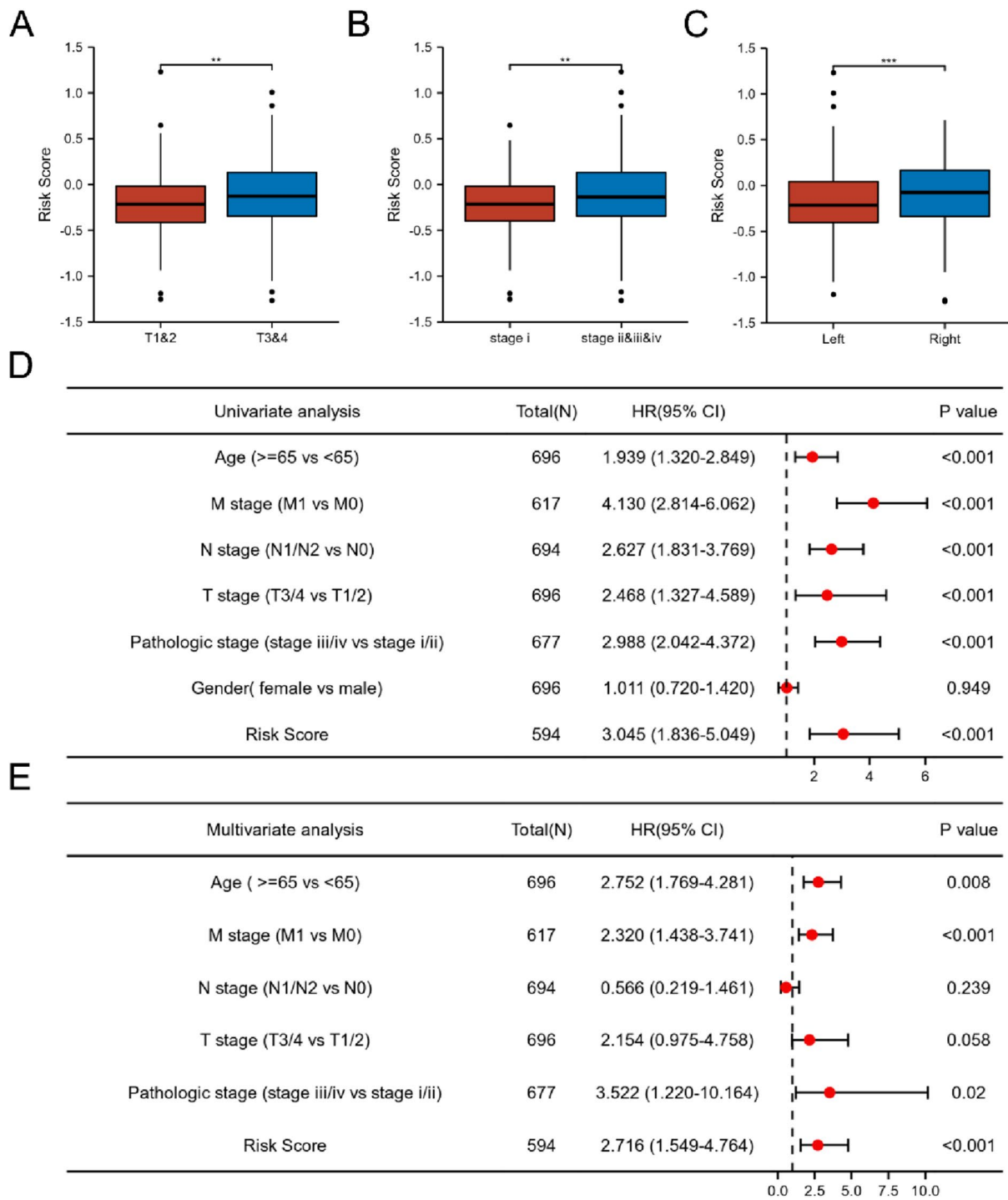


Fig. 7 The relationship between risk score and clinical feature. **A** Boxplot of risk score in T stages. **B** Boxplot of risk score in pathological stage. **C** Boxplot of risk score in tumor location. **D** Univariate analysis of risk score. **E** Multivariate analysis of risk score. ** $p < 0.01$, *** $p < 0.001$

Mutation and drug sensitivity analysis

Increasing evidence suggests that patients with high tumor mutational burden (TMB) may benefit from immunotherapy because they have higher numbers of neoantigens

(Snyder et al. 2014). Therefore, TMB scores were calculated in CRC. The CRC patients with high risk scores had higher TMB scores (Fig. 9E–F). We evaluated the variations in the distribution of the somatic mutations in the TCGA-COADREAD dataset. As shown in Fig. 9G,

