OncoTargets and Therapy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Glycosylation-Related Gene Signature Identifies MFNG as a Key Driver of Proliferation and Metastasis in Colorectal Cancer

Xinji Gao[1,*], Qiang Li[1,*], Qingshui Wang[2,*], Jun Wang[1], Lan Zhao[1], Ting Yan[1], Xiang Yu[1]

[1]General Surgery Department, the Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, People's Republic of China; [2]The Affiliated People's Hospital, College of Integrative Medicine, Fujian-Hong Kong-Macau-Taiwan Collaborative Laboratory for the Inheritance and Innovation of Traditional Chinese Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, Fujian, People's Republic of China

*These authors contributed equally to this work

Correspondence: Xiang Yu; Ting Yan, Email 278129817@qq.com; yanting@fjtcm.edu.cn

Background: Colorectal cancer (CRC) ranks as the third most common malignancy and second leading cause of cancer-related mortality worldwide. Aberrant glycosylation has emerged as a hallmark of cancer, yet systematic analyses of glycosylation-related gene expression patterns and their prognostic implications in CRC remain limited.

Methods: We conducted comprehensive analyses of 214 glycosylation-related genes using TCGA and GEO datasets. Differential expression analysis identified significantly altered genes, followed by LASSO Cox regression to construct a four-gene glycosylation-Related Gene Signature (GRGS). We validated the model across multiple independent cohorts and performed functional experiments with MFNG knockdown in CRC cell lines and zebrafish xenograft models.

Results: We identified 54 differentially expressed glycosylation-related genes in CRC tissues. The four-gene signature comprising MFNG (manic fringe), UST (uronyl 2-sulfotransferase), SLC35D1 (solute carrier family 35 member D1), and GALNT7 (polypeptide N-acetylgalactosaminyltransferase 7) demonstrated robust prognostic performance across validation cohorts. GRGS-High patients exhibited significantly shorter overall survival and were associated with advanced tumor stages. MFNG emerged as the top predictor, with high expression correlating with poor survival. Functional validation confirmed that MFNG knockdown significantly inhibited CRC cell proliferation, migration, and invasion both in vitro and in vivo.

Conclusion: Our study establishes GRGS as a reliable prognostic tool for CRC risk stratification and identifies MFNG as a promising therapeutic target. These findings provide valuable insights into glycosylation-mediated CRC progression and offer potential clinical applications for precision oncology.

Keywords: colorectal cancer, glycosylation, prognostic signature, MFNG, biomarker

Introduction

Colorectal cancer (CRC) ranks among the most prevalent malignancies worldwide, representing the third most commonly diagnosed cancer and the second leading cause of cancer-related mortality.[1–3] Despite substantial advances in screening programs and therapeutic interventions, the five-year survival rate for metastatic CRC remains alarmingly low, underscoring an urgent need for novel prognostic biomarkers and therapeutic targets to enhance patient outcomes.[4]

Glycosylation-the enzymatic attachment of carbohydrate moieties to proteins and lipids-stands as one of the most abundant and intricate post-translational modifications in mammalian cells.[5–7] This essential process regulates diverse cellular functions, including protein folding, stability, trafficking, cell-cell recognition, and signal transduction. Mounting evidence positions aberrant glycosylation as a hallmark of cancer, where tumor cells display glycan profiles distinctly altered from those in normal tissues.[8,9] Such modifications fuel various oncogenic processes, from uncontrolled proliferation and apoptosis evasion to enhanced invasion, metastasis, angiogenesis, and immune evasion.[10–12]

https://doi.org/10.2147/OTT.S572377
Received: 6 October 2025
Accepted: 4 February 2026
Published: 23 March 2026

In CRC, dysregulated glycosylation emerges as a pivotal driver of malignant transformation and progression. Perturbations in the expression of glycosyltransferases, glycosidases, and related enzymes generate tumor-associated carbohydrate antigens, such as sialyl Lewis structures, truncated O-glycans, and aberrant N-glycan branching.[13–15] These changes promote tumor cell adhesion to endothelium, facilitate epithelial-mesenchymal transition (EMT), and support immune escape via mechanisms like immune checkpoint modulation. Moreover, glycosylation alterations contribute to chemotherapy resistance, positioning them as promising therapeutic targets.[16,17]

Nevertheless, despite its acknowledged role in CRC pathogenesis, systematic analyses of glycosylation-related gene expression patterns and their prognostic implications remain scarce. Prior research has predominantly examined isolated enzymes or specific glycan structures, often overlooking a holistic assessment of the glycosylation machinery. Furthermore, the construction of reliable prognostic models incorporating these genes, alongside their interplay with the tumor immune microenvironment, warrants deeper investigation.

To address these gaps, we conducted a comprehensive analysis of glycosylation-related genes to map their differential expression in CRC tissues compared to normal controls. We developed and validated a four-gene glycosylation-Related Gene Signature (GRGS) via LASSO Cox regression. In addition, we investigated GRGS associations with clinical features, somatic mutations, Hallmark pathways, and immune infiltration patterns. Functional validation through MFNG knockdown in CRC cell lines and zebrafish xenograft models further confirmed its critical role in promoting proliferation and metastasis.

Methods

Data Acquisition and Preprocessing

RNA sequencing data for colorectal carcinoma specimens, expressed as fragments per kilobase million (FPKM) values, were obtained from The Cancer Genome Atlas (TCGA) database via the Sangerbox analytical platform (http://vip.sangerbox.com/home.html),[18] accompanied by comprehensive clinical metadata for each patient. Additional validation datasets were retrieved from the Gene Expression Omnibus (GEO) repository, specifically encompassing three independent cohorts: GSE17536, GSE39582, and GSE87211. The GSE17536 and GSE39582 cohorts were analyzed using the [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, while the GSE87211 dataset was profiled on the Agilent-026652 Whole Human Genome Microarray 4x44K v2 (Probe Name version). To ensure analytical consistency across platforms, FPKM expression values were normalized and transformed to transcripts per kilobase million (TPM) units prior to downstream computational analyses. The GSE17536 dataset comprised transcriptomic profiles from 177 colorectal cancer patients, while GSE39582 included 443 malignant colorectal specimens alongside 19 corresponding non-neoplastic colonic mucosa controls. The GSE87211 cohort provided a larger comparative framework with 203 rectal cancer samples paired against 160 normal mucosal tissue controls. In this study, the TCGA dataset served as the discovery cohort for differential expression analysis, prognostic gene screening, and GRGS model construction. The three GEO datasets (GSE17536, GSE39582, and GSE87211) were employed as independent external validation cohorts to assess the robustness and generalizability of the GRGS model. Additionally, TCGA and GSE39582 datasets, which contain paired tumor-normal samples, were utilized for comparative GRGS score analysis between tumor and normal tissues.

Identification of Glycosylation-Related Gene

Glycosylation-related genes were obtained from the GlycoGene Database (https://acgg.asia/db/ggdb/), a curated database containing comprehensive information on genes involved in glycosylation pathways. A total of 214 glycosylation-related genes were downloaded from this database for subsequent analysis. These genes encompass various functional categories involved in glycan biosynthesis, modification, and degradation processes.

Differential Gene Expression Analysis

Comparative transcriptomic profiling of glycosylation-related gene between colorectal carcinoma specimens and their corresponding adjacent non-malignant tissues in the TCGA dataset was conducted using the Linear Models for

Microarray Data (limma) statistical framework implemented in R environment. To identify genes exhibiting significant expression alterations, we applied stringent filtering criteria: genes demonstrating an absolute $\log_2$ fold change ($|\log_2 FC|$) ≥ 0.585 (equivalent to 1.5-fold expression difference) combined with a nominal p-value < 0.05 were designated as differentially expressed genes (DEGs).[19] While we acknowledge that multiple testing correction (eg., FDR) is standard, we opted for raw p-values in this exploratory phase to minimize the risk of excluding potentially relevant glycosylation-related genes before the rigorous LASSO Cox regression screening. To visualize the results, volcano plots were generated using the "ggplot2" package in R,[20] and heatmaps of DEG expression were created using the "pheatmap" package.

Functional Enrichment Analysis

To elucidate the underlying molecular mechanisms and cellular functions governed by the identified differentially expressed genes, we conducted comprehensive pathway enrichment profiling using two complementary approaches: Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping and Gene Ontology Biological Process (GO-BP) term annotation. The KEGG analysis facilitated the identification of metabolic and signaling cascades, while GO-BP assessment provided insights into specific biological functions and cellular processes. Specifically, for KEGG analysis, we obtained the latest gene annotations as the background reference using the KEGG REST API (https://www.kegg.jp/kegg/rest/keggapi.html).[21] For GO-BP analysis, the c5.go.bp.v7.4.symbols.gmt gene set was downloaded from the Molecular Signatures Database (MSigDB; http://www.gsea-msigdb.org/gsea/downloads.jsp) to serve as the background. Enrichment analyses were performed using the clusterProfiler R package (version 3.14.3).[22] The analysis parameters were set with a minimum gene set size of 5 and a maximum of 5000, and a p-value < 0.05 was considered statistically significant.

Development of GRGS in Colorectal Cancer

For the identification of glycosylation-related genes with clinical prognostic significance, we performed an initial screening using univariate Cox proportional hazards regression on the TCGA dataset. Genes demonstrating statistical significance ($p < 0.05$) were subsequently subjected to Least Absolute Shrinkage and Selection Operator (LASSO) penalized Cox regression modeling to minimize overfitting and enhance feature selection precision. The LASSO algorithm applies L1 penalty regularization, systematically shrinking coefficients of irrelevant variables toward zero while preserving those with meaningful prognostic contributions. Through this iterative process, variables retaining non-zero coefficients after regularization were designated as the final prognostic predictors. The GRGS risk score formula was constructed as a linear combination of selected gene expression levels weighted by their corresponding LASSO-derived coefficients: $GRGS = \Sigma(\beta_i \times Expression_i)$, where β_i represents the regression coefficient for gene i, and $Expression_i$ denotes the normalized expression value of the corresponding gene.[23]

Survival Analysis and Patient Stratification

For survival analysis, the optimal cutoff values for risk stratification were determined using the surv_cutpoint function from the survminer R package.[24] Based on these optimal cutoffs, patients were categorized into high-risk and low-risk groups to maximize the statistical difference in survival outcomes. Kaplan-Meier survival curves were generated using the survival (version 3.5–7) and survminer (version 0.4.9) packages. Differences in survival between groups were assessed using the Log rank test. All statistical analyses were conducted using R software (version 4.2.3).

Immune Infiltration Analysis

To comprehensively evaluate the immune cell composition within the colorectal cancer microenvironment, we utilized pre-calculated immune infiltration scores obtained from the Sangerbox 3.0 platform (http://vip.sangerbox.com/home.html).[18] Specifically, Immune Score, Stromal Score, and ESTIMATE Score were retrieved based on the ESTIMATE algorithm. The relative abundances of 22 tumor-infiltrating immune cell types were quantified using CIBERSORT (employing the LM22 signature matrix with 1,000 permutations),[25] and 64 immune and stromal cell types were assessed using xCell.[26] All immune infiltration data provided by the Sangerbox platform were derived from the TCGA transcriptomic dataset, with input matrices processed as log2(TPM+1) values to ensuring analytical consistency.".

