Robust Glycogene-Based Prognostic Signature for Proficient Mismatch Repair Colorectal Adenocarcinoma

Yixi Li[1,2†], Dehua Li[3†], Yang Chen[3†], Yongping Lu[2], Fangbin Zhou[1], Chunhong Li[1], Zhipeng Zeng[1], Wanxia Cai[1], Liewen Lin[1], Qiang Li[4], Mingjun Ye[2], Jingjing Dong[2], Lianghong Yin[2*], Donge Tang[1*], Gong Zhang[3*] and Yong Dai[1,5*]

[1] Clinical Medical Research Center, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Jinan University, Shenzhen, China, [2] Institute of Nephrology and Blood Purification, The First Affiliated Hospital of Jinan University, Jinan University, Guangzhou, China, [3] Key Laboratory of Functional Protein Research of Guangdong Higher Education Institutes and MOE Key Laboratory of Tumor Molecular Biology, Institute of Life and Health Engineering, Jinan University, Guangzhou, China, [4] Department of Nephrology, Dongguan Hospital of Guangzhou University of Traditional Chinese Medicine, Dongguan, China, [5] Guangxi Key Laboratory of Metabolic Diseases Research, Affiliated No. 924 Hospital, Southern Medical University, Guilin, China

Background: Proficient mismatch repair (pMMR) colorectal adenocarcinoma (CRAC) metastasizes to a greater extent than MMR-deficient CRAC. Prognostic biomarkers are preferred in clinical practice. However, traditional biomarkers screened directly from sequencing are often not robust and thus cannot be confidently utilized.

Methods: To circumvent the drawbacks of blind screening, we established a new strategy to identify prognostic biomarkers in the conserved and specific oncogenic pathway and its regulatory RNA network. We performed RNA sequencing (RNA-seq) for messenger RNA (mRNA) and noncoding RNA in six pMMR CRAC patients and constructed a glycosylation-related RNA regulatory network. Biomarkers were selected based on the network and their correlation with the clinicopathologic information and were validated in multiple centers (n = 775).

Results: We constructed a competing endogenous RNA (ceRNA) regulatory network using RNA-seq. Genes associated with glycosylation pathways were embedded within this scale-free network. Moreover, we further developed and validated a seven-glycogene prognosis signature, GlycoSig (*B3GNT6*, *GALNT3*, *GALNT8*, *ALG8*, *STT3B*, *SRD5A3*, and *ALG6*) that prognosticate poor-prognostic subtype for pMMR CRAC patients. This biomarker set was validated in multicenter datasets, demonstrating its robustness and wide applicability. We constructed a simple-to-use nomogram that integrated the risk score of GlycoSig and clinicopathological features of pMMR CRAC patients.

Conclusions: The seven-glycogene signature served as a novel and robust prognostic biomarker set for pMMR CRAC, highlighting the role of a dysregulated glycosylation network in poor prognosis.

Keywords: colorectal cancer, mismatch repair, glycosylation, glycogene, ceRNA, biomarker

INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer-related deaths worldwide (1). Currently, the 5-year survival rate for patients with CRC varies from over 90% in stage I to slightly greater than 10% in stage IV (2). Colorectal adenocarcinoma (CRAC) is the most common pathohistological type of CRC. DNA mismatch repair (MMR) helps maintain DNA replication fidelity (3) and is therefore a well-established biomarker for CRC. MMR-proficient (pMMR) presents in 85% of CRAC (4). The prognostic outcome of pMMR patients is worse than that of MMR-deficient (dMMR) patients, and pMMR CRAC is more likely to metastasize (5).

Numerous studies have attempted to identify prognostic biomarkers. A commonly used method is to compare macromolecules in tumor and normal tissues and to identify differentially expressed proteins or RNAs. The currently reported prognostic biomarkers for CRC include proteins HOXB5 (6), VSTM2A (7), and STYX (8); long noncoding RNAs (lncRNAs) MCHR2, AC011472.4, and AC063944.1 (9); circular RNAs (circRNAs) circFADS2 (10), circ_0026344 (11), and hsa_circ_0004831 (12); and microRNAs (miRNAs) miR-1290 (13) and five-miRNAs set (14). However, different studies have reported different biomarkers probably due to the cohort selection, genetic background, treatment, and methodological differences within each study, which present significant system errors. Consequently, no prognostic biomarkers (including biomarker sets) have been commercialized as standardized diagnostic kits in clinical practice. Due to genome instability, heterogeneity, and complexity of cancer progression, biomarkers derived from blind screening of differentially expressed genes might not be a robust method for prognosis.