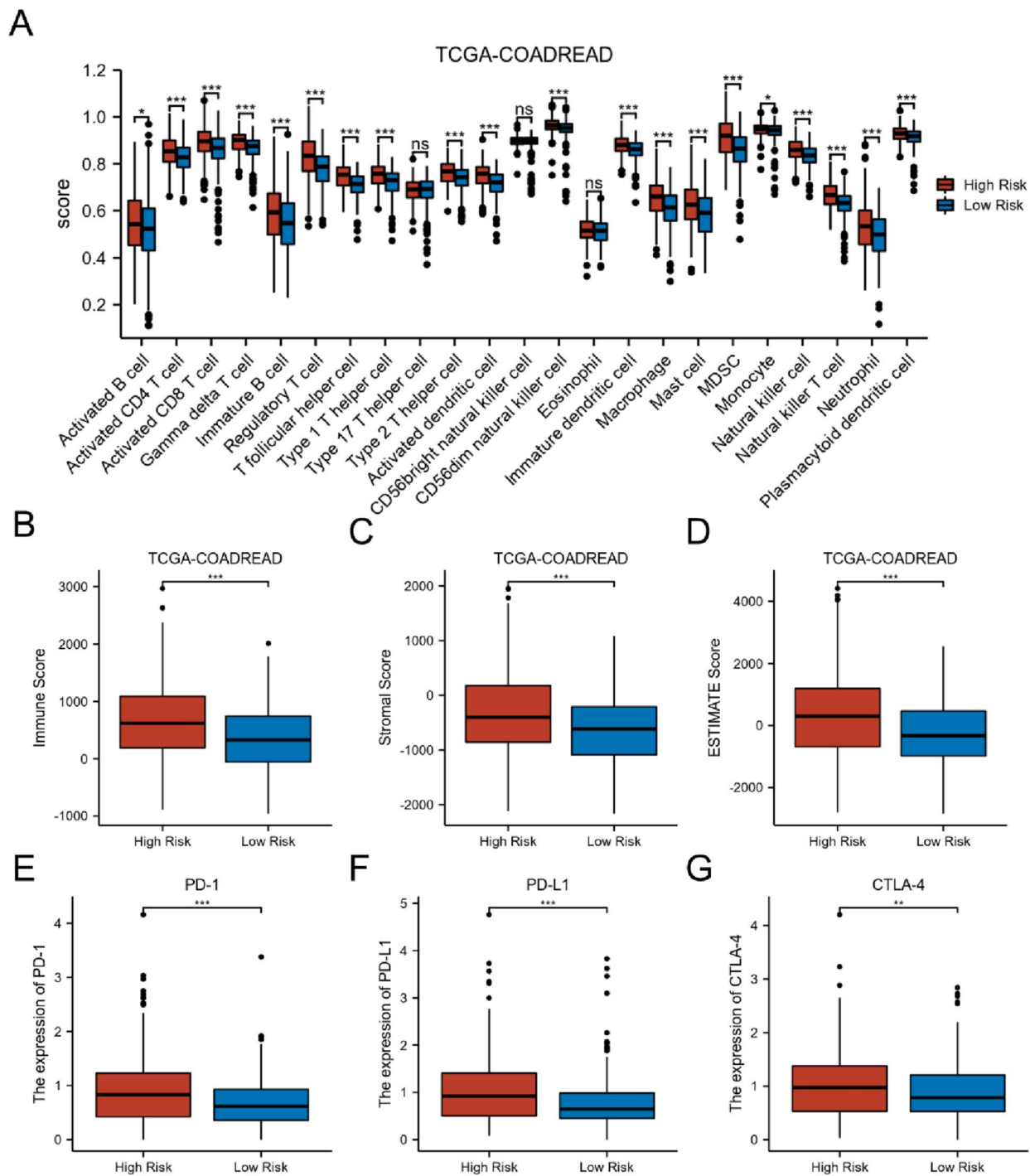


Fig. 8 Immune characteristics between high and low risk group in TCGA dataset. **A** Infiltration of 23 immune cells using ssGSEA. **B–D** Immune, stromal, and ESTIMATE scores. **E–G** The expression of

PD-1, PD-L1, and CTLA-4. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns not significant

the ten most frequently mutated genes in the high risk group were APC, TTN, TP53, KRAS, MUC16, SYNE1, PIK3CA, FAT4, RYR2, and DHAN5. However, patients with a low risk had markedly higher frequencies of APC, TP53, and KRAS mutations compared to patients with a

high risk (Fig. 9H). Furthermore, we investigated sensitivities of CRC patients to the commonly used chemotherapy drugs. We found that patients with a high risk score had a lower IC50 value for camptothecin, whereas IC50 values