Single-Cell Transcriptomic Analysis

For high-resolution characterization of cellular heterogeneity within the tumor ecosystem, we utilized the CRC_GSE146771_10X dataset retrieved from the Tumor Immune Single-cell Hub 2 (TISCH2, http://tisch.comp-genomics.org).[27] TISCH2 employs a uniform processing pipeline for quality control and data normalization to ensure comparability across datasets. In this study, we adopted the pre-defined cell-type annotations and clustering results provided directly by the TISCH2 platform without performing re-clustering or re-annotation. The specific dataset contains single-cell transcriptomic profiles of colorectal cancer tissues processed via the 10x Genomics platform. We utilized the pre-calculated Uniform Manifold Approximation and Projection (UMAP) coordinates provided by TISCH2 to visualize the distribution of major cell lineages and to examine the expression patterns of GRGS component genes at single-cell resolution.

Machine Learning

We conducted supervised classification using the Extreme Smart Analysis platform (https://www.xsmartanalysis.com/), running in R 4.2.3 and Python 3.11.4 environments. Model inputs and outcome: Normalized expression values of the four signature genes SLC35D1, MFNG, UST, and GALNT7 were used as continuous predictors; the binary outcome was tissue status (Tumor vs. Normal). Algorithms and key hyperparameters: Nine algorithms were benchmarked: XGB Classifier (objective=binary:logistic), Logistic Regression (penalty=l2, C=1.0, max_iter=100, tol=1e-4), LGBM Classifier (boosting_type=gbdt, learning_rate=0.1, n_estimators=100, max_depth=−1, num_leaves=31), Random Forest Classifier (criterion=gini, n_estimators=20, max_depth=None, min_impurity_decrease=0.0), AdaBoost Classifier (learning_rate=1.0, n_estimators=50), Decision Tree Classifier (criterion=gini, max_depth=None, min_samples_split=2, min_samples_leaf=1), Gradient Boosting Classifier (learning_rate=0.1, loss=log_loss, max_depth=3, min_samples_split=2, min_samples_leaf=1, n_estimators=100), GaussianNB (var_smoothing=1e-09), and Complement NB (alpha=1.0). Train-validation strategy and cross-validation: Within-cohort evaluation adopted 2-fold cross-validation to estimate mean ROC and SD. The platform also reported model-specific cutoffs derived from ROC Youden optimization. Class imbalance handling: Class proportions were checked at model setup; no additional resampling or class-weighting was applied by the platform in the reported runs. Evaluation metrics: We reported AUC (with 95% CI), Accuracy, Sensitivity, Specificity, PPV, NPV, F1-score, and Cohen's Kappa. Feature importance: We applied L1-regularized Logistic Regression for feature importance with parameters: penalty=l1, C≈0.9885, max_iter=100, tol=1e-4.

Cell Culture

Two human colorectal adenocarcinoma cell lines, HCT-116 and DLD1, obtained from the American Type Culture Collection (ATCC), were employed for this investigation. Cell maintenance was performed using Dulbecco's Modified Eagle's Medium (DMEM) enriched with 10% (v/v) fetal bovine serum and 1% (v/v) penicillin-streptomycin antibiotic solution. Cells were maintained under standard culture conditions at 37°C in a 5% CO_2 incubator with saturated humidity.

Quantitative Real-Time Polymerase Chain Reaction (qRT-PCR)

Total cellular RNA extraction was performed using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol. Subsequently, complementary DNA (cDNA) synthesis was carried out employing the PrimeScript RT reagent Kit (TaKaRa Bio, Kusatsu, Japan) following standard reverse transcription procedures. Gene expression quantification was executed using SYBR Green-based real-time PCR with SYBR Premix Ex Taq II (TaKaRa Bio) on a quantitative PCR system. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was employed as an endogenous housekeeping gene for normalization purposes. Relative gene expression levels were calculated using the comparative CT method ($2^{-\Delta\Delta Ct}$), with fold-change values normalized to the control group. The oligonucleotide primer sequences designed for MFNG amplification were forward primer 5′-TGCTGAGTTCGACACCTTCTT-3′ and reverse primer 5′-CCCCTTGGGTTCACATAGTTG-3′. Reference gene GAPDH was amplified using the following primer pair: forward sequence 5′-GGTGAAGGTCGGAGTCAAC-3′ and reverse sequence 5′-CAAATGAGCCCCAGCCTTC-3′.

Cell Proliferation Assay

Cell viability and proliferative capacity were determined using the Cell Counting Kit-8 (CCK-8) assay protocol. Briefly, cells were seeded into 96-well microplates at a density of 2×10^3 cells per well. Following incubation periods of 0, 24, 48, and 72 hours, CCK-8 reagent was added to individual wells and subsequently incubated for 2 hours at 37°C. Optical density measurements were obtained at 450 nm wavelength using a microplate spectrophotometer.[28]

Western Blot Analysis

Cells were lysed in RIPA buffer supplemented with protease and phosphatase inhibitors (Roche Applied Science, Mannheim, Germany). Protein concentrations were determined using the BCA Protein Assay Kit (Pierce Biotechnology, Rockford, IL, USA). Equal amounts of protein were separated by SDS-PAGE and transferred onto PVDF membranes (Millipore, Bedford, MA, USA). Membranes were blocked with 5% non-fat milk in TBST (Tris-Buffered Saline with Tween 20) and incubated overnight at 4°C with primary antibodies: Anti-MFNG (1:1000 dilution, Novus Biologicals,USA) or Anti-GAPDH (1:50000 dilution, Proteintech, China). Finally, IRDye® 680RD Goat anti-Mouse IgG or IRDye® 800CW Goat Anti-Rabbit IgG were used to quantify the proteins using the Odyssey® CLx Infrared Imaging System (LI-COR Biosciences).

Migration Assay

Cell migration capacity was assessed through an in vitro scratch wound assay. Monolayer cultures were established by growing cells to 100% confluency in 6-well culture plates. A uniform linear wound was created across the cell monolayer using a sterile 10-μL pipette tip. After gentle washing with phosphate-buffered saline (PBS) to remove cellular debris, cultures were maintained in serum-free medium. Photographic documentation was performed immediately post-wounding (0 h) and at 48 hours post-injury using an inverted phase-contrast microscope. Quantitative analysis of wound gap closure was conducted using ImageJ image analysis software.[29]

Invasion Assay

To determine cellular invasive potential, a Matrigel-coated Transwell chamber system was employed. Briefly, 2×10^4 cells suspended in serum-depleted medium were seeded into the upper compartment of the invasion chambers, while the lower compartment was replenished with complete medium supplemented with 10% fetal bovine serum serving as a chemoattractant. Following a 48-hour incubation period, non-invasive cells remaining on the upper membrane surface were gently removed using a cotton swab. Cells that had successfully penetrated through the Matrigel barrier were subsequently fixed with methanol, stained using crystal violet solution, and quantified through microscopic enumeration.[30]

Zebrafish-Based Tumor Xenograft System

A translucent zebrafish larval model was developed to assess HCT-116 colorectal adenocarcinoma cell behavior under in vivo conditions, enabling dual analysis of tumor proliferation and metastatic capacity. Wild-type zebrafish (Danio rerio) acquired from Fuzhou Bio-Service Biotechnology Co. Ltd (Fuzhou, China) were housed in recirculating aquaria at 28.5°C with standardized 14-hour light/10-hour dark photoperiods. Prior to transplantation, HCT-116 cells underwent enzymatic dissociation using trypsin solution and subsequent fluorescent tagging with 5 μM DiI (1,1′-dioctadecyl -3,3,3′,3′-tetramethylindocarbocyanine perchlorate; Meilun Biotechnology, China), a lipophilic membrane-intercalating crimson fluorophore that facilitates longitudinal tracking of cellular movement and replication. Cell microinjection was performed on 2–3 days post-fertilization (dpf) larvae using calibrated GB100T-8P glass microcapillaries coupled with FemtoJet 4i precision microinjectors (Eppendorf, Germany), with distinct anatomical sites selected based on experimental endpoints: for proliferation studies, approximately 200 fluorescently-labeled cells were precisely delivered into the nutrient-rich central yolk sac to create a confined microenvironment conducive to growth assessment, while for metastasis evaluation, an identical cell number was introduced into the circulation-accessible perivitelline space to permit systemic dissemination. Each experimental cohort comprised ten embryos, ensuring adequate statistical power for

quantitative analyses. Tumor progression was monitored via fluorescence microscopy at specific intervals: proliferation was assessed at 2 hours (baseline) and 48 hours post-implantation through whole-body imaging to document initial distribution and subsequent volumetric expansion; metastatic behavior was characterized by imaging caudal region fluorescence at 2 hours and 24 hours post-implantation to quantify distant cell migration. All procedures involving pre-feeding stage larvae (younger than 5 dpf) were exempt from formal ethics committee review according to institutional guidelines, and the experimental design followed ARRIVE standards for responsible animal research documentation.

Statistical Analysis

Data are presented as mean ± standard deviation (SD). In vitro cell culture experiments were performed with six technical replicates (n=6) per group, and in vivo zebrafish xenograft experiments included ten biological replicates (n=10) per group. The statistical correlation was assessed using Pearson's correlation coefficient. Comparisons between two groups were performed using the Student's *t*-test. For the analysis of cell proliferation measured by the CCK-8 assay, a two-way ANOVA was applied to evaluate the effects of group allocation (sh-NC vs. sh-MFNG) and time points (0, 24, 48, 72 h), as well as their interaction. This was followed by Bonferroni's post-hoc test for multiple comparisons. Survival rates were evaluated using Kaplan-Meier survival curves, with comparisons made between groups via Log rank test. A *p*-value of less than 0.05 was considered indicative of statistical significance.

Results

Analysis of Expression Patterns of Glycosylation-Related Genes in Colorectal Cancer

Dysregulated glycosylation is a hallmark of cancer progression, significantly impacting tumor biology and patient outcomes. Understanding glycosylation-related gene expression patterns in colorectal cancer is crucial for identifying novel therapeutic targets and prognostic biomarkers.

To analyze the expression patterns of glycosylation-related genes in colorectal cancer, we conducted a comprehensive analysis of TCGA database. Through differential gene expression analysis, we identified 30 significantly downregulated and 24 upregulated genes among 214 glycosylation-related genes compared to adjacent normal tissues (Figure 1A and B). KEGG pathway enrichment analysis revealed that the 24 upregulated genes were predominantly enriched in critical pathways including N-Glycan biosynthesis, Metabolic pathways, and Glycosaminoglycan biosynthesis - keratan sulfate (Figure 1C). Complementarily, GO-BP analysis demonstrated that these upregulated genes were primarily associated with Carbohydrate derivative biosynthetic process, Carbohydrate derivative metabolic process, and Organonitrogen compound biosynthetic process (Figure 1D). In contrast, KEGG pathway analysis of the 30 downregulated genes showed enrichment in Metabolic pathways, Glycosphingolipid biosynthesis-lacto and neolacto series, and Mucin type O-glycan biosynthesis (Figure 1E). Correspondingly, GO-BP analysis indicated that these downregulated genes were mainly enriched in Carbohydrate derivative metabolic process, Carbohydrate derivative biosynthetic process, and Organonitrogen compound biosynthetic process (Figure 1F). Subsequently, we performed a comprehensive mutational analysis of the 54 differentially expressed genes in colorectal cancer. To visualize the most significant genomic alterations, we selected the top ten genes with the highest mutation frequencies among the identified DEGs for oncoprint analysis. The results revealed that UGGT2, B4GALNT4, and POMGNT1 exhibited relatively high mutation frequencies, with UGGT2 showing the highest alteration rate at 23% (Figure 1G). Notably, survival analysis demonstrated that mutations in four genes-UGGT2, GCNT4, PIGZ, and GALNT12-were significantly associated with poor prognosis in colorectal cancer patients (Figure 1H–K).

