

Article

Not peer-reviewed version

Comprehensive Bioinformatics Analysis of Glycosylation-Related Genes and Potential Therapeutic Targets in Colorectal Cancer

[Po-Kai Chuang](#), [Kai-Fu Chang](#), Chih-Hsuan Chang, Ting-Yu Chen, Yueh-Jung Wu, Hui-Ru Lin, [Chi-Jen Wu](#), [Cheng-Chun Wu](#), [Yu-Cheng Ho](#), Chih-Chun Lin, Chien-Han Yuan, [Chih-Yang Wang](#), [Yung-Kuo Lee](#)^{*}, [Tung-Yuan Chen](#)^{*}

Posted Date: 13 January 2025

doi: 10.20944/preprints202501.0941.v1

Keywords: colorectal cancer; glycosylation; prognostic model; machine learning; immune microenvironment; bioinformatics; drug sensitivity analysis



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Comprehensive Bioinformatics Analysis of Glycosylation-Related Genes and Potential Therapeutic Targets in Colorectal Cancer

Po-Kai Chuang ^{1,†}, Kai-Fu Chang ^{2,†}, Chih-Hsuan Chang ^{1,2}, Ting-Yu Chen ^{1,2}, Yueh-Jung Wu ³, Hui-Ru Lin ^{4,5}, Chi-Jen Wu ^{5,6}, Cheng-Chun Wu ⁷, Yu-Cheng Ho ⁷, Chih-Chun Lin ⁸, Chien-Han Yuan ^{2,4,9,10}, Chih-Yang Wang ^{11,12}, Yung-Kuo Lee ^{2,4,*} and Tung-Yuan Chen ^{3,*}

¹ Institute of Biomedical Sciences, National Sun Yat-Sen University, Kaohsiung 80424, Taiwan

² Medical Laboratory, Medical Education and Research Center, Kaohsiung Armed Forces General Hospital, Kaohsiung 80284, Taiwan

⁴ Division of Colorectal Surgery, Department of Surgery, Kaohsiung Armed Forces General Hospital, Kaohsiung 80284, Taiwan

³ Institute of Medical Science and Technology, National Sun Yat-Sen University, Kaohsiung 80424, Taiwan

⁵ Nursing Department, Kaohsiung Armed Forces General Hospital, Kaohsiung 80284, Taiwan

⁶ College of Nursing, Kaohsiung Medical University, Kaohsiung, 80708, Taiwan

⁷ School of Medicine, College of Medicine, I-Shou University, Kaohsiung 82445, Taiwan

⁸ Department of Physical Therapy, I-Shou University, Kaohsiung 824005, Taiwan

⁹ Department of Otolaryngology, Kaohsiung Armed Forces General Hospital, Kaohsiung 80284, Taiwan

¹⁰ Department of Otolaryngology, National Defense Medical Center, Taipei 11490, Taiwan

¹¹ PhD Program for Cancer Molecular Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan

¹² Graduate Institute of Cancer Biology and Drug Discovery, College of Medical Science and Technology, Taipei Medical University, Taipei 11031, Taiwan

* Correspondence: yungkuolee@gmail.com (Y.-K.-L.) sur045@yahoo.com.tw (T.-Y.C.); Tel.: +886-7-749-6751 (ext. 726201)

Abstract: Colorectal cancer (CRC) is a leading cause of cancer-related deaths worldwide, characterized by high incidence and poor survival rates. Glycosylation, a fundamental post-translational modification, influences protein stability, signaling, and tumor progression, with aberrations implicated in immune evasion and metastasis. This study investigates the role of glycosylation-related genes (Glycosylation-RGs) in CRC using machine learning and bioinformatics. Data from The Cancer Genome Atlas (TCGA) and the Molecular Signatures Database (MSigDB) were analyzed to identify 67 differentially expressed Glycosylation-RGs. These genes were used to classify CRC patients into two subgroups with distinct survival outcomes, highlighting their prognostic value. Weighted Gene Coexpression Network Analysis (WGCNA) revealed key modules associated with CRC traits, including pathways like glycan biosynthesis and PI3K-Akt signaling. A machine learning-based prognostic model demonstrated strong predictive performance, stratifying patients into high- and low-risk groups with significant survival differences. Additionally, the model revealed correlations between risk scores and immune cell infiltration, providing insights into the tumor immune microenvironment. Drug sensitivity analysis identified potential therapeutic agents, including Trametinib, SCH772984, and Oxaliplatin, showing differential efficacy between risk groups. These findings enhance our understanding of glycosylation in CRC, identifying it as a critical factor in disease progression and a promising target for future therapeutic strategies.

Keywords: colorectal cancer; glycosylation; prognostic model; machine learning; immune microenvironment; bioinformatics; drug sensitivity analysis

1. Introduction

Colorectal cancer (CRC) is one of the leading causes of cancer-related mortality worldwide, characterized by high incidence and poor survival in advanced stages [1–3]. Despite significant advances in therapeutic strategies, the prognosis for CRC patients remains suboptimal, largely due to tumor heterogeneity, delayed detection, and the limited sensitivity of current diagnostic tools [4,5]. These challenges complicate early intervention and necessitate the development of more precise diagnostic approaches. Understanding these mechanisms is essential for improving early detection and developing personalized therapeutic approaches [5–7].

Glycosylation is a fundamental post-translational modification that plays a critical role in protein folding, stability, and signaling [8,9]. Aberrant glycosylation has been implicated in various biological processes, including tumor progression, immune evasion, and metastasis [10,11]. It affects cell-cell communication, alters the tumor microenvironment, and enables cancer cells to evade immune surveillance [10,12,13]. Recent studies have highlighted its potential as a biomarker for cancer prognosis and a target for therapeutic interventions [14,15]. However, the specific contributions of glycosylation-related genes (Glycosylation-RGs) to CRC remain inadequately explored.

The dynamic interplay between glycosylation and oncogenic pathways has garnered attention, particularly in its role in modulating epithelial-mesenchymal transition (EMT), angiogenesis, and drug resistance [16,17]. Furthermore, glycosylation patterns are known to vary across different stages of cancer progression, presenting an opportunity to refine staging and risk stratification for CRC patients [18,19].

Advancements in high-throughput sequencing technologies and bioinformatics have provided an unprecedented opportunity to investigate the molecular landscape of CRC [20–22]. Computational methods, including machine learning algorithms, enable the integration of multi-omics data to uncover novel biomarkers and predict patient outcomes with high accuracy [23,24]. Specifically, machine learning models can analyze complex datasets to predict patient-specific responses to therapies, identify early-stage biomarkers with improved sensitivity, and stratify patients into risk groups to guide personalized treatment strategies [21,25,26]. Integrating gene expression data with clinical outcomes can reveal critical pathways and identify prognostic biomarkers [27,28]. This study focuses on the systematic analysis of glycosylation-related genes in CRC, aiming to elucidate their role in tumor progression and the tumor immune microenvironment [19,29–31].

Here, we conducted a comprehensive analysis using publicly available datasets to identify glycosylation-related genes associated with CRC. We explored their differential expression, constructed a prognostic risk model, and examined their correlation with immune cell infiltration. Our findings provide valuable insights into the functional and clinical significance of glycosylation-related genes in CRC, paving the way for potential therapeutic strategies.

2. Results

2.1. Identification of Differentially Expressed Overlapping Glycosylation-Related Genes

To identify glycosylation-related genes involved in CRC, we analyzed TCGA datasets and identified 1,661 differentially expressed genes (DEGs). Using the MSigDB database, 663 glycosylation-related genes were retrieved. Among these, 67 genes were identified as differentially expressed and overlapping with glycosylation-related genes. The volcano plot shows the distribution of DEGs, with significant upregulation and downregulation patterns (Figure 1A). The Venn diagram highlights the intersection between glycosylation-related genes and DEGs, underscoring the relevance of these genes in CRC (Figure 1B).

Functional enrichment analysis of the 67 overlapping genes revealed significant involvement in biological pathways, such as glycan biosynthesis, N-glycan modification, and cell adhesion processes. These pathways are known to contribute to tumor progression and metastasis, suggesting their critical role in CRC.