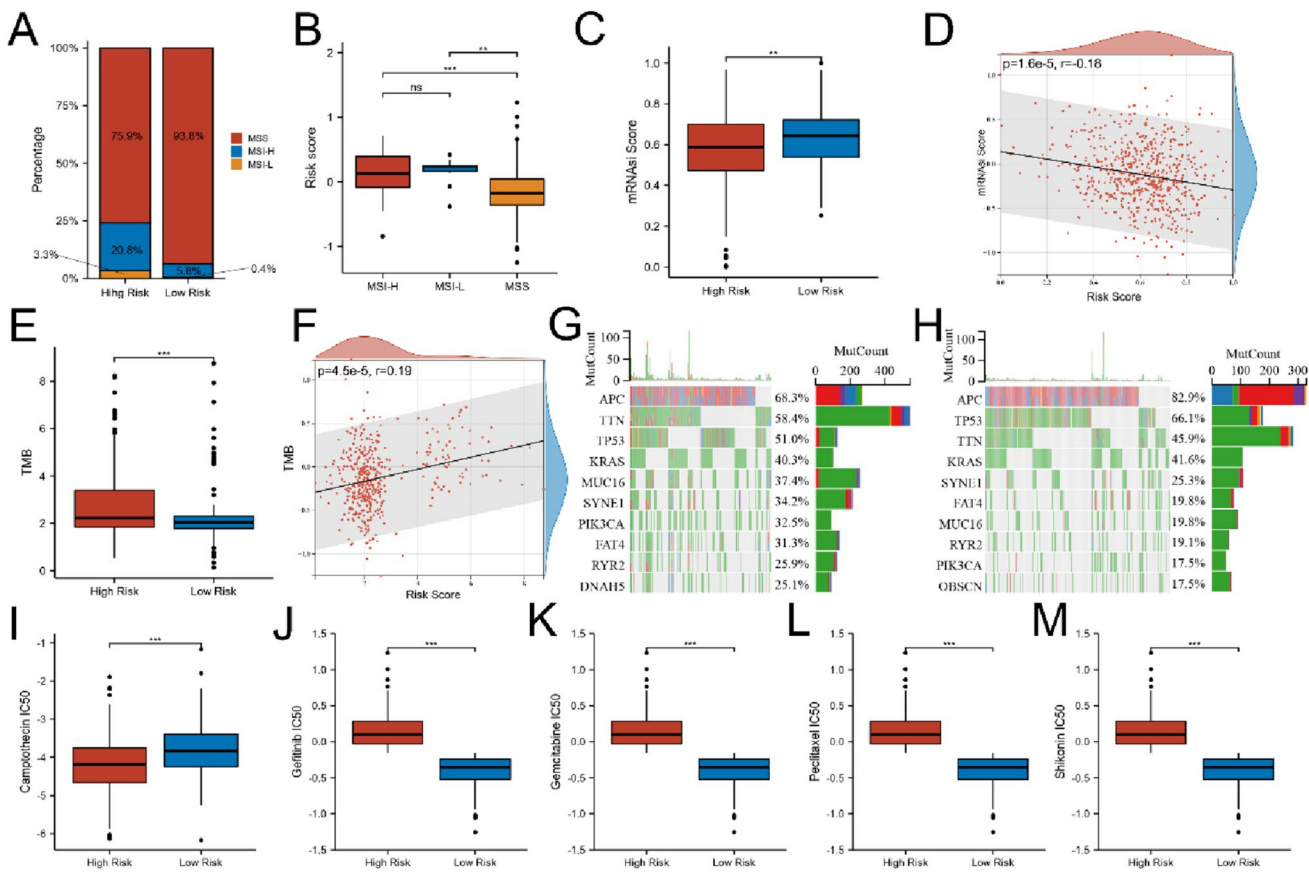


Fig. 9 Mutant and drug susceptibility analysis. **A–B** Relationship between MSI and risk score. **C–D** Relationship between mRNAseI score and risk score. **E–F** Relationship between TMB and risk score. **G–H** The waterfall plot of somatic mutation characteristics of CRC

patients with high and low risk. **I–M** Relationships between chemotherapeutic sensitivity and risk score. ****** $p < 0.01$, ******* $p < 0.001$, **ns** not significant

of gemcitabine, paclitaxel, gefitinib, and shikonin were significantly higher in high risk group (Fig. 9I–M).

Establishment of a nomogram

To establish a quantitative prediction model, a nomogram was constructed (Fig. 10A), including risk score, age, M stage, and N stage. Calibration plots indicated that the nomogram and ideal model showed high consistency in TCGA-COADREAD (Fig. 10B).

Diagnostic values of P4HA1 and PMM2 in IBD

We combined P4HA1 and PMM2 via a logistics model for ROC analysis. The results were presented in Fig. 10C–E. The AUC values of the combined model were 0.865, 0.743, and 0.822 in GSE179285, GSE126124, and GSE75214, respectively. This indicates that a combined model of P4HA1 and PMM2 performs well for diagnosing IBD.

Validation of P4H1A and PMM2 expression via IHC

To validate the proteomic expression levels of P4HA1 and PMM2 in patients with IBD and CRC, we collected twenty IBD, one hundred CRC, and fifty normal formalin-fixed, paraffin-embedded tissue samples from the Renmin Hospital of Wuhan University and performed IHC. As shown in Fig. 11A–H, the proteomic expression levels of P4HA1 and PMM2 were significantly upregulated in IBD and CRC relative to normal tissues. Our results suggest that the glycolysis-related genes P4HA1 and PMM2 are involved in the pathogenesis of IBD and CRC.

Discussion

Accumulating evidence suggests that patients with IBD have an increased risk of CRC (Beni et al. 2022, Rajamaki et al. 2021, Wijnands et al. 2021). It may be due to the development of dysplastic lesions in the colonic mucosa caused by chronic inflammation (Frigerio et al. 2021). Though the