Construction and Validation of a Prognostic Model Based on Glycosylation-Related Genes

Following our differential expression analysis, we evaluated the prognostic impact of the 54 differentially expressed glycosylation-related genes on colorectal cancer (CRC) patient outcomes. Survival analysis identified four genes-MFNG, UST, SLC35D1, and GALNT7-that were significantly associated with CRC patient prognosis (Figure 2A). Notably, high expression levels of MFNG (HR=1.44, 95% CI: 1.11–1.86) and UST (HR=1.47, 95% CI: 1.11–1.94) were associated

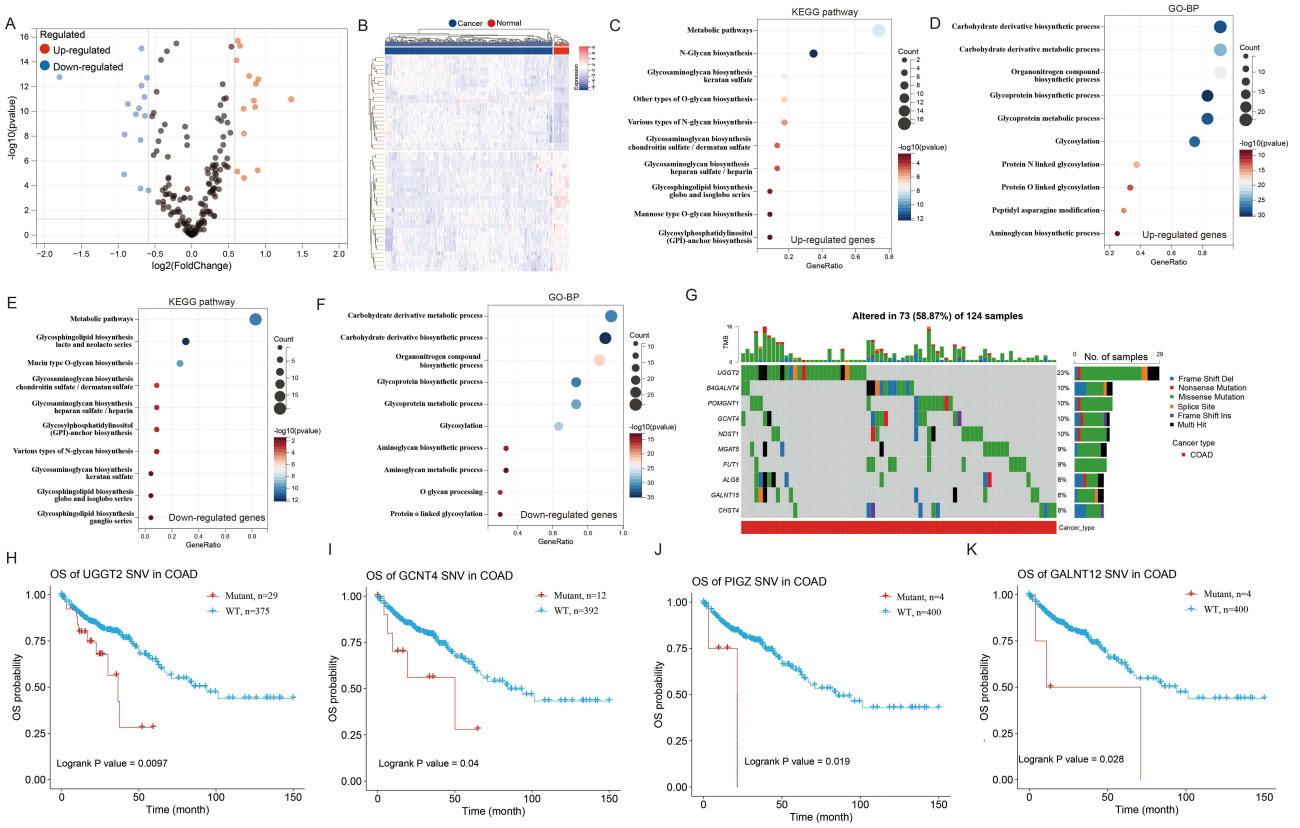

Figure 1 Expression patterns and prognostic relevance of glycosylation-related genes in colorectal cancer. (**A**) Volcano plot displaying differentially expressed glycosylation-related genes between colorectal cancer and adjacent normal tissues based on TCGA database analysis. Red dots represent upregulated genes, blue dots represent downregulated genes. (**B**) Heatmap showing hierarchical clustering of differentially expressed glycosylation-related genes in cancer versus normal samples. (**C**) KEGG pathway enrichment analysis of 24 upregulated glycosylation-related genes. (**D**) Gene Ontology Biological Process (GO-BP) enrichment analysis of upregulated glycosylation-related genes. (**E**) KEGG pathway enrichment analysis of 30 downregulated glycosylation-related genes. (**F**) GO-BP enrichment analysis of downregulated glycosylation-related genes. (**G**) Oncoprint showing mutation profiles of differentially expressed glycosylation-related genes in colorectal cancer samples. (**H–K**) Kaplan-Meier survival curves illustrating the association between mutations in UGGT2 (**H**), GCNT4 (**I**), PIGZ (**J**), and GALNT12 (**K**) genes and overall survival in colorectal cancer patients.

with unfavorable outcomes, while low expression of SLC35D1 (HR=0.68, 95% CI: 0.49–0.94) and GALNT7 (HR=0.76, 95% CI: 0.59–0.98) correlated with poor prognosis in CRC patients.

Based on these four prognostically significant genes, we established a predictive model for CRC patient outcomes using the least absolute shrinkage and selection operator (LASSO) Cox regression analysis in the TCGA training cohort (Figure 2B). This model, termed glycosylation-Related Gene Signature (GRGS), demonstrated optimal performance at λ=1.5e-3, as determined by partial likelihood deviance minimization. The regression coefficients for each gene in the signature were determined (Figure 2C), with positive coefficients for risk-associated genes (MFNG, UST) and negative coefficients for protective genes (SLC35D1, GALNT7). The risk score distribution, survival status, and expression patterns of the four signature genes for each CRC patient in the cohort are visualized in Figure 2D. Patients were stratified into high-risk and low-risk groups based on the median GRGS score. The heatmap clearly illustrates that risk-associated genes (MFNG and UST) showed elevated expression in the high-risk group, whereas protective genes (SLC35D1 and GALNT7) exhibited higher expression in the low-risk group. To evaluate the diagnostic potential of the model, we compared the GRGS scores between tumor and normal tissues. The results showed that GRGS scores were significantly elevated in tumor samples compared to normal controls in both the TCGA (Figure 2E) and GSE39582 (Figure 2F) cohorts. Kaplan-Meier survival analysis demonstrated that patients in the GRGS-High group had significantly shorter overall survival compared to those in the GRGS-Low group (Figure 2G).

To rigorously assess the robustness and generalizability of our GRGS, we conducted external validation using independent cohorts. Consistent with our findings in the discovery cohort, patients classified as GRGS-High exhibited significantly poorer overall survival in the GSE17536 (Figure 2H) and GSE39582 (Figure 2I) datasets. We further

https://doi.org/10.2147/OTT.S572377

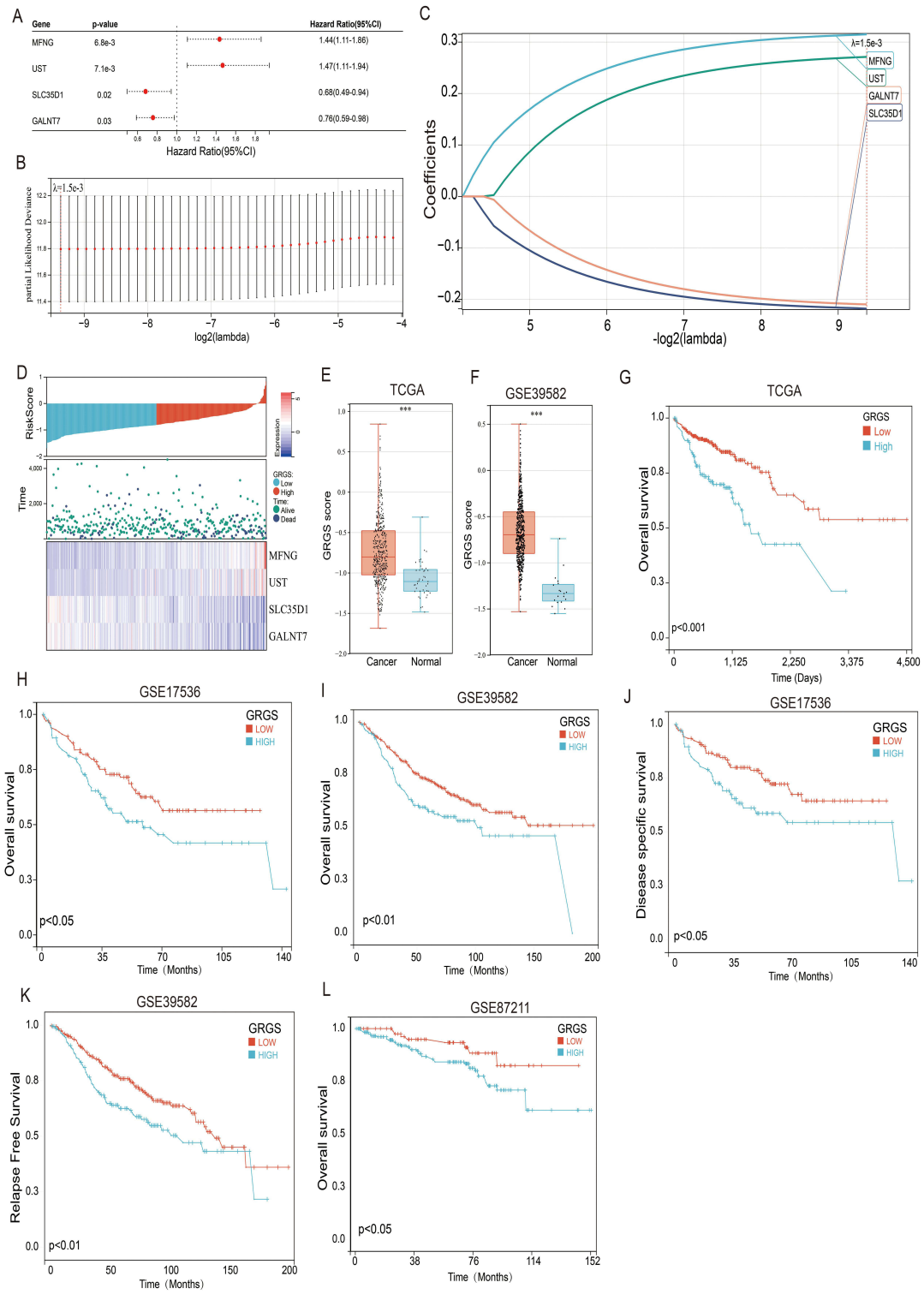

Figure 2 Construction of a prognostic model based on glycosylation-related genes in colorectal cancer. (**A**)Forest plot showing the association between four glycosylation-related genes and overall survival in colorectal cancer patients, including hazard ratios with 95% confidence intervals and p-values. (**B**) Tuning parameter (λ) selection in the LASSO regression model by minimum criteria. The partial likelihood deviance curve was plotted versus log2(λ), with the optimal λ value of 0.0015 determined by 10-fold cross-validation. (**C**) LASSO coefficient profiles of the four prognostic glycosylation-related genes at the optimal λ value. (**D**) Distribution of GRGS risk scores, survival status, and expression patterns of the four signature genes in the TCGA colorectal cancer cohort. (**E** and **F**) Boxplots comparing GRGS risk scores between tumor tissues and normal tissues in the TCGA (**E**) and GSE39582 (**F**) cohorts. (**G**) Kaplan-Meier curves for overall survival comparing the high-risk and low-risk groups stratified by the GRGS model. (**H** and **I**) Kaplan-Meier curves for overall survival in (**H**) GSE17536 and (**I**) GSE39582 datasets comparing patients in GRGS-High and GRGS-Low risk groups. (**J**) Kaplan-Meier curves for disease-specific survival in GSE17536 dataset comparing patients in GRGS-High and GRGS-Low risk groups. (**K**) Kaplan-Meier curves for relapse-free survival in GSE39582 dataset comparing patients in GRGS-High and GRGS-Low risk groups. (**L**) Kaplan-Meier curves for overall survival in GSE87211 datasets comparing patients in GRGS-High and GRGS-Low risk groups. ***, $p < 0.001$.