Although individual genes in tumors can mutate quickly, oncogenic pathways are much more conserved. Inspired by this fact, we changed the strategy of biomarker identification by identifying biomarkers from common oncogenic pathways of pMMR CRAC. For example, glycosylation is a key posttranslational modification that regulates single sugar moieties for targeted molecules, leading to glycan elongation (15). Tumor-associated glycosylation changes stem from a series of molecular causes, as glycan biosynthesis is a non-template-driven process (16). The genes involved in their biosynthesis, degradation, transport, and recognition are known as glycogenes (17). Notably, alterations in glycogenes have been shown to be associated with pMMR CRAC. Galectin-3 is often overexpressed in pMMR tissues compared to dMMR tissues (18). The Tn antigen is frequently upregulated in dMMR CRC compared with pMMR (19). The lack of GALNT6 was observed in more than 50% of dMMR tumors, but it was less frequent (<12%) in pMMR tumors (20). Downregulation of GALNT6 occurs during the transition from precancerous neoplasia to invasive carcinoma in a certain subset of tumors that frequently exhibit dMMR (21). Therefore, we aimed to discover prognostic biomarkers of pMMR CRAC based on the glycogene network.

MATERIALS AND METHODS

Patients and Samples

A total of six pMMR CRAC tissue and paired adjacent noncancerous tissue specimens were collected from patients who underwent operations in Shenzhen People's Hospital (Shenzhen, China), and their clinical information is shown in **Table 1**. The new diagnosis of pMMR CRAC was confirmed by surgery and histopathologic examination, and patients had not received radiotherapy, chemotherapy, or biological immunotherapy before recruitment. The paired adjacent noncancerous tissue was defined as tissue that was at a 2.0-cm distance from the edge of the tumor and was free of tumor cells based on an experienced pathologist's evaluation. All resected tissue samples were snap-frozen in liquid nitrogen and stored at -80°C until use. The study was approved by the Medical Ethics Committee of Shenzhen People's Hospital, and consent was obtained from all participants. The workflow of the experimental process is illustrated in **Figure 1A**.

Total RNA Extraction and Quality Testing

Total RNA was extracted from pMMR CRAC tissues and paired non-tumorous tissues using TRIzol reagent (Invitrogen, USA). The RNA purity of each sample was quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher, USA). Total RNA quantity and integrity were assessed using a Bioanalyzer 2100 and RNA 6000 Nano LabChip Kit (Agilent, USA) with RNA integrity number (RIN) >7.0.

Construction of the Small RNA Library and Sequencing

Approximately 1 µg of total RNA was used to prepare a small RNA library according to the TruSeq Small RNA Sample Prep Kit (Illumina, USA). Then, we performed single-end sequencing (50 bp) on an Illumina Hiseq 2500 (LC Sciences, China) following the manufacturer's protocol.

TABLE 1 | Patients and tumor characteristics of the sequencing set.

Case No.	Gender	Age	Primary tumor location	TNM 8th	Stage	MMR status
1	M	55	Rectum	pT3/N1/M0	IIIB	pMMR
2	F	63	Left colon	pT3/N0/M0	IIA	pMMR
3	F	44	Left colon	pT3/N2a/M0	IIIB	pMMR
4	F	62	Right colon	pT2/N0/M0	I	pMMR
5	F	59	Rectum	pT3/N0/M0	IIA	pMMR
6	F	55	Rectum	pT4b/N1/M0	IIIC	pMMR