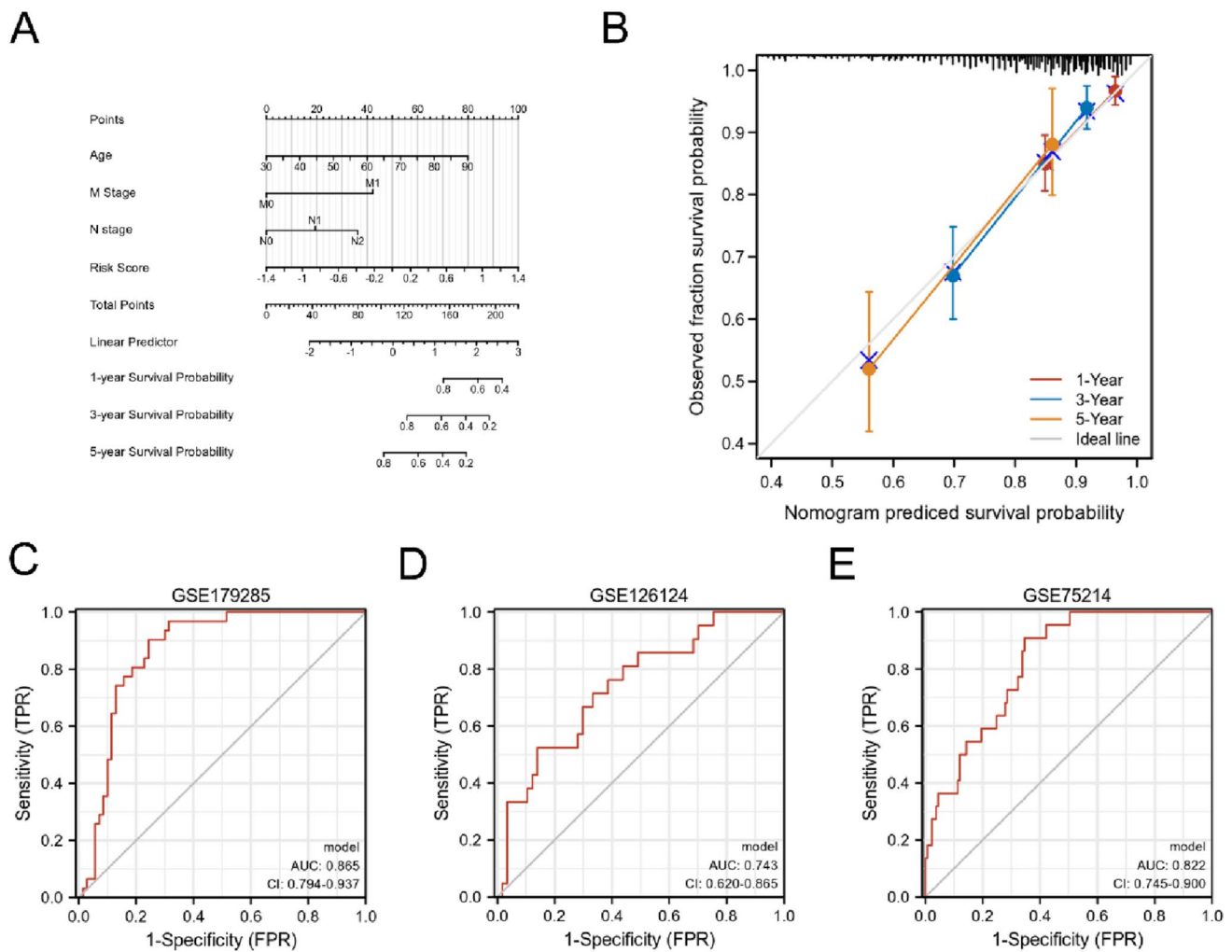


Fig. 10 Predict and diagnostic values of signature in CRC and IBD. **A** A nomogram for predicting 1-, 3-, and 5-year OS of CRC patients. **B** Calibration curves of the nomogram for predicting 1-, 3-, and

5-year OS of CRC patients. **C–E** ROC of logistics model of P4HA1 and PMM2 in GSE179285, GSE12624, and GSE75214

incidence of IBD-associated CRC is low relative to sporadic CRC, patients with IBD have a mortality rate of 10 to 15% due to CRC (Eluri et al. 2017). Glycolysis is one of the major metabolic pathways essential for providing energy for cellular processes, and it is involved in the regulation of innate and adaptive immune cells in the inflammatory response (Soto-Herederó et al. 2020). Studies have shown that macrophages regulate glycolysis via the B-cell adapter for PI3K to improve experimental colitis (Irizarry-Caro et al. 2020). Moreover, N6-methyladenosine modifications of HK2 and GLUT1 regulate glycolysis to enhance colorectal cancer progression (Shen et al. 2020).

IBD and CRC may have overlapping pathogenic pathways involving glycolysis. The main purpose of this study was to reveal glycolytic cross-talk genes in IBD and CRC. We firstly identified 252 cross-talk DE-GRGs between IBD and CRC. Further, WGCNA was used to screen for IBD-related

modules, and the module with high trait correlation was intersected with DE-GRGs for the next step in the analysis. Finally, P4HA1 and PMM2 were identified as potential prognostic markers in CRC using LASSO, SVM-RFE, and Cox regressions. Therefore, we speculated that P4HA1 and PMM2 might be two glycolytic cross-talk genes. By further verification, we found that four TFs that participate in a coordinated manner in the regulation of P4HA1 and PMM2 were highly expressed in IBD and CRC, including GRHL3, CEBPB, TCF3, and SUPT5H.

One of two glycolytic cross-talk genes identified in this study, P4HA1, is a key enzyme in collagen synthesis. Upregulation of P4HA1 contributes to glycolysis by improving the stability of HIF1- α in breast cancer and pancreatic cancer (Cao et al. 2019, Xiong et al. 2018). Studies have confirmed P4HA1 is overexpressed in CRC. Agarwal et al. reported that knockdown of P4HA1 decreased CRC cell proliferation,

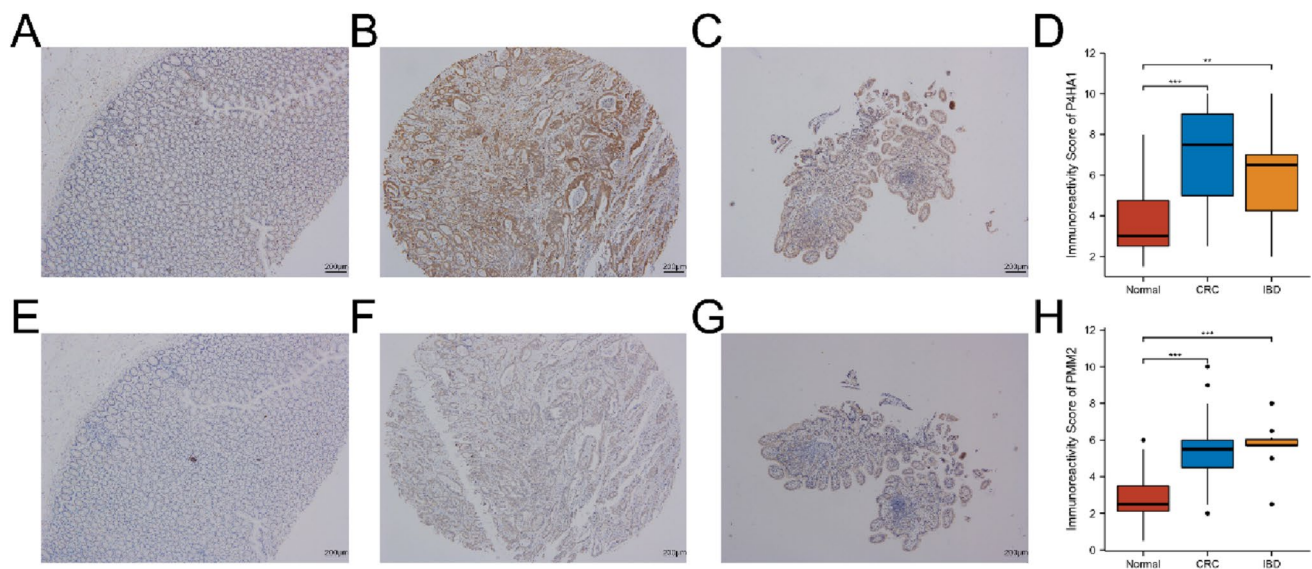


Fig. 11 IHC validation (magnification $\times 40$). **A–C** IHC of P4HA in normal, CRC, and IBD, respectively. **D** Immunoreactivity score of P4HA1 in normal, CRC, and IBD, respectively. **E–G** IHC of PMM2