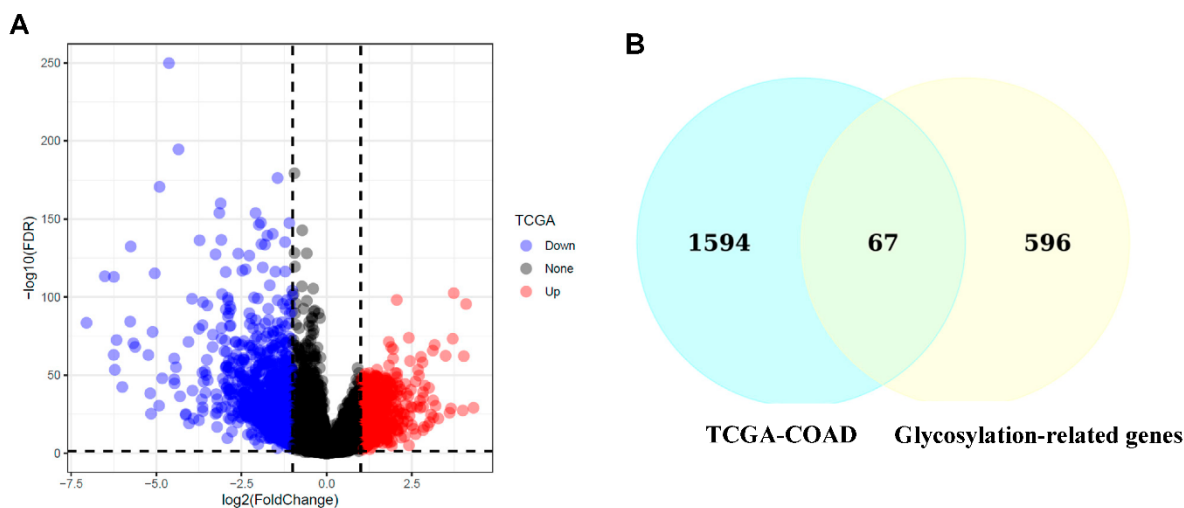


Figure 1. Identification of differentially expressed glycosylation-related genes in CRC. (A) Volcano plot displaying the distribution of differentially expressed genes in the TCGA-COAD dataset. Blue dots indicate downregulated genes, red dots indicate upregulated genes, and black dots represent non-significant genes. The x-axis shows log₂ (Fold Change), while the y-axis shows -log₁₀ (FDR). (B) Venn diagram illustrating the overlap between glycosylation-related genes and differentially expressed genes in TCGA-COAD. The intersection identifies 67 overlapping genes critical for further analysis.

2.2. Identification of Glycosylation-Related Subgroups

Unsupervised consensus clustering was performed using the expression profiles of the 67 glycosylation-related genes, classifying CRC patients into two distinct subgroups, C1 and C2. The optimal number of clusters (k = 2) was determined using the consensus cumulative distribution function (CDF) (Figure 2A). Principal Component Analysis (PCA) confirmed the clear separation between the two subgroups, indicating distinct molecular characteristics (Figure 2B). Kaplan-Meier survival analysis revealed that patients in subgroup C2 exhibited significantly poorer survival outcomes compared to those in subgroup C1 (p < 0.001, Figure 2C). The heatmap of gene expression further illustrated distinct glycosylation-related gene expression patterns between the two subgroups, correlating with clinical traits such as tumor stage and metastasis (Figure 2D).

The subgroup-specific expression of glycosylation-related genes suggests potential molecular mechanisms driving CRC progression. For example, subgroup C2 showed higher expression of genes associated with immune evasion and angiogenesis, which may contribute to the poor prognosis observed in this cohort.

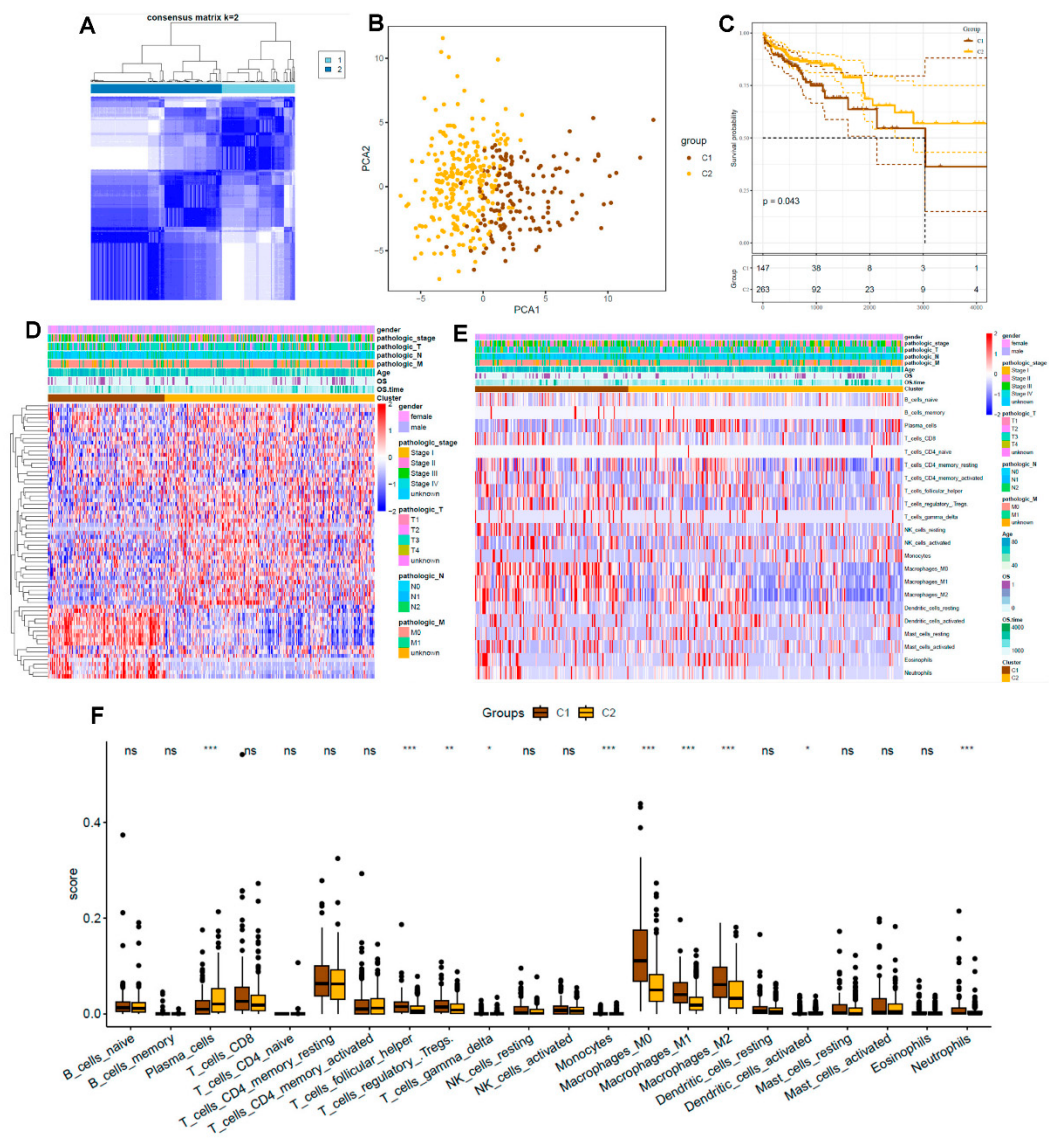


Figure 2. Identification of glycosylation-related subgroups in CRC. (A) Consensus matrix heatmap demonstrating optimal clustering into two subgroups (C1 and C2) using glycosylation-related genes. (B) Principal Component Analysis (PCA) showing the separation between subgroups C1 and C2. (C) Kaplan-Meier survival curves illustrating overall survival differences between the subgroups, with subgroup C2 showing poorer survival outcomes (p = 0.043). (D) Heatmap displaying glycosylation-related gene expression in subgroups C1 and C2, alongside clinicopathological characteristics. (E) Heatmap depicting immune cell infiltration patterns in the two subgroups. (F) Boxplots showing significant differences in the abundance of specific immune cell types between subgroups C1 and C2 (e.g., macrophages M0, dendritic cells, and activated NK cells, p < 0.05).