MMR, mismatch repair; pMMR, proficient mismatch repair.
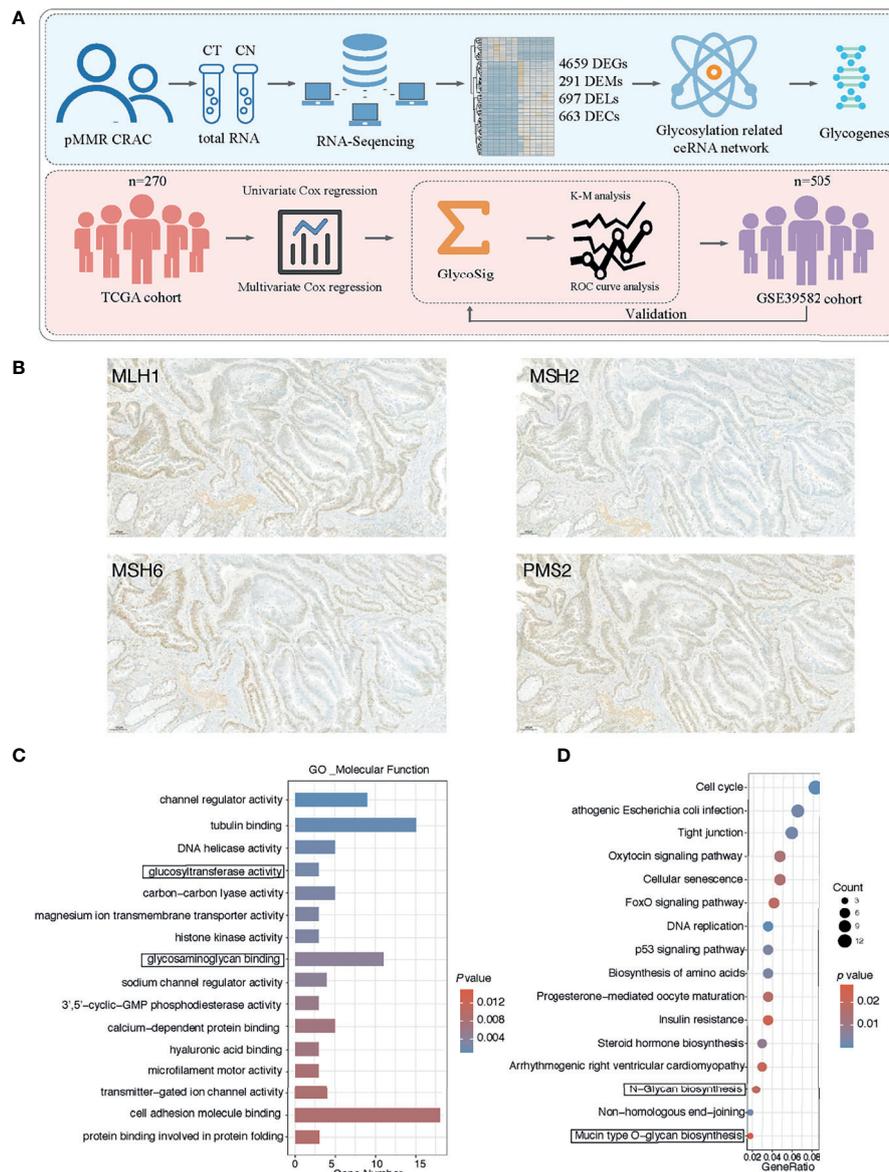
FIGURE 1 | Cancer-specific messenger RNAs (mRNAs) in proficient mismatch repair (pMMR) colorectal adenocarcinoma (CRAC) and functional enrichment analysis. **(A)** Workflow of the experiment process. **(B)** The representative pattern of mismatch repair (MMR) status for CRAC tissues. **(C)** Top 16 enriched Gene Ontology (GO) terms of molecular function; y axis represents GO terms, and x axis represents gene number. Color of the bars represent enrichment significance. **(D)** Top 16 enriched Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways; y axis represents pathway names, and x axis represents rich factor. Size and color of the bubble represent number of differentially expressed mRNAs (DEGs) enriched in the pathway and enrichment significance, respectively.

Read Alignment for miRNAs and Analysis of Differentially Expressed miRNAs

Raw data were analyzed as previously published (22). Raw reads were subjected to ACGT101-miR (LC Sciences, China) to remove adapter dimers, junk, low complexity, common RNA families, and repeats. Subsequently, unique sequences with lengths of ~18–26 nucleotides were mapped in miRBase 22.0 by BLAST search to identify known miRNAs and novel miRNAs. Differentially expressed miRNAs (DEMs) were selected with | log2 FC| ≥1 and p < 0.05 using the R package Ballgown.

Construction of the Ribo-Zero RNA Library and Sequencing

Ribosomal RNA was removed from approximately 5 μg of total RNA using the Ribo-Zero Gold rRNA Removal Kit (Illumina, USA). After purification, the remaining RNA was fragmented (Magnesium RNA Fragmentation Module, NEB, USA) into small pieces using magnesium ions at 94°C. The cleaved RNA fragments were reverse-transcribed to create the final cDNA library in accordance with the protocols for commercial Illumina library preparation kits (Illumina, USA) (23). For reverse

transcription, random primers were applied to the rRNA-depleted RNA to synthesize first-stranded RNA. Next, we synthesized U-labeled second-stranded DNAs using *Escherichia coli* DNA polymerase I (NEB, USA), RNase H (NEB, USA), and dUTP solution (Thermo Fisher, USA). An A-base was then added to the blunt ends of each strand, preparing them for ligation to the indexed adapters. Each adapter contained a T-base overhang to ligate the adapter to the A-tailed fragmented DNA. Single- or dual-index adapters were ligated to the fragments, and size selection (300–600 bp) was performed using AMPureXP beads. After heat-labile UDG enzyme (NEB, USA) treatment of the U-labeled second-stranded DNAs, the ligated products were amplified by PCR under the following conditions: initial denaturation at 95°C for 3 min; eight cycles of denaturation at 98°C for 15 s, annealing at 60°C for 15 s, extension at 72°C for 30 s; and a final extension at 72°C for 5 min. The average insert size of the final cDNA library was 300 ± 50 bp. Finally, we performed 2 × 150 bp paired-end sequencing on an Illumina Hiseq 4000 (LC-Bio Technology, China) following the vendor's recommended protocol. Ribo-Zero RNA sequencing was used to analyze not only messenger RNAs (mRNAs) but also lncRNAs and circRNAs.