in normal, CRC, and IBD, respectively. **H** Immunoreactivity score of PMM2 in normal, CRC, and IBD, respectively. $^{***}p < 0.01$, $^{**}p < 0.001$

invasion, and migration and inhibited tumor growth as well as metastases. P4HA1 inhibitors have been shown to be an effective therapeutic strategy for aggressive CRC (Agarwal et al. 2020). The other glycolytic cross-talk gene identified was PMM2. PMM2 catalyzes the isomerization of mannose 6-phosphate to mannose 1-phosphate, which is a precursor of GDP-mannose necessary for the synthesis of dolichol-P-oligosaccharides. Mutations of PMM2 result in the PMM2-congenital disorder of glycosylation, which is an inherited condition that affects many parts of the body (Witters et al. 2019). Studies have indicated that the overexpression of PMM2 could promote malignancy of renal cell carcinoma (Yamada et al. 2018).

To further explore the role of P4HA1 and PMM2, we constructed a risk signature based on P4HA1 and PMM2 in CRC. Our results suggest that CRC patients with a high risk score have a poor prognosis. Moreover, patients with low and high risk scores showed significantly different clinical and pathological characteristics, mutations, TMEs, immune checkpoints, MSI, cancer stemness, and drug susceptibility. Finally, a nomogram was constructed by integrating the risk score, tumor stage, and age which further facilitated the use of risk signature. These findings may improve the prognosis stratification of patients with CRC and provide new ideas for targeted therapies. Despite recent advances in immunotherapy for CRC, different patients still show different sensitivity to treatment, highlighting the crucial role of TME in CRC tumorigenesis and progression (Chen et al. 2021). PD-1/PD-L1 and CTLA-4 blockades have demonstrated their safety and efficacy and have been used to treat CRC (Monjazeb et al. 2021, Yaghoubi et al. 2019).

Our results showed that patients with a high risk score had higher infiltration of immune cells and high expression levels of immune checkpoints. Moreover, TMB score and the proportion of patients with MSI-H were higher in the high risk group. Previous studies have shown that patients with a high TMB or MSI-H may be more sensitive to immunotherapy (Ganesh, Stadler, Cercek, Mendelsohn, Shia, Segal and Diaz 2019, Snyder, Makarov, Merghoub, Yuan, Zaretsky, Desrichard, Walsh, Postow, Wong, Ho, Hollmann, Bruggeman, Kannan, Li, Elipenahli, Liu, Harbison, Wang, Ribas, Wolchok and Chan 2014). This suggests that CRC patients with a high risk score are more likely to benefit from immunotherapy. Furthermore, we evaluated the diagnostic values of P4HA1 and PMM2 in IBD. Our results reveal that a combined model based on P4HA and PMM2 had an excellent diagnostic accuracy for IBD.

Nevertheless, this study had several limitations. Firstly, although we applied IHC to validate the expression levels of P4HA1 and PMM2 in patients with IBD and CRC, it lacks some clinical information, such as disease activity, previous therapies, and prognosis, which may have affected expression levels. Secondly, we failed to find an expression profile data of IBD-associated CRC to further validate our findings.

Conclusion

Our comprehensive analyses revealed glycolysis may be involved in IBD developing CRC. Furthermore, glycolysis-related genes P4HA1 and PMM2 are potential cross-talk genes and biomarkers in IBD and CRC. This study provides

new insights for the further study of the molecular mechanism of IBD-associated CRC.

Method

Sample collection and immunohistochemistry analysis

Formalin-fixed, paraffin-embedded tissue samples were collected from twenty IBD, one hundred CRC, and fifty healthy patients for IHC analysis from the Renmin Hospital of Wuhan University from 2020 to 2021. This study was approved by the Ethics Committee of Renmin Hospital of Wuhan University (WDRY2021-KS066); all methods were performed in accordance with the relevant guidelines and the Declaration of Helsinki. Immunohistochemical staining was performed using a standard EnVision complex method (Kammerer et al. 2001). The primary antibodies used were P4HA1 (Proteintech, Wuhan, China) and PMM2 (Proteintech, Wuhan, China). Five high-power fields (400×) were randomly selected, and examination and scoring were performed independently by three experienced pathologists who were unaware of the clinical information. The expression of P4HA1 and PMM2 was calculated by multiplying the mean signal intensity (on a scale of 0–3: 0, no staining; 1, light staining; 2, moderate staining; and 3, strong staining) and the proportion of positive tumor cells (on a scale of 0–4: 0, 0%; 1, 1–25%; 2, 26–50%; 3, 51–75%; and 4, 76–100%). The final immunoreactive scores were the mean of scores from the three pathologists.