https://doi.org/10.2147/OTT.S572377

evaluated the model's performance across different survival endpoints, observing significantly shorter disease-specific survival (DSS) in the GSE17536 dataset (Figure 2J) and inferior relapse-free survival (RFS) in the GSE39582 cohort (Figure 2K) for high-risk patients. Notably, to further verify the broad applicability of our model across different anatomical subsites of colorectal cancer, we validated the GRGS in the GSE87211 dataset, which consists specifically of rectal cancer samples. The results confirmed that high GRGS scores were significantly associated with poor overall survival (Figure 2L) in rectal cancer patients, thereby reinforcing the robustness and clinical utility of our signature.

Correlation Analysis Between GRGS and Clinical Characteristics of Colorectal Cancer

Subsequently, we conducted a comprehensive analysis to investigate the associations between GRGS and various clinical parameters in colorectal cancer patients, including sex, histological type, age, cancer status, and tumor stage. The results demonstrated that patients with Stage III&IV tumors exhibited significantly higher GRGS values compared to those with Stage I&II tumors (Figure 3A–E). Notably, this finding suggests that GRGS may serve as a potential indicator for advanced disease progression and tumor aggressiveness in colorectal cancer patients.

To further elucidate the molecular landscape underlying different GRGS subgroups, we performed comprehensive genetic mutation analysis. In the GRGS-Low subgroup, the ten most frequently mutated genes were APC (67.4%), TP53 (55.4%), TTN (53.8%), KRAS (38.0%), SYNE1 (34.8%), MUC16 (31.5%), PIK3CA (30.4%), RYR2 (27.2%), ZFHX4 (26.6%), and OBSCN (25.0%) (Figure 3F). In the GRGS-High subgroup, the ten most frequently mutated genes were

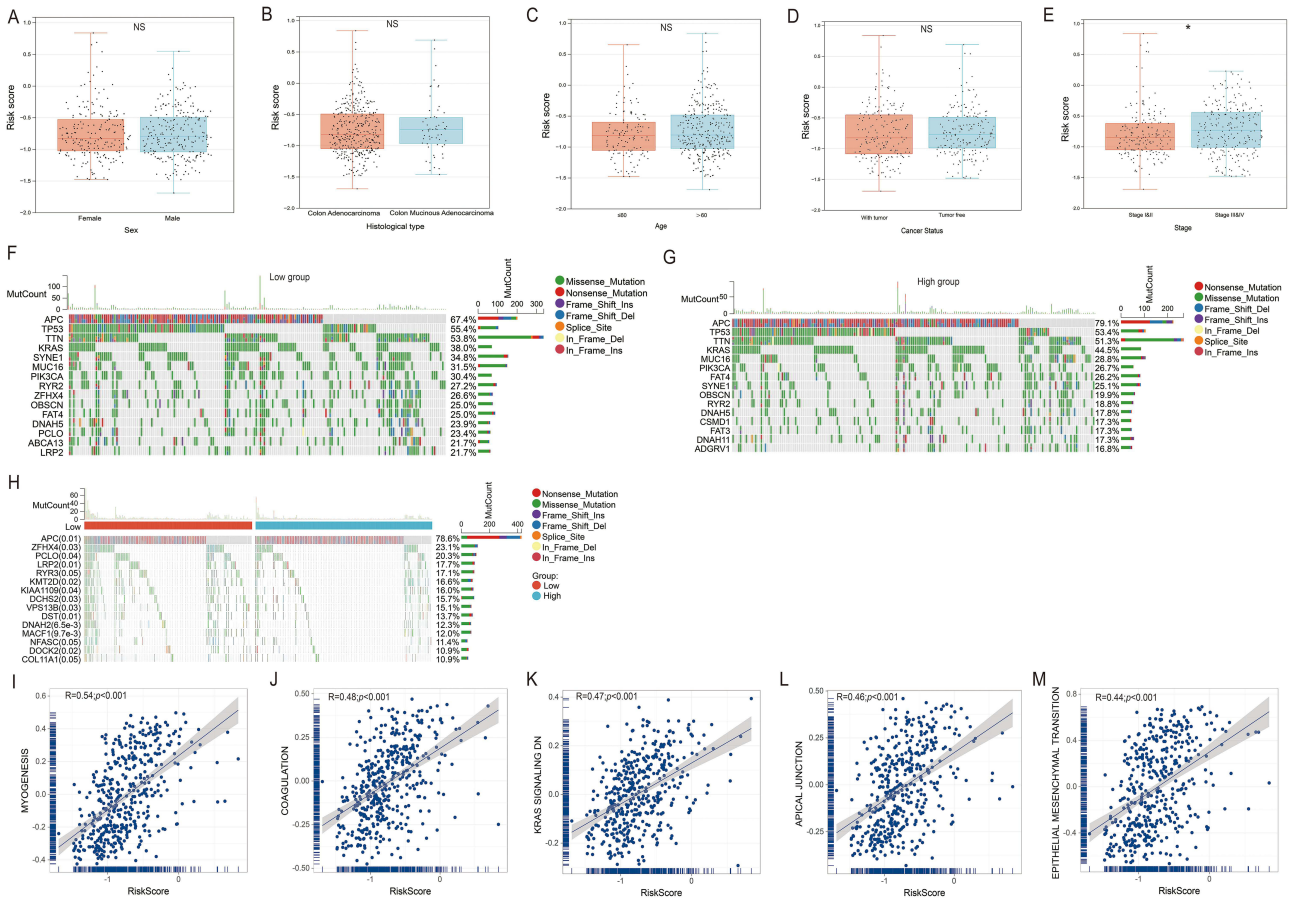

Figure 3 Correlation analysis of GRGS with clinical features, mutation profiles, and hallmark pathways in colorectal cancer. (**A**–**E**) GRGS distribution across clinical subgroups: sex, histological type, age, cancer status, and tumor stage. (**F**) Top 10 mutated genes in GRGS-Low subgroup. (**G**) Top 10 mutated genes in GRGS-High subgroup. (**H**) Comparative analysis of APC mutation rates between GRGS subgroups. (**I**–**M**) Correlation scatter plots between GRGS and hallmark pathways: Myogenesis (**I**), Coagulation (**J**), KRAS Signaling down (**K**), Apical Junction (**L**), Epithelial Mesenchymal Transition (**M**). *, $p < 0.05$.
Abbreviation: NS, Not Significant.

https://doi.org/10.2147/OTT.S572377

APC (79.1%), TP53 (53.4%), TTN (51.3%), KRAS (44.5%), MUC16 (28.8%), PIK3CA (26.7%), FAT4 (26.2%), SYNE1 (25.1%), OBSCN (19.9%), and RYR2 (18.8%) (Figure 3G). Comparative analysis between the GRGS-Low and GRGS-High subgroups revealed that the GRGS-High subgroup exhibited significantly higher APC gene mutation rates (Figure 3H).

Furthermore, we explored the functional implications of GRGS by analyzing its correlations with Hallmark pathway signatures. The results revealed that GRGS demonstrated significant positive correlations with multiple critical biological processes, including Myogenesis (R=0.54), Coagulation (R=0.48), KRAS Signaling down (R=0.47), Apical Junction (R=0.46), and Epithelial Mesenchymal Transition (R=0.44) pathways (Figure 3I–M). These findings suggest that higher GRGS values are associated with enhanced activation of pathways involved in tumor progression, metastasis, and therapeutic resistance, providing valuable insights into the biological mechanisms underlying genomic risk stratification in colorectal cancer.

Correlation of GRGS with the Immune Microenvironment in Colorectal Cancer

To build upon our previous findings regarding the prognostic significance of the GRGS, we next examined its association with the tumor immune microenvironment. Using the ESTIMATE algorithm, we calculated Immune Score, Stromal Score, and ESTIMATE Score for each colorectal cancer sample. As shown in Figure 4A, patients in the GRGS-High group exhibited significantly higher Immune Scores, Stromal Scores, and combined ESTIMATE Scores than those in the GRGS-Low group.

We then applied CIBERSORT to deconvolute the relative abundances of 22 immune cell types. The heatmap in Figure 4A summarizes the landscape of immune infiltration across GRGS groups. Quantitative comparisons revealed that GRGS-High tumors harbored markedly increased naïve B cells and M2 macrophages (Figure 4B), whereas resting $CD4^+$ memory T cells and activated mast cells were significantly more abundant in the GRGS-Low group. To validate M2 macrophage enrichment, we performed xCELL analysis (Figure 4C), which confirmed a pronounced increase in M2 macrophage scores in the GRGS-High cohort.

To dissect gene expression patterns underpinning these differences at single-cell resolution, we analyzed the CRC_GSE146771_10X scRNA-seq dataset. The t-SNE plot in Figure 4D delineates major immune and stromal lineages-B cells, conventional $CD4^+$ T cells (CD4Tconv), $CD8^+$ T cells (CD8T), exhausted $CD8^+$ T cells (CD8Tex), mast cells, monocytes/macrophages (Mono/Macro), NK cells, plasma cells, and proliferating T cells (Tprolif). Violin plots (Figure 4E) compare expression of four GRGS component genes (GALNT7, MFNG, SLC35D1, UST) across these clusters. MFNG consistently shows the highest expression levels among all four genes in B cells, CD4Tconv, CD8T, CD8Tex, mast cells, Mono/Macro, NK cells, plasma cells, and Tprolif cells. Finally, gene-specific t-SNE overlays (Figures 4F–I) highlight the spatial enrichment of SLC35D1, MFNG, GALNT7, and UST within the tumor microenvironment.