2.3. Identification of Highly Correlated Gene Modules in CRC

Using WGCNA, we constructed a gene coexpression network to identify modules associated with CRC traits. The turquoise module exhibited the strongest correlation with tumor traits and included several glycosylation-related genes (Figure 3A-D). Functional enrichment analysis of the turquoise module genes highlighted pathways such as cGMP-PKG signaling, cytoskeleton in muscle cells, extracellular matrix organization, and PI3K-Akt signaling (Figure 3E). Gene significance and module membership analysis revealed that key glycosylation-related genes. These genes are known to influence tumor invasiveness and immune suppression [32].

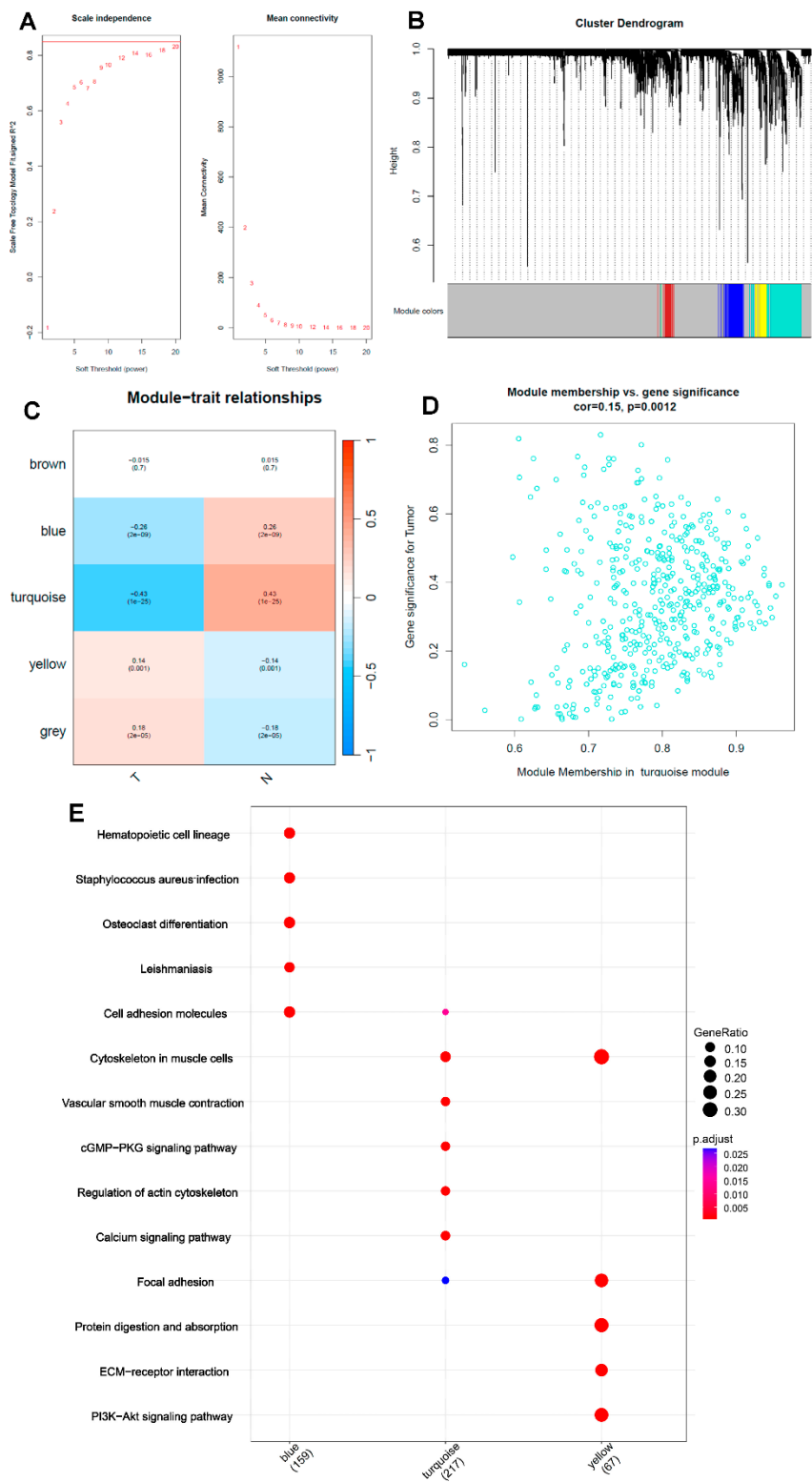


Figure 3. Identification of highly correlated gene modules using Weighted Gene Coexpression Network Analysis (WGCNA). (A) Soft thresholding power determination for constructing a scale-free network, showing the scale-free topology fit index and mean connectivity for different power values. (B) Cluster dendrogram of glycosylation-related genes grouped into distinct modules, with colors representing module membership. (C) Module-trait relationships illustrating correlations between modules (e.g., turquoise, blue, brown) and clinical traits, such as tumor presence. (D) Scatter plot showing the correlation between module membership and gene significance in the turquoise module, which is highly associated with CRC. (E) KEGG pathway enrichment analysis for genes in the turquoise module, highlighting pathways such as "Focal adhesion," "PI3K-Akt signaling" and "ECM-receptor interaction."

2.4. Identification of Glycosylation-Related Gene Clusters

Consensus clustering was extended to identify four distinct gene clusters (C1-C4) using glycosylation-related genes (Figure 4A). PCA showed clear separation of these clusters (Figure 4B), while Kaplan-Meier analysis revealed significant survival differences, with cluster C4 associated with the poorest prognosis (Figure 4C). The heatmap illustrated the differential expression of glycosylation-related genes across clusters, along with clinicopathological annotations (Figure 4D).

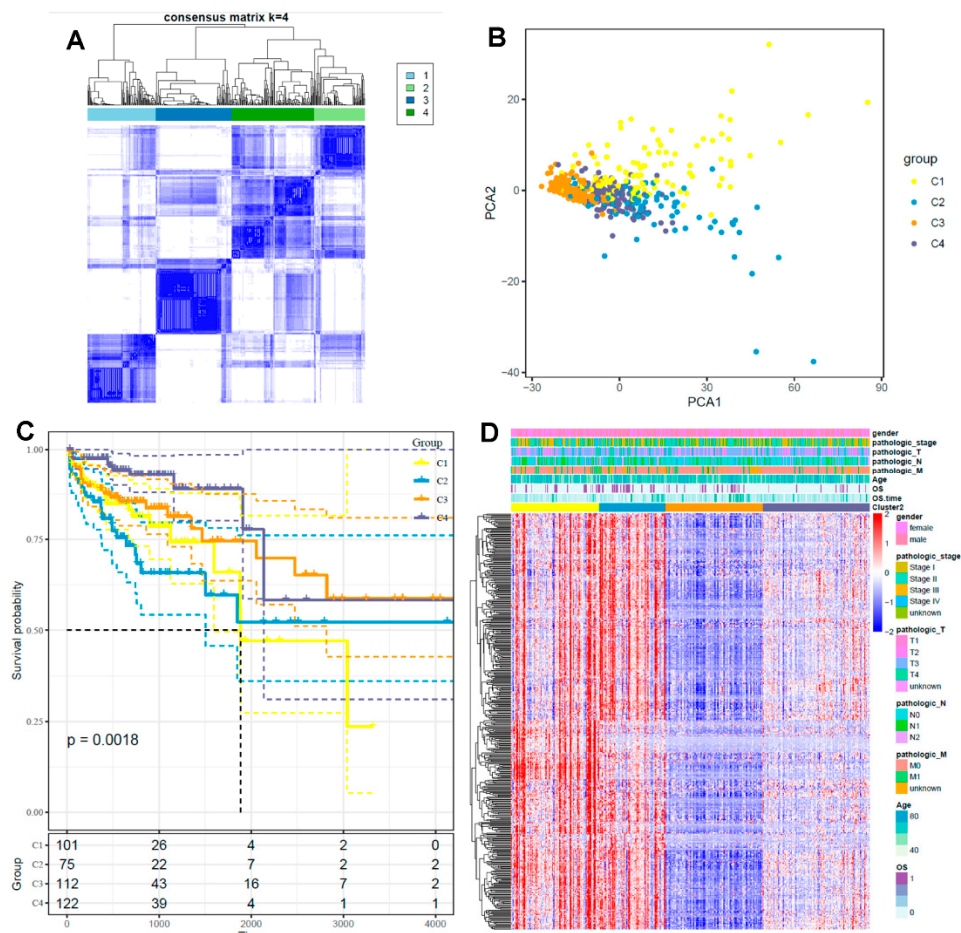


Figure 4. Identification of glycosylation-related gene clusters in CRC. (A) Consensus matrix heatmap defining four gene clusters (C1, C2, C3, and C4) based on glycosylation-related gene expression. (B) Principal Component Analysis (PCA) visualizing the separation of samples into four clusters. (C) Kaplan-Meier survival curves showing significant differences in overall survival among the four clusters ($p = 0.0018$), with cluster C4 demonstrating the poorest survival. (D) Heatmap depicting the expression patterns of glycosylation-related genes across the four clusters, alongside associated clinicopathological characteristics.