Data collection and gene set enrichment analysis

The RNA sequencing data and clinical data of 644 CRC patients were downloaded from TCGA. The expression array profiles of CRC patients were obtained from GEO, namely, GSE161158 ($n = 192$) and GSE17536 ($n = 178$). The IBD datasets GSE179285 (70 inflamed IBD tissues and 31 healthy controls), GSE75214 (133 inflamed IBD tissues and 22 healthy controls), and GSE126124 (57 inflamed IBD tissues and 21 healthy controls) were used. Gene set enrichment analysis (GSEA) was performed via GSEA/MSigDB and visualized by the R package ggplot2 (Subramanian et al. 2005, Wickham 2009). Glycolysis-related pathways from GSEA were used to obtain glycolysis-related genes (GRGs).

Differential expression and weighted correlation network analysis

Differential expression analyses were performed using the R package limma with a false discovery rate of less than 0.05 (Ritchie et al. 2015), then genes were defined

as IBD-differential expression genes (DEGs) and CRC-DEGs. Weighted correlation network analysis (WGCNA) was applied to assess the relative importance and module membership of genes (Langfelder and Horvath 2008). The minimum number of module genes was set at 30. The hierarchical clustering dendrogram summarizes the gene modules with different colors.

Construction and evaluation of the prognostic model

The least absolute shrinkage and selection operator (LASSO) regression and support vector machine recursive feature elimination (SVM-RFE) were applied to screen CRC prognosis and to construct a prognosis-predicting model using the glmnet R package (Friedman et al. 2010). The risk score was calculated using the following formula: Risk score = $\sum_{i=1}^n coef * id$ (Chen et al. 2007). The survminer and survival R packages were used to conduct the Kaplan–Meier survival analysis to assess the difference in survival between high and low risk score groups. We estimated the performance of the gene risk model using the time-dependent receiver operating characteristic (ROC) curve estimated by the R package survivalROC (Lorent et al. 2014). Then, Cox regression analyses were conducted to estimate the independent prognostic values of signature and other clinical characteristics. The proteomic expression of prognostic cross-talk genes was confirmed by the Clinical Proteomic Tumor Analysis Consortium (CPTAC) and the Human Protein Atlas (HPA) databases (Uhlen et al. 2015). To estimate survival likelihood, a nomogram was constructed based on the risk score and clinical characteristics using the R package rms, and the prognostic ability of the nomogram was assessed using calibration plots.

Prediction of co-regulated transcription factors

The hTFtarget database was utilized to predict co-regulated transcription factors (TFs) (Zhang et al. 2020). hTFtarget is an online tool that has a curated and comprehensive database of TF-target regulators from large-scale ChIP-Seq data from human TFs (7190 experimental samples of 659 TFs), in 569 conditions (399 types of cell line, 129 classes of tissues or cells, and 141 kinds of treatments). The co-regulation network was constructed and visualized via Cytoscape.

Evaluation of immune cell infiltration and tumor microenvironment

Single-sample GSEA (ssGSEA) was applied to estimate the 23 common immune cell infiltrations (Charoentong et al. 2017). The R package estimate was used to predict the presence of stromal or immune cells in tumor tissues

(Yoshihara et al. 2013). The one-class logistic regression algorithm constructed by Malta et al. was used to evaluate cancer stemness based on the mRNA expression-based stemness index (mRNAsi) (Malta et al. 2018).

Mutations and drug susceptibility analysis

Files in the mutation annotation format from the TCGA database were generated by the R package maftools for mutation analysis and the calculation of tumor mutation burden (TMB) score (Mayakonda et al. 2018). The semi-inhibitory concentration (IC50) values of chemotherapeutic drugs commonly used to treat CRC were obtained using the R package pRRophetic (Geeleher et al. 2014).

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10142-023-01170-5>.

Author contribution Conceptualization: Chenglin Ye, Sizhe Zhu, and Jingping Yuan; methodology: Chenglin Ye and Yabing Huang; visualization: Chenglin Ye and Yuan Gao; writing—original draft preparation: Chenglin Ye; writing—review and editing: Sizhe Zhu and Jingping Yuan.

Funding This work was supported by the National Natural Science Foundation of China (No. 82202524 and 62276191) and Open Foundation of Hubei Key Laboratory (2021KFY004 and 2021KFY002).

Data availability The data used to support the findings of this study are available from supplementary materials and the corresponding author upon request.

Declarations

Ethics approval and consent to participate The studies involving human participants were reviewed and approved by Ethics Committee of Renmin Hospital of Wuhan University (WDRY2021-KS066). All methods were performed in accordance with the relevant guidelines and the Declaration of Helsinki. The age of all participants is greater than 16 years old, and written informed consent to participate in this study was provided by the participants.