MFNG Overexpression Predicts Poor Prognosis in Colorectal Cancer

To refine our diagnostic panel and identify the most clinically relevant marker, we compared nine machine-learning classifiers-XXGBoost, Logistic Regression, LightGBM, Random Forest, AdaBoost, Decision Tree, GBDT, Gaussian Naive Bayes (GNB), and Complement Naive Bayes (CNB)-using expression levels of GALNT7, MFNG, SLC35D1, and UST to distinguish tumor from adjacent normal tissue. Although XGBoost achieved perfect discrimination in the training set (AUC = 1.000; Figure 5A), it underperformed in the independent validation cohort, whereas logistic regression maintained high accuracy (AUC = 0.998; Figure 5B). This discrepancy suggests that XGBoost may be overfitted, while logistic regression offers greater stability. Feature-importance analysis via L1-regularized logistic regression ranked MFNG as the top predictor (Figure 5C), leading us to focus subsequent analyses on this gene.

Kaplan-Meier survival curves revealed that high MFNG expression was consistently associated with worse overall survival (OS) in GSE39582, GSE72970, GSE87211, and TCGA database (Figure 5D–G). Furthermore, patients with elevated MFNG levels exhibited significantly shorter disease-free survival (DFS) in GSE161158 and TCGA database (Figure 5H and I), reduced disease-specific survival (DSS) in GSE87211 and TCGA database (Figure 5J and K), and poorer progression-free survival (PFS) in TCGA database (Figure 5L). In the TCGA-CRC cohort, MFNG expression was

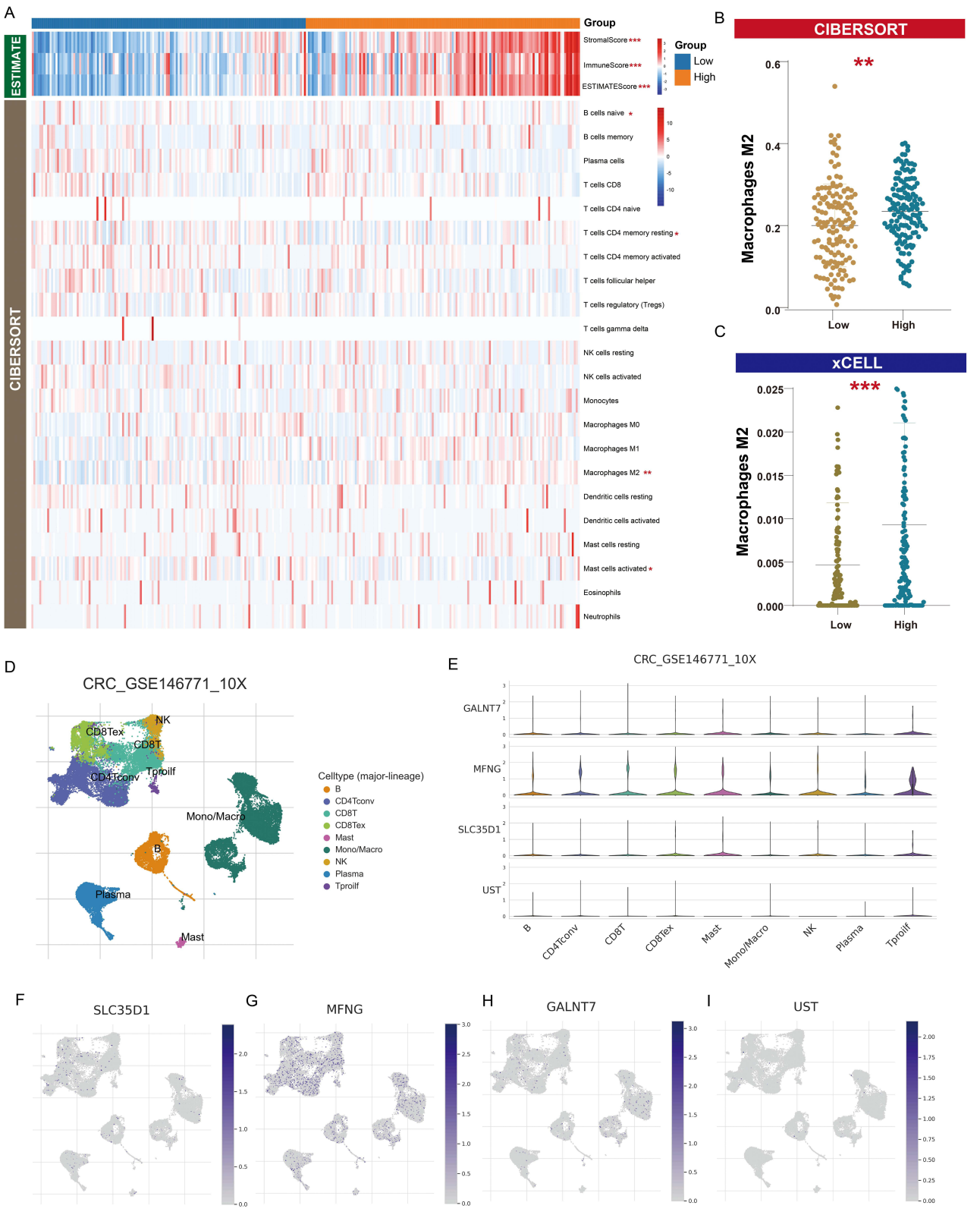

Figure 4 Correlation of GRGS with the immune microenvironment in colorectal cancer. (**A**) Heatmap showing ESTIMATE-derived Immune Score, Stromal Score and ESTIMATE Score alongside CIBERSORT-inferred proportions of 22 immune cell types in GRGS-High versus GRGS-Low groups. (**B** and **C**) Boxplots of and M2 macrophage abundances estimated by (**B**) CIBERSORT and (**C**) xCELL. (**D**) t-SNE plot of major immune and stromal cell clusters in the CRC_GSE146771_10X scRNA-seq dataset. (**E**) Violin plots comparing expression levels of GALNT7, MFNG, SLC35D1 and UST across identified cell lineages. (**F**–**I**) Gene-specific t-SNE overlays showing the spatial distribution of SLC35D1 (**F**), MFNG (**G**), GALNT7 (**H**) and UST (**I**) within the single-cell atlas. **, $p < 0.01$; ***, $p < 0.001$.

https://doi.org/10.2147/OTT.S572377

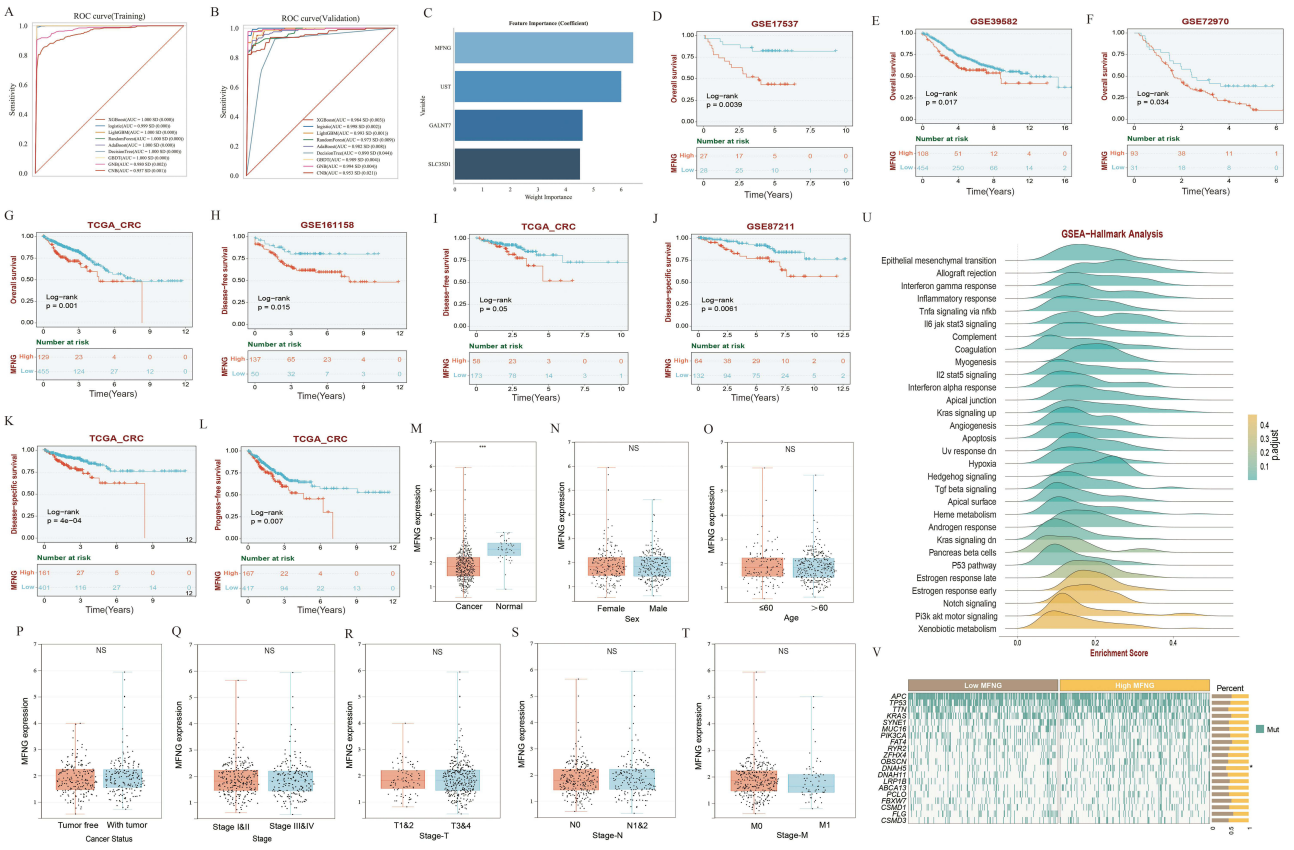

Figure 5 Prognostic evaluation of MFNG expression in colorectal cancer. (**A**) ROC curves of nine machine-learning classifiers in the training cohort. (**B**) ROC curves of the same classifiers in the independent validation cohort. (**C**) Feature importance ranking by L1-regularized logistic regression. (**D**–**G**) Kaplan–Meier curves for overall survival in GSE39582, GSE72970, GSE87211 and TCGA cohorts. (**H** and **I**) Kaplan–Meier curves for disease-free survival in GSE161158 and TCGA cohorts. (**J** and **K**) Kaplan–Meier curves for disease-specific survival in GSE87211 and TCGA cohorts. (**L**) Kaplan–Meier curve for progression-free survival in the TCGA cohort. (**M**) Box-plots comparing MFNG expression between tumor and normal tissuesin TCGA database. (**N**–**T**) Association between MFNG expression and (**N**) sex, (**O**) age, (**P**) recurrence, (**Q**) stage, (**R**) T, (**S**) N, (**T**) M in TCGA database. (**U**) Gene Set Enrichment Analysis (Hallmark) of high versus low MFNG expression tumors. (**V**) Comparison of somatic mutation frequencies between high- and low-MFNG groups. ***, $p < 0.001$.

Abbreviation:NS, Not Significant.

significantly lower in tumor tissues than in normal tissues (Figure 5M). Clinical correlation analysis further indicated no significant association between MFNG expression and sex, age, tumor recurrence, tumor stage, deeper tumor invasion (T stage), lymph node metastasis (N stage), and distant metastasis (M stage) (Figure 5N–T).