2.5. Construction of a Glycosylation-Related Prognostic Risk Model

LASSO regression and Cox proportional hazards analysis were employed to construct a prognostic risk model based on glycosylation-related genes. Four genes (TUB, TCF7L1, MPP2, and TMEM59L) were selected for the final model, which stratified patients into high- and low-risk groups (Figure 5A-B). Kaplan-Meier survival analysis demonstrated that high-risk patients had significantly worse survival outcomes ($p < 0.00013$, Figure 5C). Patients in the high-risk group exhibited poorer survival outcomes compared to those in the low-risk group. Furthermore, the distribution of risk scores and survival outcomes (Figure 5D). The top panel displays the stratification of patients by risk score, while the middle panel illustrates survival status, where green dots denote living patients and orange triangles represent deceased individuals. The heatmap in the bottom panel highlights the

expression profiles of the selected prognostic genes across the high- and low-risk groups, demonstrating distinct expression patterns correlating with risk stratification.

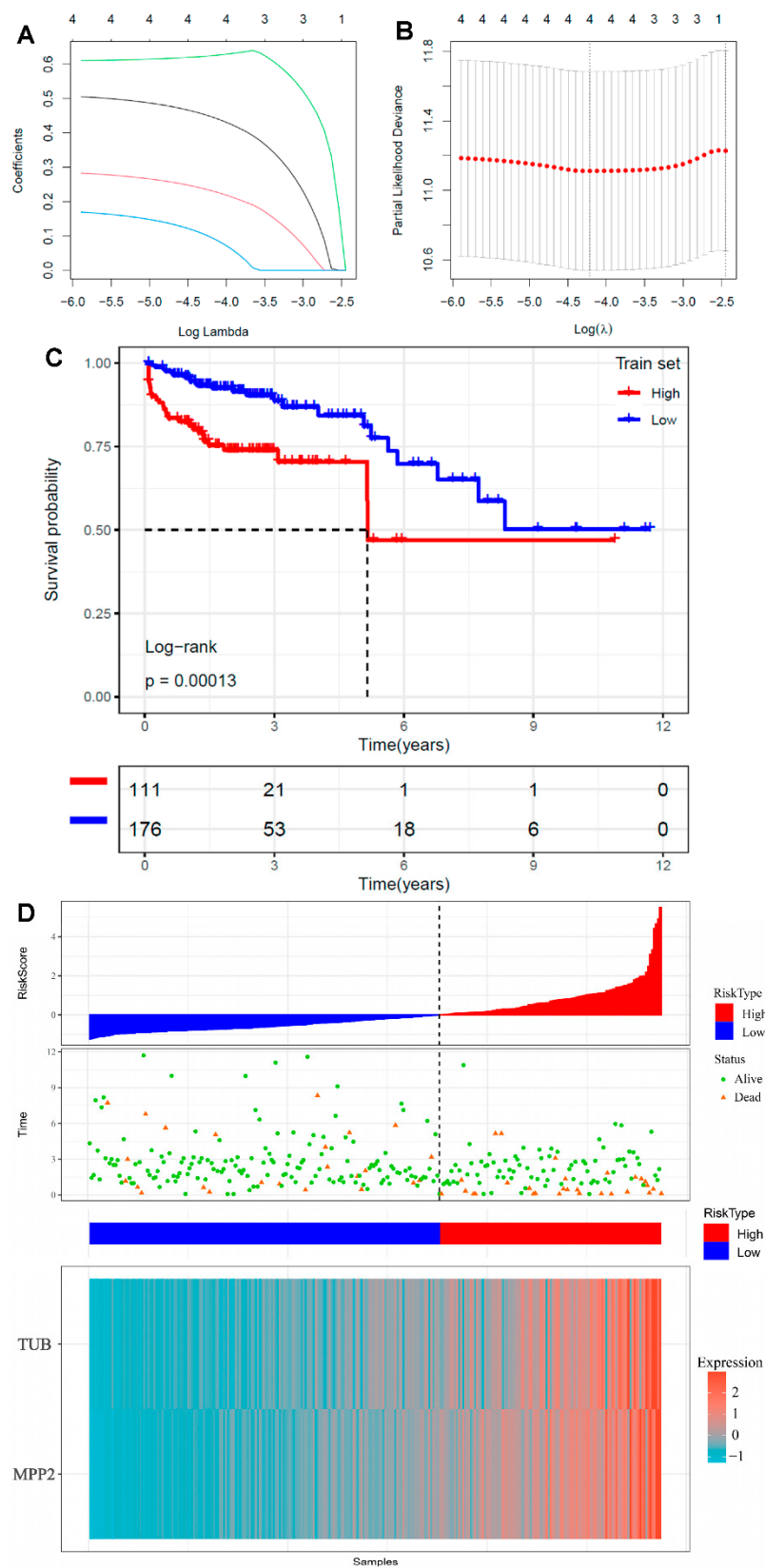


Figure 5. Construction of a glycosylation-related prognostic risk model for CRC. (A) Coefficient profiles of selected genes during LASSO regression. Different colored lines represent individual genes. (B) Partial likelihood deviance plot to determine the optimal lambda value for the model, with the minimum value marked

by a dotted line. (C) Kaplan-Meier survival curves showing a significant difference in overall survival between high-risk (red) and low-risk (blue) groups in the training dataset ($p = 0.00013$). (D) Distribution of risk scores, survival status, and gene expression profiles in the high- and low-risk groups. The top panel shows risk scores, the middle panel displays survival outcomes (green dots: alive; orange triangles: deceased), and the bottom heatmap illustrates the expression patterns of selected prognostic genes.

2.6. Validation of the Glycosylation-Related Prognostic Risk Model

The glycosylation-related prognostic risk model for CRC was validated using time-dependent ROC curves, as shown in Figure 6A. The model demonstrated good predictive accuracy for overall survival, with AUC values of 0.7, 0.7, and 0.74 for 1-year, 3-year, and 5-year survival, respectively. Figure 6B presents boxplots comparing the expression levels of the key prognostic genes, TUB and MPP2, between high- and low-risk groups. Both genes were significantly overexpressed in the high-risk group ($***p < 0.001$), supporting their relevance to the risk stratification. Kaplan-Meier survival analysis for individual genes (Figure 6C) revealed that high expression levels of TUB and MPP2 were significantly associated with poorer overall survival (log-rank $p = 0.019$ and $p = 0.00094$, respectively). These findings further confirm the robustness and prognostic value of the model in predicting CRC outcomes.

Figure 6. Validation of the glycosylation-related prognostic risk model for CRC. (A) Time-dependent ROC curves for predicting overall survival at 1, 3, and 5 years, showing AUC values of 0.7, 0.7, and 0.74, respectively. (B) Boxplots comparing the expression levels of the key prognostic genes (TUB and MPP2) between high- and low-risk groups, indicating significant differences ($***p < 0.001$). (C) Kaplan-Meier survival curves for individual genes TUB (left) and MPP2 (right), with high expression associated with poorer overall survival ($p = 0.019$ and $p = 0.00094$, respectively).

2.7. Correlation Between Risk Scores and Immune Cell Infiltration

CIBERSORT analysis revealed significant correlations between glycosylation-related risk scores and immune cell infiltration. High-risk patients exhibited increased infiltration of suppressive immune cells, such as regulatory T cells and M2 macrophages, alongside reduced activation of cytotoxic T cells and NK cells. These findings suggest that aberrant glycosylation may contribute to an immunosuppressive tumor microenvironment (Figure 7A-K).