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

References

- Abboud G, Choi SC, Kanda N, Zeumer-Spataro L, Roopenian DC, Morel L (2018) Inhibition of glycolysis reduces disease severity in an autoimmune model of rheumatoid arthritis. *Front Immunol* 9:1973. <https://doi.org/10.3389/fimmu.2018.01973>
- Agarwal S, Behring M, Kim HG, Bajpai P, Chakravarthi B, Gupta N, Elkholy A, Al Diffalha S, Varambally S, Manne U (2020) Targeting P4HA1 with a small molecule inhibitor in a colorectal cancer PDX model. *Transl Oncol* 13:100754. <https://doi.org/10.1016/j.tranon.2020.100754>
- Arnold M, Abnet CC, Neale RE, Vignat J, Giovannucci EL, McGlynn KA, Bray F (2020) Global burden of 5 major types of gastrointestinal cancer. *Gastroenterol* 159(335–349):e315. <https://doi.org/10.1053/j.gastro.2020.02.068>
- Beni FA, Kazemi M, Dianat-Moghadam H, Behjati M (2022) MicroRNAs regulating Wnt signaling pathway in colorectal cancer: biological implications and clinical potentials. *Funct Integr Genomics* 22:1073–1088. <https://doi.org/10.1007/s10142-022-00908-x>
- Billir LH, Schrag D (2021) Diagnosis and treatment of metastatic colorectal cancer: a review. *JAMA* 325:669–685. <https://doi.org/10.1001/jama.2021.0106>
- Cao XP, Cao Y, Li WJ, Zhang HH, Zhu ZM (2019) P4HA1/HIF1alpha feedback loop drives the glycolytic and malignant phenotypes of pancreatic cancer. *Biochem Biophys Res Commun* 516:606–612. <https://doi.org/10.1016/j.bbrc.2019.06.096>
- Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z (2017) Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep* 18:248–262. <https://doi.org/10.1016/j.celrep.2016.12.019>
- Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF, Chan WK, Li HN, Liu CC, Singh S, Chen WJ, Chen JJ, Yang PC (2007) A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356:11–20. <https://doi.org/10.1056/NEJMoa060096>
- Chen Y, Zheng X, Wu C (2021) The role of the tumor microenvironment and treatment strategies in colorectal cancer. *Front Immunol* 12:792691. <https://doi.org/10.3389/fimmu.2021.792691>
- Duan X, Tang X, Nair MS, Zhang T, Qiu Y, Zhang W, Wang P, Huang Y, Xiang J, Wang H, Schwartz RE, Ho DD, Evans T, Chen S (2021) An airway organoid-based screen identifies a role for the HIF1alpha-glycolysis axis in SARS-CoV-2 infection. *Cell Rep* 37:109920. <https://doi.org/10.1016/j.celrep.2021.109920>
- Eluri S, Parian AM, Limketkai BN, Ha CY, Brant SR, Dudley-Brown S, Efron JE, Fang SG, Gearhart SL, Marohn MR, Meltzer SJ, Bashar S, Truta B, Montgomery EA, Lazarev MG (2017) Nearly a third of high-grade dysplasia and colorectal cancer is undetected in patients with inflammatory bowel disease. *Dig Dis Sci* 62:3586–3593. <https://doi.org/10.1007/s10620-017-4652-5>
- Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33:1–22
- Frigerio S, Lartey DA, D'Haens GR, Grootjans J (2021) The role of the immune system in IBD-associated colorectal cancer: from pro to anti-tumorigenic mechanisms. *Int J Mol Sci* 22:12739. <https://doi.org/10.3390/ijms222312739>
- Ganesh K, Stadler ZK, Cercek A, Mendelsohn RB, Shia J, Segal NH, Diaz LA Jr (2019) Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol* 16:361–375. <https://doi.org/10.1038/s41575-019-0126-x>