To explore underlying mechanisms, we performed Gene Set Enrichment Analysis (GSEA) using Hallmark gene sets. High MFNG tumors were enriched for pathways that drive tumor progression and immune modulation, including Epithelial Mesenchymal Transition, inflammatory response, TNF-α signaling via NF-κB, IL6-JAK-STAT3 signaling, complement activation, and coagulation (Figure 5U). Finally, somatic mutation profiling showed that MFNG-high tumors harbored a significantly higher mutation frequency in DNAH5 compared to MFNG-low cases (Figure 5V).

MFNG Knockdown Suppresses Proliferation and Metastatic Capacity in Colorectal Cancer Cells

Given the established prognostic significance of MFNG in colorectal cancer, we sought to elucidate its functional contribution to oncogenic processes through loss-of-function studies. We generated stable MFNG-knockdown HCT-116 cells using two independent shRNA constructs (sh-MFNG1 and sh-MFNG2). The knockdown efficiency was rigorously verified at both the transcriptional and translational levels. qRT-PCR analysis confirmed a significant reduction in MFNG mRNA levels (Figure 6A), and Western blotting further demonstrated a substantial decrease in MFNG protein expression compared to the negative control (Figure 6B). Functional characterization revealed profound effects on multiple cancer

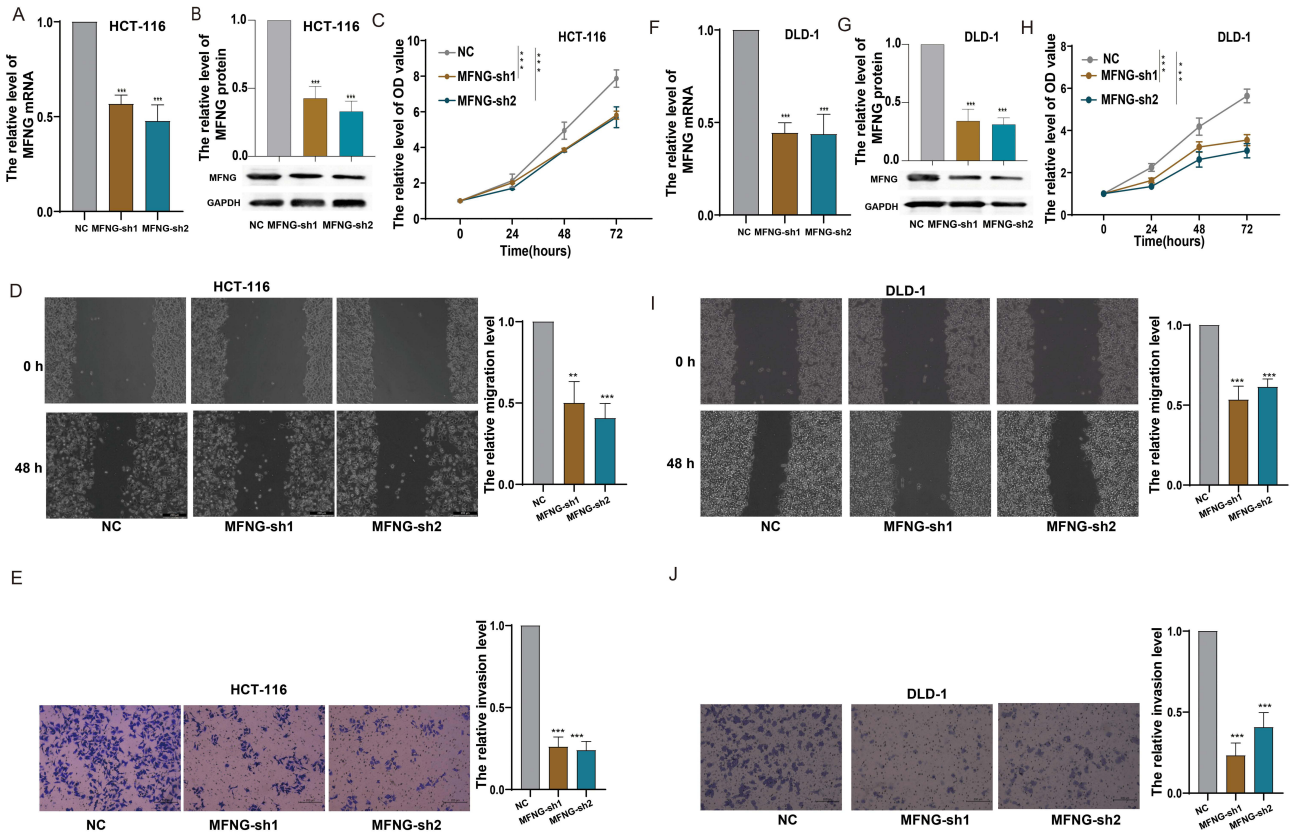

Figure 6 MFNG knockdown inhibits proliferation and metastatic capacity in colorectal cancer cells. (**A**) qRT-PCR analysis of MFNG expression in HCT-116 cells transduced with control shRNA or MFNG shRNA. (**B**) Western blot analysis and quantification of MFNG protein levels, confirming knockdown efficiency. (**C**) CCK-8 proliferation assay of HCT-116 cells with MFNG knockdown. (**D**) Wound healing assay showing migration of HCT-116 cells at 0 and 48 hours after scratch. (**E**) Transwell invasion assay of HCT-116 cells through Matrigel-coated membranes. (**F**) qRT-PCR analysis of MFNG expression in DLD-1 cells with control or MFNG knockdown. (**G**) Western blot analysis and quantification of MFNG protein expression. (**H–J**) CCK-8 proliferation (**H**), wound healing migration (**I**), and Transwell invasion (**J**) assays in DLD-1 cells with MFNG knockdown. Data represent mean ± SD from six independent experiments. **, $p < 0.01$; ***, $p < 0.001$.

hallmarks. CCK-8 proliferation assays demonstrated that MFNG depletion significantly attenuated the growth kinetics of HCT-116 cells compared to scrambled control cells (Figure 6C). Wound healing assays showed markedly impaired migratory capacity in MFNG-knockdown cells, with substantially reduced wound closure rates at 48 hours post-scratch (Figure 6D). Moreover, Transwell invasion assays confirmed that MFNG knockdown significantly diminished the invasive potential of HCT-116 cells through Matrigel-coated membranes (Figure 6E). To validate these findings across different colorectal cancer models, we established MFNG-knockdown DLD-1 stable cell lines. Consistent with HCT-116 results, we observed successful suppression of MFNG expression at both mRNA (Figure 6F) and protein levels (Figure 6G) in DLD-1 cells. Subsequent functional experiments demonstrated that MFNG suppression significantly inhibited cellular proliferation (Figure 6H), migration (Figure 6I), and invasion capabilities (Figure 6J). These convergent results across two independent cell line models establish MFNG as a critical driver of aggressive phenotypes in colorectal cancer.

To validate our in vitro findings in a physiologically relevant context, we established a zebrafish xenograft model using fluorescently labeled HCT-116 cells to assess tumor proliferation and metastatic potential in vivo. We first evaluated tumor cell proliferation by quantifying fluorescence intensity in the yolk sac region over time (Figure 7A). At 2 hours post-transplantation (hpt), both MFNG-knockdown groups showed comparable fluorescence areas to the negative control, indicating similar initial tumor cell distribution and successful engraftment. However, by 48 hours post-transplantation, both MFNG-knockdown groups exhibited significantly reduced yolk fluorescence areas compared to controls, demonstrating that MFNG knockdown effectively suppressed tumor cell proliferation within the zebrafish host environment. To evaluate metastatic capacity, we analyzed tumor cell dissemination to the tail region as a surrogate

https://doi.org/10.2147/OTT.S572377

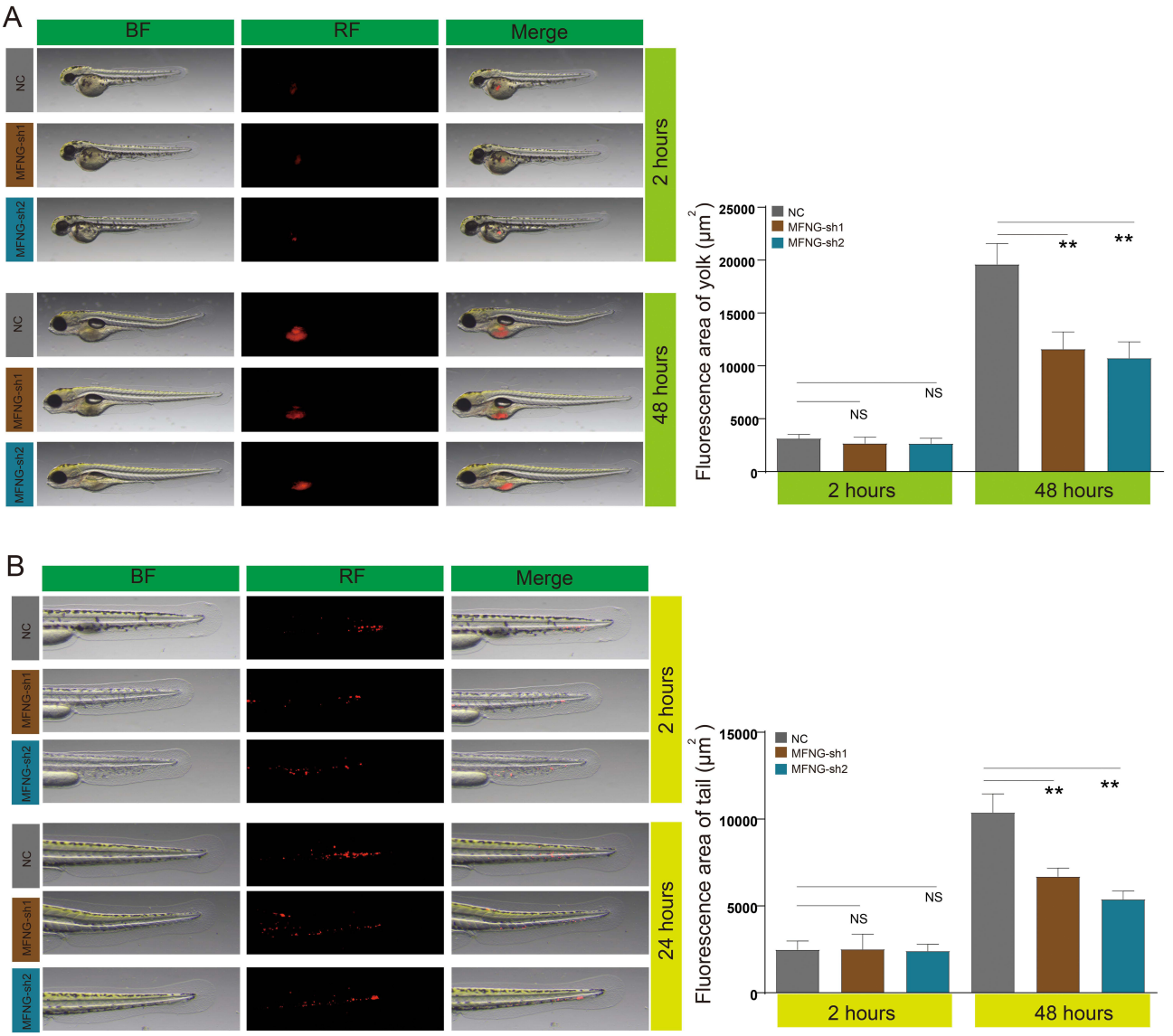

Figure 7 Zebrafish xenograft model for assessing MFNG function in tumor proliferation and metastasis. (**A**) Representative images of zebrafish embryos injected with fluorescently labeled HCT-116 cells under bright field (BF), red fluorescence (RF), and merged conditions. Quantification of yolk fluorescence area at 2- and 48-hours post-transplantation. (**B**) Representative images showing tumor cell dissemination to tail region. Quantification of tail fluorescence area at 2- and 24-hours post-injection. **, $p < 0.01$.