Heatmap analysis further illustrated the relationship between key glycosylation-related genes and immune cell infiltration, highlighting their potential role in modulating immune escape mechanisms in CRC. These findings provide a comprehensive view of the molecular and immunological landscape shaped by glycosylation in CRC progression.

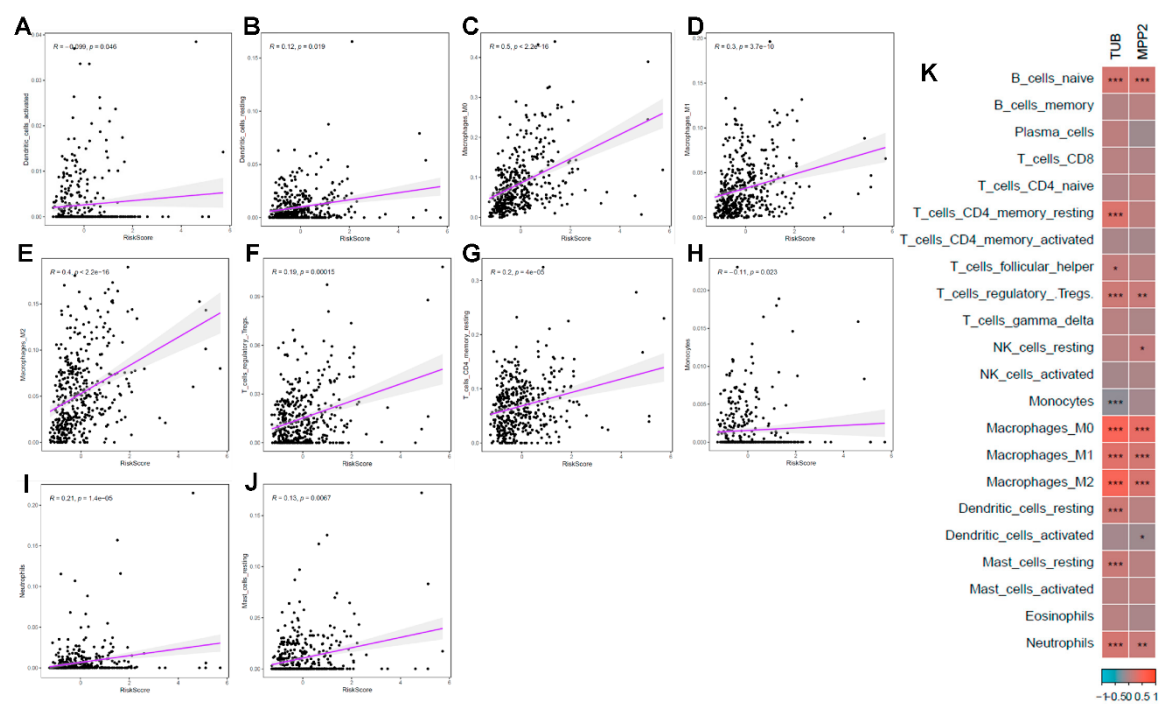


Figure 7. Correlation between immune cell infiltration and glycosylation-related risk scores in CRC. (A-J) Scatter plots showing significant correlations between the glycosylation-related risk score and the infiltration levels of various immune cell types, including activated (A) dendritic cells, (B) resting dendritic cells, (C) macrophages M0, (D) macrophages M1, (E) macrophages M2, (F) regulatory T cells, (G) resting memory CD4+ T cells, (H) monocytes, (I) neutrophils, and (J) resting mast cells. Each purple line indicates the linear regression trend, with shaded areas representing the 95% confidence interval. (K) Heatmap illustrating the relationship between the expression of key prognostic genes (TUB and MPP2) and immune cell infiltration. Red indicates positive correlation, and blue indicates negative correlation.

2.8. Drug Sensitivity Prediction in High- and Low-Risk Groups

To investigate potential therapeutic strategies, drug sensitivity analysis was conducted using the oncoPredict algorithm. The analysis identified the top five drugs with the largest and smallest differences in predicted sensitivity between high-risk and low-risk groups.

The low-risk group exhibited significantly higher sensitivity to drugs targeting the MAPK pathway, including Trametinib (Figure 8A) and SCH772984 (Figure 8B), as well as the chemotherapy agent Oxaliplatin (Figure 8C). Acetalax, a BCL-2 inhibitor, and VX-11e, an ERK inhibitor, also showed significantly enhanced efficacy in the low-risk group (Figures 8D and 8E). These findings highlight the potential for these agents to provide improved therapeutic outcomes in patients with favorable risk profiles.

In contrast, the drugs AZD8055, Doxorubicin, Axitinib, NU7441, and ZM447439 demonstrated lower sensitivity in predicted sensitivity between high- and low-risk groups (Figures 8F-J). These results suggest that these agents may not exhibit significant stratified efficacy in CRC treatment based on the identified glycosylation-related risk model. These findings provide a basis for further exploring the translational application of these drugs in stratified CRC treatment, enabling personalized therapeutic strategies to improve patient outcomes.

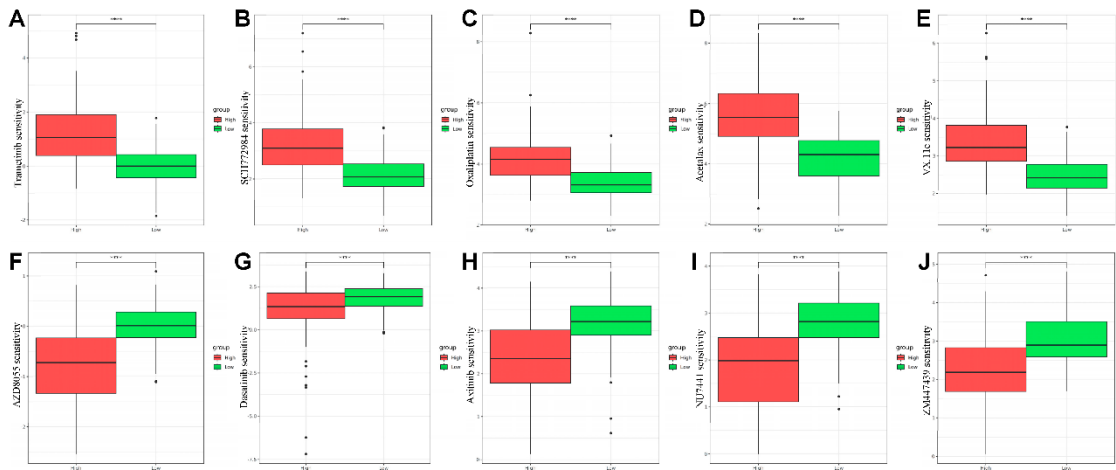


Figure 8. Drug sensitivity analysis (IC50) of high- and low-risk groups predicted by oncoPredict. Box plots display the top five drugs with the largest differences in predicted sensitivity between high-risk (red) and low-risk (green) groups and the top five drugs with the smallest differences. (A) Trametinib, (B) SCH772984, (C) Oxaliplatin, (D) Acetalax, and (E) VX-11e are the top five drugs with significantly higher sensitivity in the low-risk group. Conversely, (F) AZD8055, (G) Doxorubicin, (H) Axitinib, (I) NU7441, and (J) ZM447439 showed lower differences in sensitivity between the groups.

3. Discussion

This study systematically analyzed glycosylation-related genes in colorectal cancer (CRC) and established their potential role in tumor progression, prognosis, and the tumor immune microenvironment. Differential expression analysis revealed that 67 glycosylation-related genes were significantly dysregulated in CRC, suggesting their functional involvement in oncogenic pathways (Figure 1A-B) [33,34].

The identification of CRC subgroups based on glycosylation-related gene expression profiles highlighted distinct clinical and molecular characteristics. Subgroup C2 demonstrated significantly worse survival outcomes compared to C1, emphasizing the prognostic value of glycosylation-related genes in CRC (Figure 2A-D). These findings align with previous reports implicating glycosylation in tumor heterogeneity and aggressiveness [11,35]. Notably, these subgroups provide a framework for personalized treatment strategies, underscoring the potential of molecular classification in optimizing patient management [36].