- Geeleher P, Cox N, Huang RS (2014) pRRophetic: an R package for prediction of clinical chemotherapeutic response from tumor gene expression levels. *PLoS One* 9:e107468. <https://doi.org/10.1371/journal.pone.0107468>
- Hipkiss AR (2019) Aging, Alzheimer's disease and dysfunctional glycolysis; similar effects of too much and too little. *Aging Dis* 10:1328–1331. <https://doi.org/10.14336/AD.2019.0611>
- Hnatyszyn A, Hryhorowicz S, Kaczmarek-Rys M, Lis E, Slomski R, Scott RJ, Plawski A (2019) Colorectal carcinoma in the course of inflammatory bowel diseases. *Hered Cancer Clin Pract* 17:18. <https://doi.org/10.1186/s13053-019-0118-4>
- Irizarry-Caro RA, McDaniel MM, Overcast GR, Jain VG, Troutman TD, Pasare C (2020) TLR signaling adapter BCAP regulates inflammatory to reparatory macrophage transition by promoting histone lactylation. *Proc Natl Acad Sci U S A* 117:30628–30638. <https://doi.org/10.1073/pnas.2009778117>
- Kammerer U, Kapp M, Gassel AM, Richter T, Tank C, Dietl J, Ruck P (2001) A new rapid immunohistochemical staining technique using the EnVision antibody complex. *J Histochem Cytochem* 49:623–630. <https://doi.org/10.1177/002215540104900509>
- Kim N, Gim JA, Lee BJ, Choi BI, Park SB, Yoon HS, Kang SH, Kim SH, Joo MK, Park JJ, Kim C, Kim HK (2021) RNA-sequencing identification and validation of genes differentially expressed in high-risk adenoma, advanced colorectal cancer, and normal controls. *Funct Integr Genomics* 21:513–521. <https://doi.org/10.1007/s10142-021-00795-8>
- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform* 9:559. <https://doi.org/10.1186/1471-2105-9-559>
- Lorent M, Giral M, Foucher Y (2014) Net time-dependent ROC curves: a solution for evaluating the accuracy of a marker to predict disease-related mortality. *Stat Med* 33:2379–2389. <https://doi.org/10.1002/sim.6079>
- Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, Colaprico A, Czerwinska P, Mazurek S, Mishra L, Heyn H, Krasnitz A, Godwin AK, Lazar AJ, Cancer Genome Atlas Research N et al (2018) Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* 173(338–354):e315. <https://doi.org/10.1016/j.cell.2018.03.034>
- Mao C, Gao Y, Wan M, Xu N (2022) Identification of glycolysis-associated long non-coding RNA regulatory subtypes and construction of prognostic signatures by transcriptomics for bladder cancer. *Funct Integr Genomics* 22:597–609. <https://doi.org/10.1007/s10142-022-00845-9>
- Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP (2018) Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 28:1747–1756. <https://doi.org/10.1101/gr.239244.118>
- Monjabez AM, Giobbie-Hurder A, Lako A, Thrash EM, Brennick RC, Kao KZ, Manuszak C, Gentzler RD, Tesfaye A, Jabbour SK, Alese OB, Rahma OE, Cleary JM, Sharon E, Mamon HJ, Cho M, Streicher H, Chen HX, Ahmed MM et al (2021) A randomized trial of combined PD-L1 and CTLA-4 inhibition with targeted low-dose or hypofractionated radiation for patients with metastatic colorectal cancer. *Clin Cancer Res* 27:2470–2480. <https://doi.org/10.1158/1078-0432.CCR-20-4632>
- Nadeem MS, Kumar V, Al-Abbasi FA, Kamal MA, Anwar F (2020) Risk of colorectal cancer in inflammatory bowel diseases. *Semin Cancer Biol* 64:51–60. <https://doi.org/10.1016/j.semcancer.2019.05.001>
- Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, Panaccione R, Ghosh S, Wu JCY, Chan FKL, Sung JY, Kaplan GG (2017) Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 390:2769–2778. [https://doi.org/10.1016/S0140-6736\(17\)32448-0](https://doi.org/10.1016/S0140-6736(17)32448-0)
- Rajamaki K, Taira A, Katainen R, Valimaki N, Kuosmanen A, Plaketti RM, Seppala TT, Ahtiainen M, Wirta EV, Vartiainen E, Sulo P, Ravanti J, Lehtipuro S, Granberg KJ, Nykter M, Tanskanen T, Ristimaki A, Koskensalo S, Renkonen-Sinisalo L et al (2021) Genetic and epigenetic characteristics of inflammatory bowel disease-associated colorectal cancer. *Gastroenterol* 161:592–607. <https://doi.org/10.1053/j.gastro.2021.04.042>
- Riffelmacher T, Giles DA, Zahner S, Dicker M, Andreyev AY, McArdle S, Perez-Jeldres T, van der Gracht E, Murray MP, Hartmann N, Tumanov AV, Kronenberg M (2021) Metabolic activation and colitis pathogenesis is prevented by lymphotoxin beta receptor expression in neutrophils. *Mucosal Immunol* 14:679–690. <https://doi.org/10.1038/s41385-021-00378-7>
- Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43:e47. <https://doi.org/10.1093/nar/gkv007>
- Shen C, Xuan B, Yan T, Ma Y, Xu P, Tian X, Zhang X, Cao Y, Ma D, Zhu X, Zhang Y, Fang JY, Chen H, Hong J (2020) m(6)A-dependent glycolysis enhances colorectal cancer progression. *Mol Cancer* 19:72. <https://doi.org/10.1186/s12943-020-01190-w>
- Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, Hollmann TJ, Bruggeman C, Kannan K, Li Y, Elipenahli C, Liu C, Harbison CT, Wang L, Ribas A et al (2014) Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* 371:2189–2199. <https://doi.org/10.1056/NEJMoa1406498>
- Soto-Herederó G, de Las G, Heras MM, Gabande-Rodríguez E, Oller J, Mittelbrunn M (2020) Glycolysis – a key player in the inflammatory response. *FEBS J* 287:3350–3369. <https://doi.org/10.1111/febs.15327>
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545–15550. <https://doi.org/10.1073/pnas.0506580102>
- Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S, Szgyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S et al (2015) Proteomics. Tissue-based map of the human proteome. *Sci* 347:1260419. <https://doi.org/10.1126/science.1260419>
- Warburg O (1956) On the origin of cancer cells. *Sci* 123:309–314. <https://doi.org/10.1126/science.123.3191.309>
- Wickham H (2009) ggplot2: elegant graphics for data analysis (use R!), vol 10. Springer, New York, pp 978–970
- Wijnands AM, de Jong ME, Lutgens M, Hoentjen F, Elias SG, Oldenburg B, Dutch Initiative on Crohn and Colitis (ICC) (2021) Prognostic factors for advanced colorectal neoplasia in inflammatory bowel disease: systematic review and meta-analysis. *Gastroenterol* 160:1584–1598. <https://doi.org/10.1053/j.gastro.2020.12.036>
- Witters P, Honzik T, Bauchart E, Altassan R, Pascreau T, Bruneel A, Vuillaumier S, Seta N, Borgel D, Matthijs G, Jaeken J, Meersseman W, Cassiman D, Pascale de L, Morava E (2019) Long-term follow-up in PMM2-CDG: are we ready to start treatment trials? *Genet Med* 21:1181–1188. <https://doi.org/10.1038/s41436-018-0301-4>
- Xiong G, Stewart RL, Chen J, Gao T, Scott TL, Samayoa LM, O'Connor K, Lane AN, Xu R (2018) Collagen prolyl 4-hydroxylase 1 is essential for HIF-1 α stabilization and TNBC chemoresistance. *Nat Commun* 9:4456. <https://doi.org/10.1038/s41467-018-06893-9>

- Yaghoubi N, Soltani A, Ghazvini K, Hassanian SM, Hashemy SI (2019) PD-1/PD-L1 blockade as a novel treatment for colorectal cancer. *Biomed Pharmacother* 110:312–318. <https://doi.org/10.1016/j.biopha.2018.11.105>
- Yamada Y, Arai T, Sugawara S, Okato A, Kato M, Kojima S, Yamazaki K, Naya Y, Ichikawa T, Seki N (2018) Impact of novel oncogenic pathways regulated by antitumor miR-451a in renal cell carcinoma. *Cancer Sci* 109:1239–1253. <https://doi.org/10.1111/cas.13526>
- Yang Y, Han Z, Li X, Huang A, Shi J, Gu J (2020) Epidemiology and risk factors of colorectal cancer in China. *Chin J Cancer Res* 32:729–741. <https://doi.org/10.21147/j.issn.1000-9604.2020.06.06>
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, Trevino V, Shen H, Laird PW, Levine DA, Carter SL, Getz G, Stemke-Hale K, Mills GB, Verhaak RG (2013) Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 4:2612. <https://doi.org/10.1038/ncomms3612>
- Zhang Q, Liu W, Zhang HM, Xie GY, Miao YR, Xia M, Guo AY (2020) hTFtarget: a comprehensive database for regulations of human transcription factors and their targets. *Genom Proteom Bioinform* 18:120–128. <https://doi.org/10.1016/j.gpb.2019.09.006>

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