Abbreviation: NS, Not Significant.

marker for cancer cell migration and invasion (Figure 7B). At 2 hours post-injection, tail fluorescence areas showed no statistically significant differences between MFNG-knockdown and control groups, confirming uniform initial cell distribution. Remarkably, by 24 hours, both MFNG-knockdown groups displayed significantly reduced tail fluorescence areas compared to controls, indicating that MFNG depletion markedly impaired tumor cell migration and dissemination capacity. These comprehensive in vivo findings corroborate our in vitro observations and provide compelling evidence that MFNG serves as a critical regulator of colorectal cancer progression.

Discussion

This study provides a comprehensive investigation into the role of glycosylation-related genes in colorectal cancer, establishing a robust prognostic signature and elucidating the functional significance of MFNG in tumor progression. Our findings reveal that dysregulated glycosylation is a hallmark of CRC pathogenesis, characterized by coordinated

 https://doi.org/10.2147/OTT.S572377

transcriptomic and genomic alterations that collectively drive tumor aggressiveness and impact patient outcomes. The GRGS demonstrates remarkable prognostic accuracy across multiple independent cohorts, while functional validation confirms MFNG as a critical oncogenic driver with therapeutic potential.

Our systematic analysis began by mapping the landscape of glycosylation dysregulation in CRC. The differential expression analysis identified 54 glycosylation-related genes with significant alterations in CRC tissues compared to adjacent normal tissues, including 30 downregulated and 24 upregulated genes. This transcriptomic dysregulation was functionally characterized by enrichment in critical pathways such as N-Glycan biosynthesis and Glycosphingolipid biosynthesis, suggesting profound metabolic reprogramming of the glycosylation machinery in tumor cells. Importantly, these expression alterations were complemented by genomic evidence showing that mutations in key glycosylation genes-specifically UGGT2, GCNT4, PIGZ, and GALNT12-were significantly associated with poor prognosis. This convergence of transcriptomic and genomic aberrations indicates that glycosylation defects in CRC arise through multiple molecular mechanisms and are not merely epiphenomena but rather intrinsic drivers of aggressive tumor behavior.

However, given the substantial heterogeneity of CRC and the complexity of interconnected glycosylation pathways, relying on individual gene alterations-whether at the expression or mutation level-is insufficient for robust clinical risk stratification. Therefore, to translate these multifaceted molecular perturbations into a clinically actionable tool, we employed an integrative approach to construct a multigene prognostic signature. The construction of GRGS using LASSO Cox regression represents this translational step, distilling the system-wide glycosylation dysregulation into a parsimonious four-gene model. The signature incorporates MFNG and UST as risk-enhancing genes and SLC35D1 and GALNT7 as protective factors, reflecting the complex interplay between different glycosylation pathways in CRC biology. The model's consistent performance across TCGA and three validation cohorts (GSE17536, GSE39582, GSE87211) for overall survival, disease-specific survival, and relapse-free survival underscores its robustness and suggests that GRGS captures fundamental biological processes independent of population-specific variables. This progression-from identifying widespread glycosylation dysregulation and its genetic underpinnings to constructing an integrated prognostic tool-provides a conceptually unified framework for understanding and clinically applying glycosylation alterations in CRC.

While recent studies have also explored glycosylation-related prognostic models in CRC, our study offers distinct contributions. For instance, Chuang et al constructed a risk model emphasizing drug sensitivity and immune infiltration but lacked direct experimental validation of the identified biomarkers.[31] Chen et al utilized single-cell sequencing and machine learning to develop a glycosyltransferase-associated signature (GARS), yet their study primarily focused on algorithmic benchmarking rather than investigating the biological function of specific driver genes.[32] Similarly, Li et al established a seven-gene signature specific to pMMR CRC patients based on a ceRNA network.[33] In contrast to these purely bioinformatic or subtype-specific approaches, our study not only establishes a robust GRGS applicable to general CRC cohorts but, crucially, integrates lab validation. We experimentally confirmed that MFNG, the core risk gene in our model, functions as a key oncogenic driver promoting CRC proliferation and metastasis in vitro and in vivo (zebrafish xenografts). This "bench-to-bedside" approach bridges the gap between computational prediction and biological mechanism.

Recent research has significantly expanded our understanding of the four GRGS component genes in cancer biology. MFNG, a member of the glycosyltransferase Fringe family, modulates Notch signaling through O-fucose glycosylation and has emerged as a context-dependent oncogene in multiple malignancies.[34,35] Studies demonstrate that MFNG overexpression activates oncogenic Notch3 signaling in breast cancer and correlates with poor prognosis across various cancer types, including pancreatic adenocarcinoma.[36] UST, encoding uronyl 2-sulfotransferase, regulates glycosaminoglycan sulfation patterns and has been implicated in melanoma metastasis, where UST knockdown significantly reduces metastatic potential through alterations in chondroitin sulfate/dermatan sulfate modifications.[37] SLC35D1 is a critical nucleotide sugar transporter located in the endoplasmic reticulum membrane. This protein plays essential roles in cellular glycosylation processes and skeletal development. SLC35D1 functions as a UDP-sugar transporter, facilitating the transport of UDP-glucuronic acid and UDP-N-acetylgalactosamine into the ER lumen for various biosynthetic pathways.[38,39] GALNT7, a polypeptide N-acetylgalactosaminyltransferase, exhibits dual roles depending on cancer type-

functioning as an oncogene in prostate cancer where it promotes tumor growth through O-glycosylation modifications,[40] while serving as a tumor suppressor in melanoma where its downregulation by miR-30b/30d enhances metastatic behavior.[41]

Clinically, GRGS showed strong association with advanced tumor stage (III&IV), indicating its value as an indicator of disease progression. The significantly higher APC mutation rate in GRGS-High subgroups further suggests a link between glycosylation dysregulation and canonical Wnt pathway activation.[42–44] This connection is biologically plausible given APC's role in regulating cell adhesion and migration-processes modulated by glycosylation. The positive correlations between GRGS and Hallmark pathways like EMT and KRAS Signaling provide mechanistic explanations for its prognostic power, revealing how glycosylation alterations promote metastasis and therapeutic resistance.

Our immune microenvironment analysis revealed that GRGS-High tumors exhibit significantly elevated immune and stromal scores. Paradoxically, this inflamed microenvironment correlates with poor prognosis, likely due to the predominance of immunosuppressive M2 macrophages and naïve B cells in GRGS-High tumors. Single-cell RNA sequencing confirmed that MFNG-the top predictor in our diagnostic panel-is widely expressed across immune cell types, suggesting its role in shaping the tumor-immune crosstalk. These findings have important implications for immunotherapy, as GRGS stratification could identify patients less responsive to checkpoint inhibitors due to immunosuppressive microenvironments.

The central role of MFNG in CRC progression was substantiated through multiple lines of evidence. Machine-learning classifiers identified MFNG as the most significant diagnostic marker, with high expression consistently predicting poor survival across all endpoints (OS, DFS, DSS, PFS) in multiple cohorts. GSEA revealed that MFNG-high tumors are enriched in EMT, inflammatory response, and IL6-JAK-STAT3 pathways-all critical drivers of metastasis and therapy resistance. The higher DNAH5 mutation frequency in MFNG-high tumors suggests potential links to genomic instability, warranting further investigation into MFNG's role in DNA repair mechanisms.

Functional validation provided compelling evidence for MFNG's oncogenic properties. Knockdown experiments in HCT-116 and DLD-1 cell lines demonstrated that MFNG depletion significantly inhibits proliferation, migration, and invasion. The zebrafish xenograft model further confirmed these findings in vivo, showing reduced tumor growth and metastatic dissemination upon MFNG knockdown. This concordance across two cell lines and an in vivo model strengthens the conclusion that MFNG is a bona fide therapeutic target. The zebrafish model's particular strength lies in its ability to track real-time metastatic behavior, providing insights unavailable through conventional assays. Interestingly, our differential expression analysis in the TCGA cohort revealed that MFNG mRNA levels were significantly lower in CRC tissues compared to adjacent normal mucosa. This finding presents an apparent paradox, as our survival analysis consistently identified high MFNG expression as a predictor of poor prognosis. This discrepancy suggests that MFNG may play a complex, stage-dependent role in CRC progression. While its downregulation might be associated with the initial loss of normal epithelial architecture during tumorigenesis, higher expression levels in established tumors may be required to drive aggressive phenotypes, such as EMT and metastasis. Thus, MFNG likely functions as a "context-dependent" factor, where its re-expression or maintenance in tumor cells confers a survival and migratory advantage in later disease stages.

Several limitations warrant consideration. First, In the initial differential expression analysis, we utilized raw p-values rather than FDR-adjusted q-values to maximize the pool of candidate genes for downstream modeling. While this strategy reduces the risk of type II errors (false negatives), it may inadvertently increase type I errors (false positives); however, the subsequent rigorous LASSO regression and multi-cohort validation steps were employed to mitigate this issue. Second, we acknowledge a methodological limitation regarding the machine learning analysis used for feature prioritization. In the initial screening phase, we employed the four LASSO-selected genes to train additional classifiers (eg., XGBoost, Random Forest) on the same dataset. We recognize that performing feature selection and model training on the same dataset introduces feature selection bias and may result in optimistically inflated performance metrics (overfitting). We clarify that these machine learning models served primarily as an exploratory strategy for feature importance ranking to systematically identify the most critical driver gene (MFNG) for downstream experimental validation, rather than to propose a definitive multi-gene diagnostic classifier for clinical deployment. The robustness of MFNG as a biomarker was subsequently validated through survival analyses in independent cohorts and functional

assays, mitigating the reliance on the initial training set performance. Third, the retrospective nature of our clinical data necessitates prospective validation of GRGS in randomized trials. Fourth, while bioinformatic analyses suggest strong immune microenvironment associations, these require experimental confirmation using spatial transcriptomics or multiplex immunohistochemistry. Finally, the precise molecular mechanisms through which MFNG influences metastasis-particularly its downstream targets-remain incompletely characterized. Future studies should explore whether MFNG modulates glycosylation of specific integrins or growth factor receptors involved in invasion.

Conclusion

In conclusion, this study establishes glycosylation dysregulation as a central determinant of CRC aggressiveness. GRGS provides a clinically applicable tool for prognosis prediction, while MFNG emerges as both a biomarker and therapeutic target. Our integrated approach-combining bioinformatics, machine learning, and functional validation-offers a blueprint for translating molecular signatures into clinical practice. Future work should focus on developing MFNG-targeted therapies and validating GRGS in prospective interventional trials to advance precision oncology in CRC.

Data Sharing Statement

The datasets used and/or analyzed during the current study are available from the corresponding author (Xiang Yu) on reasonable request.