Using WGCNA, the turquoise module was identified as a key coexpression network associated with CRC traits. Functional enrichment analysis highlighted the role of this module in critical pathways such as glycan biosynthesis, focal adhesion, and PI3K-Akt signaling, supporting its biological relevance in CRC progression (Figure 3A-G). The strong correlation between these pathways and tumor invasiveness, metastasis, and immune evasion further validates the significance of glycosylation in the broader context of cancer biology [37].

The construction of a prognostic risk model based on glycosylation-related genes represents a significant advance in CRC management. High-risk patients identified by the model exhibited markedly poorer survival, and the model demonstrated robust predictive performance across multiple datasets (Figure 5A-D, Figure 6A-C). These results underscore the clinical utility of glycosylation-related genes as prognostic biomarkers. Additionally, the machine learning approach used to build this model exemplifies how computational tools can optimize biomarker selection and stratification processes, ensuring higher accuracy and reproducibility [38,39].

Furthermore, this study revealed significant associations between glycosylation-related risk scores and immune cell infiltration, including macrophages, regulatory T cells, and neutrophils. These findings suggest that glycosylation may influence the tumor immune microenvironment, offering new insights into potential immunotherapeutic strategies (Figure 7A-K). Specifically,

glycosylation patterns could affect the expression of immune checkpoints or the recruitment of suppressive immune cells, which are critical in shaping the tumor microenvironment [40].

The translational potential of targeting glycosylation pathways in CRC treatment is immense. Aberrant glycosylation could serve as both a diagnostic marker and a therapeutic target. For instance, glycosylation inhibitors or monoclonal antibodies targeting glycan-modified proteins might enhance the efficacy of existing therapies, including immune checkpoint inhibitors and chemotherapy [41]. In addition, the interplay between glycosylation and immune cell recruitment presents opportunities to modulate the tumor immune microenvironment, potentially enhancing anti-tumor immunity. The integration of glycosylation-targeted therapies with existing treatment regimens could lead to synergistic effects, improving patient outcomes [42].

Machine learning played a pivotal role in this study, enabling the integration of high-dimensional data and the identification of robust biomarkers. Its application to risk stratification and survival prediction highlights its value in precision oncology. Future research should incorporate additional omics layers, such as metabolomics and proteomics, to develop more comprehensive models. Integrating these models with clinical data could further refine prognostic predictions and guide individualized therapeutic strategies. Moreover, exploring deep learning techniques could further enhance the predictive capabilities of models used for biomarker discovery [43,44].

Despite its contributions, this study has limitations. The reliance on publicly available datasets may introduce biases related to cohort selection and data quality. Experimental validation of identified biomarkers in larger, more diverse cohorts is necessary to confirm their clinical utility [45]. Additionally, while the prognostic model showed high predictive accuracy, external validation in prospective studies is required to establish its generalizability. A further limitation is the lack of mechanistic studies to directly validate the functional role of glycosylation in CRC progression and immune modulation, which should be a focus for future research [46].

The integration of drug sensitivity analysis represents a significant expansion of this study, providing novel insights into potential therapeutic strategies for CRC patients stratified by glycosylation-related risk groups. Using oncoPredict, we identified five drugs (Trametinib, SCH772984, Oxaliplatin, Acetalax, and VX-11e) with significant differential sensitivity between high- and low-risk groups (Figures 8A-E). These findings underscore the translational potential of tailoring therapeutic approaches to molecularly defined subgroups. The higher sensitivity of the low-risk group to Trametinib and SCH772984 highlights the relevance of targeting the MAPK signaling pathway, which is frequently dysregulated in CRC. This pathway, activated downstream of cGMP-dependent protein kinases (PKG) in response to cGMP signaling (Figures 3E), has been implicated in increasing tumor cell stemness and metastasis [47]. The cGMP-PKG-MAPK cascade underscores a critical molecular mechanism by which tumor progression and therapy resistance may be mediated, providing further rationale for therapeutic intervention in this pathway.

Additionally, the increased efficacy of Oxaliplatin in the low-risk group suggests that stratification based on glycosylation-related risk scores may refine patient selection for standard chemotherapy regimens. Acetalax, a BCL-2 inhibitor, demonstrated enhanced sensitivity in the low-risk group, pointing to the potential of apoptosis-targeting therapies in this cohort. Similarly, VX-11e further supports the utility of targeting the ERK pathway. In contrast, drugs such as AZD8055, Doxorubicin, Axitinib, NU7441, and ZM447439 showed lower differences in predicted sensitivity between high- and low-risk groups (Figures 8F-J). This suggests that these agents may not exhibit substantial stratified efficacy in CRC treatment based on glycosylation-related risk models. The identification of these drugs also underscores the specificity of certain therapeutic agents to molecularly distinct subgroups.

These drug sensitivity predictions extend the utility of the glycosylation-related prognostic model beyond survival forecasting to therapeutic guidance. By integrating molecular stratification with drug efficacy predictions, this study provides a framework for precision oncology. Future experimental validation of these predictions and exploration of combination therapies targeting glycosylation-related pathways and tumor-specific vulnerabilities are warranted.

In conclusion, our findings highlight the critical role of glycosylation in CRC progression and its potential as a biomarker and therapeutic target. By leveraging machine learning and advanced bioinformatics, this study provides a robust framework for understanding glycosylation's contributions to CRC and opens avenues for innovative therapeutic interventions. Further experimental and clinical investigations are warranted to fully realize the potential of glycosylation-related therapies in CRC management.

4. Materials and Methods

4.1. Data Collection and Preprocessing

Gene expression profiles and corresponding clinical data for colorectal cancer (CRC) were downloaded from The Cancer Genome Atlas (TCGA) database. The dataset included RNA-seq data normalized as fragments per kilobase of transcript per million mapped reads (FPKM) values. Clinical information included patient demographics, tumor stage, and survival outcomes.

4.2. Identification of Glycosylation-Related Genes (Glycosylation-RGs)

To investigate the role of glycosylation-related genes in CRC, a comprehensive list of glycosylation-related genes was obtained by searching the Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/human/>) using the keyword "Glycosylation." A total of 663 glycosylation-related genes were retrieved for further analysis.

4.3. Differential Gene Expression Analysis

Differentially expressed genes (DEGs) between CRC and adjacent normal tissue samples were identified using the "limma" R package. Genes with $|\log_2 \text{fold change}| > 1$ and adjusted p-value < 0.05 were considered significantly differentially expressed. Overlapping genes between DEGs and the 663 glycosylation-related genes were identified using Venn diagram analysis.

4.4. Consensus Clustering

Unsupervised consensus clustering was performed using the "ConsensusClusterPlus" R package to classify CRC samples based on the expression profiles of overlapping glycosylation-related genes. The optimal number of clusters was determined using the consensus cumulative distribution function (CDF).

4.5. Weighted Gene Coexpression Network Analysis (WGCNA)

WGCNA was conducted to construct coexpression networks and identify gene modules associated with CRC traits. A soft thresholding power was selected to achieve scale-free topology, and modules were correlated with clinical traits to identify the most relevant module for downstream analysis.

4.6. Prognostic Model Construction and Validation

LASSO regression and multivariate Cox proportional hazards analysis were used to develop a glycosylation-related prognostic risk model. The risk score for each patient was calculated as a weighted sum of the expression levels of selected genes. Patients were divided into high- and low-risk groups based on the median risk score. The model was validated using Kaplan-Meier survival analysis and time-dependent receiver operating characteristic (ROC) curves.

4.7. Immune Infiltration Analysis

The relative abundance of immune cell types in CRC samples was estimated using the CIBERSORT algorithm. Correlation analysis was performed to evaluate the relationship between

glycosylation-related risk scores and immune cell infiltration levels. A heatmap was generated to visualize the association between key prognostic genes and immune cell infiltration.

4.8. Functional Enrichment Analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed for genes in the turquoise module identified by WGCNA. Enrichment analyses were conducted using the "clusterProfiler" R package, and pathways with adjusted p-values < 0.05 were considered significantly enriched.

4.9. Drug Sensitivity Analysis

Drug sensitivity predictions were performed using the oncoPredict R package, an extension of the earlier pRRophetic tool developed by Paul Geeleher and colleagues at the University of Minnesota. OncoPredict predicts in vivo or cancer patient drug response based on gene expression data and IC50 values derived from cell line screening data. This package builds on the oncoPredict algorithm [48], which was introduced in 2014 and employs ridge regression to model relationships between baseline gene expression and in vitro drug sensitivity in cell lines [49].

4.10. Statistical Analysis

All statistical analyses were performed in R software (version 4.4.1) and a p-value < 0.05 was considered statistically significant.

5. Conclusions

This study systematically highlights the significance of glycosylation-related genes in colorectal cancer (CRC) progression and prognosis. By integrating bioinformatics and machine learning approaches, we identified key glycosylation-related genes and constructed a robust prognostic risk model capable of stratifying patients into high- and low-risk groups. These findings not only deepen our understanding of the molecular mechanisms underlying CRC but also underscore the pivotal role of glycosylation in modulating the tumor immune microenvironment.

The addition of drug sensitivity analysis further expands the translational potential of this work. By leveraging oncoPredict, we identified five key drugs (Trametinib, SCH772984, Oxaliplatin, Acetalax, and VX-11e) with significant differences in predicted efficacy between high- and low-risk groups. These findings suggest that the identified risk model can serve not only as a prognostic tool but also as a guide for personalized therapeutic strategies. The low-risk group demonstrated higher sensitivity to targeted agents such as Trametinib and SCH772984, emphasizing the relevance of MAPK pathway inhibitors, while also exhibiting enhanced efficacy with standard chemotherapy agents like Oxaliplatin. These insights pave the way for integrating molecular stratification with personalized drug selection to optimize CRC treatment outcomes.

Furthermore, the study identified five additional drugs (AZD8055, Doxorubicin, Axitinib, NU-7441, and ZM447439) that showed lower sensitivity differences between risk groups. While these agents may not have stratified efficacy based on the current risk model, their identification provides additional context for exploring broader therapeutic applications in CRC treatment. The combined insights from prognostic modeling and drug sensitivity analysis highlight the value of integrating molecular biomarkers with therapeutic strategies to refine treatment paradigms.

The study's results suggest that targeting aberrant glycosylation could hold promise for improving therapeutic strategies, including enhancing the efficacy of immune checkpoint inhibitors and chemotherapy. Furthermore, the integration of glycosylation-related biomarkers into clinical workflows may refine CRC staging, improve early detection, and inform tailored treatment regimens. Future studies should focus on validating these findings in diverse cohorts and experimental settings to fully realize the clinical potential of glycosylation-related therapies in CRC.

Author Contributions: Conceptualization, P.-K. C., K.-F. C., Y.-K. L., and T.-Y. C.; methodology, P.-K. C., K.-F. C., C.-Y. W., and Y.-K. L.; software, K.-F. C., C.-Y. W., and Y.-K. L.; formal analysis, C.-H. C., T.-Y. C., Y.-J. W., H.-R. L., and C.-J. W.; investigation, C.-H. Y., and C.-Y. W.; data curation, C.-C. W., Y.-C. H., and C.-C., L.; writing—original draft preparation, P.-K. C., K.-F. C., and Y.-K. L.; writing—review and editing, P.-K. C., K.-F. C., Y.-K. L., and T.-Y. C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Science and Technology Council (NSTC) of Taiwan, grant number 113-2320-B-393-001 and by Kaohsiung Armed Forces General Hospital grant number, KAFGH_D_113019, KAFGH_D_114022, and KAFGH_D_114053. The APC was funded by Kaohsiung Armed Forces General Hospital.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data involved in this study are available from the corresponding author on request. Molecular Signatures Database (MSigDB, <https://www.gsea-msigdb.org/gsea/msigdb/human/>) using the keyword "Glycosylation for analysis in this study."

Acknowledgments: We would like to acknowledge Yi-Ting Wu, Chien-Cheng Chao, Yun- Yu Lin, Kai-Fu Chang and Yueh-Yuan Shieh for their excellent technical support at Laboratory of Research and Medical Education and Research Center, Kaohsiung Armed Forces General Hospital.

Conflicts of Interest: Not applicable.

Abbreviations

The following abbreviations are used in this manuscript:

CRC	Colorectal cancer
TCGA	The Cancer Genome Atlas
FPKM	Fragments per kilobase of transcript per million mapped
MSigDB	Molecular Signatures Database
WGCNA	Weighted Gene Coexpression Network Analysis
CDF	consensus cumulative distribution function
ROC	receiver operating characteristic

References

1. Xi, Y.; Xu, P. Global colorectal cancer burden in 2020 and projections to 2040. *Translational oncology* **2021**, *14*, 101174.
2. Arnold, M.; Sierra, M.S.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **2017**, *66*, 683-691.
3. Hossain, M.S.; Karuniawati, H.; Jairoun, A.A.; Urbi, Z.; Ooi, D.J.; John, A.; Lim, Y.C.; Kibria, K.K.; Mohiuddin, A.; Ming, L.C. Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. *Cancers* **2022**, *14*, 1732.
4. Andrei, P.; Battuello, P.; Grasso, G.; Rovera, E.; Tesio, N.; Bardelli, A. Integrated approaches for precision oncology in colorectal cancer: The more you know, the better. In *Proceedings of the Seminars in cancer biology*, 2022; pp. 199-213.
5. Beniwal, S.S.; Lamo, P.; Kaushik, A.; Lorenzo-Villegas, D.L.; Liu, Y.; MohanaSundaram, A. Current status and emerging trends in colorectal cancer screening and diagnostics. *Biosensors* **2023**, *13*, 926.
6. Das, V.; Kalita, J.; Pal, M. Predictive and prognostic biomarkers in colorectal cancer: A systematic review of recent advances and challenges. *Biomedicine & Pharmacotherapy* **2017**, *87*, 8-19.
7. Dey, A.; Mitra, A.; Pathak, S.; Prasad, S.; Zhang, A.S.; Zhang, H.; Sun, X.-F.; Banerjee, A. Recent advancements, limitations, and future perspectives of the use of personalized medicine in treatment of colon cancer. *Technology in Cancer Research & Treatment* **2023**, *22*, 15330338231178403.
8. Jayaprakash, N.G.; Surolia, A. Role of glycosylation in nucleating protein folding and stability. *Biochemical Journal* **2017**, *474*, 2333-2347.