Ethical Statement

The datasets analyzed in this study were retrieved from publicly accessible repositories, including The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO). All data were fully anonymized and collected in accordance with established ethical standards. As this research involved only retrospective analysis of de-identified data and did not include any direct interaction with human participants, it was reviewed by the Ethics Committee of the Second Affiliated Hospital of Fujian University of Traditional Chinese Medicine and determined to be exempt from ethical approval. This exemption is consistent with national regulations, specifically Items (1) and (2) of Article 32 of the Measures for the Ethical Review of Life Science and Medical Research Involving Human Subjects (Revised on February 18, 2023, China). Accordingly, this study raises no ethical concerns and presents no conflicts of interest. In addition, experiments conducted on zebrafish larvae younger than five days post-fertilization do not require ethical approval. All animal-related procedures were reported in compliance with the ARRIVE guidelines.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work was supported by grants from Fujian Provincial Department of Finance (C2024004) and Fujian University of Traditional Chinese Medicine (X2024037).

Disclosure

The authors declare that they have no competing interests in this work.

References

1. Abedizadeh R, Majidi F, Khorasani HR, Abedi H, Sabour D. Colorectal cancer: a comprehensive review of carcinogenesis, diagnosis, and novel strategies for classified treatments. *Cancer Metastasis Rev*. 2024;43(2):729–753. doi:10.1007/s10555-023-10158-3
2. Underwood PW, Ruff SM, Pawlik TM. Update on targeted therapy and immunotherapy for metastatic colorectal cancer. *Cells*. 2024; 13.

3. Suydam CR, Schlussel AT. Management of oligometastatic colorectal cancer. *Surg Clin North Am*. 2024;104(3):619–629. doi:10.1016/j.suc.2023.11.011
4. Haas S, Mikkelsen AH, Kronborg CJSCJS, et al. Management of treatment-related sequelae following colorectal cancer. *Colorectal Dis*. 2023;25(3):458–488. doi:10.1111/codi.16299
5. Chatham JC, Patel RP. Protein glycosylation in cardiovascular health and disease. *Nat Rev Cardiol*. 2024;21(8):525–544. doi:10.1038/s41569-024-00998-z
6. Čaval T, Alisson-Silva F, Schwarz F. Roles of glycosylation at the cancer cell surface: opportunities for large scale glycoproteomics. *Theranostics*. 2023;13(8):2605–2615. doi:10.7150/thno.81760
7. Gędaj A, Gregorczyk P, Żukowska D, et al. Glycosylation of FGF/FGFR: an underrated sweet code regulating cellular signaling programs. *Cytokine Growth Factor Rev*. 2024;77:39–55. doi:10.1016/j.cytogfr.2024.04.001
8. Pinho SS, Reis CA. Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer*. 2015;15(9):540–555. doi:10.1038/nrc3982
9. Silsirivanit A. Glycosylation markers in cancer. *Adv Clin Chem*. 2019;89:189–213.
10. Xu X, Peng Q, Jiang X, et al. Altered glycosylation in cancer: molecular functions and therapeutic potential. *Cancer Commun*. 2024;44(11):1316–1336. doi:10.1002/cac2.12610
11. Thomas D, Rathinavel AK, Radhakrishnan P. Altered glycosylation in cancer: a promising target for biomarkers and therapeutics. *Biochim Biophys Acta Rev Cancer*. 2021;1875(1):188464. doi:10.1016/j.bbcan.2020.188464
12. Sun R, Kim AMJ, Lim SO. Glycosylation of Immune Receptors in Cancer. *Cells*. 2021; 10.
13. Zhang L, Ten Hagen KG. Pleiotropic effects of O-glycosylation in colon cancer. *J Biol Chem*. 2018;293(4):1315–1316. doi:10.1074/jbc.H117.812826
14. Krug J, Rodrian G, Petter K, et al. N-glycosylation regulates intrinsic IFN-γ resistance in colorectal cancer: implications for immunotherapy. *Gastroenterology*. 2023;164(3):392–406.e5. doi:10.1053/j.gastro.2022.11.018
15. Gao T, Wen T, Ge Y, et al. Disruption of Core 1-mediated O-glycosylation oppositely regulates CD44 expression in human colon cancer cells and tumor-derived exosomes. *Biochem Biophys Res Commun*. 2020;521(2):514–520. doi:10.1016/j.bbrc.2019.10.149
16. Zhu L, Chen Y, Du H, et al. N-glycosylation of CD82 at Asn157 is required for suppressing migration and invasion by reversing EMT via Wnt/β-catenin pathway in colon cancer. *Biochem Biophys Res Commun*. 2022;629:121–127. doi:10.1016/j.bbrc.2022.08.079
17. Hung J-S, Huang J, Lin Y-C, et al. C1GALT1 overexpression promotes the invasive behavior of colon cancer cells through modifying O-glycosylation of FGFR2. *Oncotarget*. 2014;5(8):2096–2106. doi:10.18632/oncotarget.1815
18. Chen D, Xu L, Xing H, et al. Sangerbox 2: enhanced functionalities and update for a comprehensive clinical bioinformatics data analysis platform. *Imeta*. 2024;3(5):e238. doi:10.1002/imt2.238
19. Wang Q, Weng S, Zhong W, et al. Modulation of DAPK1 expression by its alternative splice variant DAPK1-215 in cancer. *J Transl Med*. 2025;23(1):85. doi:10.1186/s12967-025-06127-9
20. Cao T, Li Q, Huang Y, Li A. plotnineSeqSuite: a python package for visualizing sequence data using ggplot2 style. *BMC Genomics*. 2023;24(1):585. doi:10.1186/s12864-023-09677-8
21. Huckvale E, Moseley HNB. kegg_pull: a software package for the restful access and pulling from the Kyoto encyclopedia of gene and genomes. *BMC Bioinf*. 2023;24(1):78. doi:10.1186/s12859-023-05208-0
22. Wang Q, Liu J, Li R, et al. Assessing the role of programmed cell death signatures and related gene TOP2A in progression and prognostic prediction of clear cell renal cell carcinoma. *Cancer Cell Int*. 2024;24(1):164. doi:10.1186/s12935-024-03346-w
23. Wang M, Guo Y, Xu Y, et al. Unraveling the role of programmed cell death gene signature and THBS1 in gastric cancer progression and therapy response. *J Gastroenterol Hepatol*. 2025;40(7):1825–1837. doi:10.1111/jgh.16987
24. Li X, Zhang L, Liu C, et al. Construction of mitochondrial quality regulation genes-related prognostic model based on bulk- RNA -seq analysis in multiple myeloma. *Biofactors*. 2025;51(1):e2135. doi:10.1002/biof.2135
25. Chen B, Khodadoust MS, Liu CL, Newman AM, Alizadeh AA. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods Mol Biol*. 2018;1711:243–259.
26. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol*. 2017;18(1):220. doi:10.1186/s13059-017-1349-1
27. Han Y, Wang Y, Dong X, et al. TISCH2: expanded datasets and new tools for single-cell transcriptome analyses of the tumor microenvironment. *Nucleic Acids Res*. 2023;51(D1):D1425–d31. doi:10.1093/nar/gkac959
28. Wang Q, Li S, Xu Y, et al. UBC9 overexpression promotes proliferation and metastasis in gastric cancer via ATF2. *World J Surg Oncol*. 2025;23(1):270. doi:10.1186/s12957-025-03922-y
29. Li F, Dai Y, Tang C, et al. Elevated UBC9 expression and its oncogenic role in colorectal cancer progression and chemoresistance. *Sci Rep*. 2025;15(1):9123. doi:10.1038/s41598-025-93868-8
30. Chen W, Wang Q, Li S. NPR3 promotes colorectal cancer cell proliferation, migration, invasion, and chemotherapy resistance. *Biochim Biophys Acta Gen Subj*. 2026;1870(2):130895. doi:10.1016/j.bbagen.2025.130895
31. Chuang P-K, Chang K-F, Chang C-H, et al. Comprehensive bioinformatics analysis of glycosylation-related genes and potential therapeutic targets in colorectal cancer. *Int J Mol Sci*. 2025;26(4):1648. doi:10.3390/ijms26041648
32. Chen X, Zhang D, Ou H, Su J, Wang Y, Zhou F. Bulk and single-cell RNA sequencing analyses coupled with multiple machine learning to develop a glycosyltransferase associated signature in colorectal cancer. *Transl Oncol*. 2024;49:102093. doi:10.1016/j.tranon.2024.102093
33. Li Y, Li D, Chen Y, et al. Robust glycogene-based prognostic signature for proficient mismatch repair colorectal adenocarcinoma. *Front Oncol*. 2021;11:727752. doi:10.3389/fonc.2021.727752
34. Mugisha S, Di X, Wen D, Zhao Y, Wu X, Zhang S, Jiang H. Upregulated GATA3/miR205-5p axis inhibits MFNG transcription and reduces the malignancy of triple-negative breast cancer. *Cancers*. 2022; 14.
35. Gao Y, Luo L, Qu Y, Zhou Q. MFNG is an independent prognostic marker for osteosarcoma. *Eur J Med Res*. 2023;28(1):256. doi:10.1186/s40001-023-01139-x
36. Zhang S, Chung W-C, Wu G, Egan SE, Miele L, Xu K. Manic fringe promotes a claudin-low breast cancer phenotype through notch-mediated PIK3CG induction. *Cancer Res*. 2015;75(10):1936–1943. doi:10.1158/0008-5472.CAN-14-3303

37. Nikolovska K, Spillmann D, Haier J, Ladányi A, Stock C, Seidler DG. Melanoma cell adhesion and migration is modulated by the uronyl 2-O sulfotransferase. *PLoS One*. 2017;12(1):e0170054. doi:10.1371/journal.pone.0170054
38. Lipiński P, Stępień KM, Ciara E, Tylki-Szymańska A, Jezela-Stanek A. Skeletal and bone mineral density features, genetic profile in congenital disorders of glycosylation: review. *Diagnostics*. 2021; 11.
39. Ondo K, Arakawa H, Nakano M, Fukami T, Nakajima M. SLC35B1 significantly contributes to the uptake of UDPGA into the endoplasmic reticulum for glucuronidation catalyzed by UDP-glucuronosyltransferases. *Biochem Pharmacol*. 2020;175:113916. doi:10.1016/j.bcp.2020.113916
40. Scott E, Hodgson K, Calle B, et al. Upregulation of GALNT7 in prostate cancer modifies O-glycosylation and promotes tumour growth. *Oncogene*. 2023;42(12):926–937. doi:10.1038/s41388-023-02604-x
41. Gaziel-Sovran A, Segura MF, Di Micco R, et al. miR-30b/30d regulation of GalNAc transferases enhances invasion and immunosuppression during metastasis. *Cancer Cell*. 2011;20(1):104–118. doi:10.1016/j.ccr.2011.05.027
42. Zhang L, Shay JW. Multiple Roles of APC and its therapeutic implications in colorectal cancer. *J Natl Cancer Inst*. 2017;109(8). doi:10.1093/jnci/djw332
43. Tian Y, Wang X, Cramer Z, et al. APC and P53 mutations synergise to create a therapeutic vulnerability to NOTUM inhibition in advanced colorectal cancer. *Gut*. 2023;72(12):2294–2306. doi:10.1136/gutjnl-2022-329140
44. Parker TW, Neufeld KL. APC controls Wnt-induced β-catenin destruction complex recruitment in human colonocytes. *Sci Rep*. 2020;10(1):2957. doi:10.1038/s41598-020-59899-z