9. Hansson, K.; Stenflo, J. Post-translational modifications in proteins involved in blood coagulation. *Journal of thrombosis and haemostasis* **2005**, *3*, 2633-2648.
10. Rodrigues, J.G.; Balmaña, M.; Macedo, J.A.; Poças, J.; Fernandes, Â.; de-Freitas-Junior, J.C.M.; Pinho, S.S.; Gomes, J.; Magalhães, A.; Gomes, C. Glycosylation in cancer: Selected roles in tumour progression, immune modulation and metastasis. *Cellular immunology* **2018**, *333*, 46-57.
11. Pinho, S.S.; Reis, C.A. Glycosylation in cancer: mechanisms and clinical implications. *Nature Reviews Cancer* **2015**, *15*, 540-555.
12. Läubli, H.; Borsig, L. Altered cell adhesion and glycosylation promote cancer immune suppression and metastasis. *Frontiers in immunology* **2019**, *10*, 2120.
13. Peixoto, A.; Relvas-Santos, M.; Azevedo, R.; Santos, L.L.; Ferreira, J.A. Protein glycosylation and tumor microenvironment alterations driving cancer hallmarks. *Frontiers in oncology* **2019**, *9*, 380.
14. Thomas, D.; Rathinavel, A.K.; Radhakrishnan, P. Altered glycosylation in cancer: A promising target for biomarkers and therapeutics. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* **2021**, *1875*, 188464.
15. Mereiter, S.; Balmaña, M.; Campos, D.; Gomes, J.; Reis, C.A. Glycosylation in the era of cancer-targeted therapy: where are we heading? *Cancer cell* **2019**, *36*, 6-16.
16. Fonseca, L.M.d.; Silva, V.A.d.; Freire-de-Lima, L.; Previato, J.O.; Mendonca-Previato, L.; Capella, M.A.M. Glycosylation in cancer: interplay between multidrug resistance and epithelial-to-mesenchymal transition? *Frontiers in oncology* **2016**, *6*, 158.
17. Abdelmaksoud, N.M.; Abulsoud, A.I.; Abdelghany, T.M.; Elshaer, S.S.; Rizk, S.M.; Senousy, M.A. Mitochondrial remodeling in colorectal cancer initiation, progression, metastasis, and therapy: A review. *Pathology-Research and Practice* **2023**, *246*, 154509.
18. Holst, S.; Wuhler, M.; Rombouts, Y. Glycosylation characteristics of colorectal cancer. *Advances in cancer research* **2015**, *126*, 203-256.
19. Fernandes, E.; Sores, J.; Cotton, S.; Peixoto, A.; Ferreira, D.; Freitas, R.; Reis, C.A.; Santos, L.L.; Ferreira, J.A. Esophageal, gastric and colorectal cancers: Looking beyond classical serological biomarkers towards glycoproteomics-assisted precision oncology. *Theranostics* **2020**, *10*, 4903.
20. Ullah, I.; Yang, L.; Yin, F.-T.; Sun, Y.; Li, X.-H.; Li, J.; Wang, X.-J. Multi-omics approaches in colorectal cancer screening and diagnosis, recent updates and future perspectives. *Cancers* **2022**, *14*, 5545.
21. Ahluwalia, P.; Ballur, K.; Leeman, T.; Vashisht, A.; Singh, H.; Omar, N.; Mondal, A.K.; Vaibhav, K.; Baban, B.; Kolhe, R. Incorporating Novel Technologies in Precision Oncology for Colorectal Cancer: Advancing Personalized Medicine. *Cancers* **2024**, *16*, 480.
22. Satam, H.; Joshi, K.; Mangrolia, U.; Waghoo, S.; Zaidi, G.; Rawool, S.; Thakare, R.P.; Banday, S.; Mishra, A.K.; Das, G. Next-generation sequencing technology: current trends and advancements. *Biology* **2023**, *12*, 997.
23. Zafari, N.; Bathaei, P.; Velayati, M.; Khojasteh-Leylakoochi, F.; Khazaei, M.; Fiuji, H.; Nassiri, M.; Hassanian, S.M.; Ferns, G.A.; Nazari, E. Integrated analysis of multi-omics data for the discovery of biomarkers and therapeutic targets for colorectal cancer. *Computers in Biology and Medicine* **2023**, *155*, 106639.
24. Xu, Z.; Li, W.; Dong, X.; Chen, Y.; Zhang, D.; Wang, J.; Zhou, L.; He, G. Precision medicine in colorectal cancer: Leveraging multi-omics, spatial omics, and artificial intelligence. *Clinica Chimica Acta* **2024**, *119686*.
25. Nascimben, M. Machine learning approaches for personalized medicine. **2024**.
26. Mathema, V.B.; Sen, P.; Lamichhane, S.; Orešič, M.; Khoomrung, S. Deep learning facilitates multi-data type analysis and predictive biomarker discovery in cancer precision medicine. *Computational and Structural Biotechnology Journal* **2023**, *21*, 1372-1382.
27. Butz, H.; Szabo, P.M.; Nofech-Mozes, R.; Rotondo, F.; Kovacs, K.; Mirham, L.; Girgis, H.; Boles, D.; Patocs, A.; Yousef, G.M. Integrative bioinformatics analysis reveals new prognostic biomarkers of clear cell renal cell carcinoma. *Clinical Chemistry* **2014**, *60*, 1314-1326.
28. Sheng, K.L.; Kang, L.; Pridham, K.J.; Dunkenberger, L.E.; Sheng, Z.; Varghese, R.T. An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression. *Scientific reports* **2020**, *10*, 21246.

29. Ma, H.; Xiong, L.; Zhao, B.; Hahan, Z.; Wei, M.; Shi, H.; Yang, S.; Ren, Q. Comprehensive investigation into the influence of glycosylation on head and neck squamous cell carcinoma and development of a prognostic model for risk assessment and anticipating immunotherapy. *Frontiers in Immunology* **2024**, *15*, 1364082.
30. Liu, X.; Shi, J.; Tian, L.; Xiao, B.; Zhang, K.; Zhu, Y.; Zhang, Y.; Jiang, K.; Zhu, Y.; Yuan, H. Comprehensive prognostic and immune analysis of a glycosylation related risk model in pancreatic cancer. *BMC cancer* **2023**, *23*, 1229.
31. Gao, Y.; Li, W.; Guo, H.; Hao, Y.; Lu, L.; Li, J.; Piao, S. Construction of an abnormal glycosylation risk model and its application in predicting the prognosis of patients with head and neck cancer. *Scientific Reports* **2024**, *14*, 1310.
32. Shih, P.C.; Chen, H.P.; Hsu, C.C.; Lin, C.H.; Ko, C.Y.; Hsueh, C.W.; Huang, C.Y.; Chu, T.H.; Wu, C.C.; Ho, Y.C.; et al. Long-term DEHP/MEHP exposure promotes colorectal cancer stemness associated with glycosylation alterations. *Environ Pollut* **2023**, *327*, 121476, doi:10.1016/j.envpol.2023.121476.
33. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. *CA: a cancer journal for clinicians* **2018**, *68*, 7-30.
34. Arnold, M.; Abnet, C.C.; Neale, R.E.; Vignat, J.; Giovannucci, E.L.; McGlynn, K.A.; Bray, F. Global burden of 5 major types of gastrointestinal cancer. *Gastroenterology* **2020**, *159*, 335-349. e315.
35. Varki, A. Biological roles of glycans. *Glycobiology* **2017**, *27*, 3-49.
36. Marisa, L.; de Reyniès, A.; Duval, A.; Selves, J.; Gaub, M.P.; Vescovo, L.; Etienne-Grimaldi, M.C.; Schiappa, R.; Guenot, D.; Ayadi, M.; et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* **2013**, *10*, e1001453, doi:10.1371/journal.pmed.1001453.
37. Dube, D.H.; Bertozzi, C.R. Glycans in cancer and inflammation--potential for therapeutics and diagnostics. *Nat Rev Drug Discov* **2005**, *4*, 477-488, doi:10.1038/nrd1751.
38. Hastie, T. The elements of statistical learning: data mining, inference, and prediction. **2009**.
39. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **1996**, *58*, 267-288.
40. Binnewies, M.; Roberts, E.W.; Kersten, K.; Chan, V.; Fearon, D.F.; Merad, M.; Coussens, L.M.; Gabrilovich, D.I.; Ostrand-Rosenberg, S.; Hedrick, C.C. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature medicine* **2018**, *24*, 541-550.
41. Fuster, M.M.; Esko, J.D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature Reviews Cancer* **2005**, *5*, 526-542.
42. Zare, I.; Kiadeh, S.Z.H.; Varol, A.; Varol, T.Ö.; Varol, M.; Sezen, S.; Zarepour, A.; Mostafavi, E.; Nasab, S.Z.; Rahi, A. Glycosylated nanoplateforms: From glycosylation strategies to implications and opportunities for cancer theranostics. *Journal of Controlled Release* **2024**, *371*, 158-178.
43. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nature medicine* **2019**, *25*, 24-29.
44. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine* **2019**, *25*, 44-56.
45. Moqri, M.; Herzog, C.; Poganik, J.R.; Ying, K.; Justice, J.N.; Belsky, D.W.; Higgins-Chen, A.T.; Chen, B.H.; Cohen, A.A.; Fuellen, G. Validation of biomarkers of aging. *Nature medicine* **2024**, *30*, 360-372.
46. de Freitas Junior, J.C.M.; Morgado-Díaz, J.A. The role of N-glycans in colorectal cancer progression: Potential biomarkers and therapeutic applications. *Oncotarget* **2015**, *7*, 19395.
47. Lv, Y.; Wang, X.; Li, X.; Xu, G.; Bai, Y.; Wu, J.; Piao, Y.; Shi, Y.; Xiang, R.; Wang, L. Nucleotide de novo synthesis increases breast cancer stemness and metastasis via cGMP-PKG-MAPK signaling pathway. *PLoS Biol* **2020**, *18*, e3000872, doi:10.1371/journal.pbio.3000872.

48. Maeser, D.; Gruener, R.F.; Huang, R.S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data. *Brief Bioinform* **2021**, *22*, doi:10.1093/bib/bbab260.
49. Geeleher, P.; Cox, N.J.; Huang, R.S. Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biology* **2014**, *15*, R47, doi:10.1186/gb-2014-15-3-r47.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.