Transcript Assembly and Analysis of Differentially Expressed mRNAs

Cutadapt (24) (https://cutadapt.readthedocs.io/en/stable) was used to remove reads containing adaptor contamination, low-quality bases, and undetermined bases. Sequence quality was verified using FastQC (25) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). We used Hisat2 (26) (https://ccb.jhu.edu/software/hisat2) to map reads to the genome of *Homo sapiens*. The mapped reads of each sample were assembled using StringTie (27) (http://ccb.jhu.edu/software/stringtie). Then, all transcriptomes from the samples were merged to reconstruct a comprehensive transcriptome using Gffcompare (https://github.com/gpertea/gffcompare). After the final transcriptome was generated, StringTie was used to determine mRNA expression levels by calculating FPKM (28). The differentially expressed mRNAs (DEGs) were selected with |log2 FC| ≥1 and p < 0.05 using the R package Ballgown. Driver genes for the DEGs were identified by using OncoVar analysis (29), and the result was shown in **Table S1**.

LncRNA Identification and Analysis for Differentially Expressed lncRNAs

First, StringTie assembled and quantified the reads mapped to the genome of *Homo sapiens*. Second, transcripts that overlapped with known mRNAs and transcripts shorter than 200 bp were discarded. This is because lncRNAs are usually defined as RNAs that are >200 nt in length and do not encode proteins. We then utilized CPC (30) (http://cpc2.cbi.pku.edu.cn) and CNCI (31) (http://www.bioinfo.org/software/cnci) to predict transcripts with coding potential. All transcripts with CPC score <-1 and CNCI score <0 were retained without coding potential, and the remaining transcripts were considered lncRNAs. Differentially expressed lncRNAs (DELs) were selected with |log2 FC| ≥1 and p < 0.05 using the R package Ballgown.

CircRNA Identification and Analysis for Differentially Expressed circRNAs

Cutadapt (24) and FastQC (25) were used as previously described. We first used TopHat2 (32) to map reads to the genome of *Homo sapiens*, which were removed as linear RNAs. Unmapped reads in the first step were mapped to the genome using the Tophat-fusion algorithm according to the splicing site features, such as GU/AG, GC/AG, and AU/AC. Back-spliced reads were identified in unmapped reads using the Tophat-fusion algorithm and CIRCExplorer (33). All samples were processed separately to identify circRNAs. circRNA expression from different samples was calculated using scripts in the house. Differentially expressed circRNAs (DECs) were selected with |log2 FC | ≥1 and p < 0.05 using the R package Ballgown.

The Cancer Genome Atlas and Gene Expression Omnibus Data Mining

Gene expression and clinical information of COAD and READ patients from The Cancer Genome Atlas (TCGA) datasets were downloaded from UCSC xena (http://xenabrowser.net/), and the GSE39582 data were downloaded from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/). The clinical data of the two cohorts are shown in **Table 2**. We selected all eligible CRAC samples with MMR status of the pMMR. First, each candidate gene was evaluated using univariate Cox regression analysis. Only candidate genes that correlated with prognosis were used for subsequent analysis. Then, all the significant genes in the univariate Cox regression were used to construct the prognostic GlycoSig model by multivariate Cox regression analysis. Kaplan–Meier (K-M) survival analysis and receiver operating characteristic (ROC) curve analysis were used to assess the predictive value of this GlycoSig model. The R packages on above workflow include "survival", "survminer", "GEOquery", "ggplot2", "Hmisc", "grid", "lattice", "Formula", and "rms."

Functional Enrichment

Enrichment of Gene Ontology (GO) analysis: DEGs were classified by GO annotation into three categories: biological process, cellular component, and molecular function. For each category, a two-tailed Fisher's exact test was used to test the enrichment of DEGs against all the identified mRNAs. GO with a corrected p < 0.05 was considered significant. For the Kyoto Encyclopedia of Gene and Genomes (KEGG) pathway enrichment analysis, survival analysis for all DEGs was first performed using the K-M algorithm combined with TCGA datasets. The enriched pathways were then tested for prognosis-related DEGs using the clusterProfiler R package. The pathway with a corrected p < 0.05 was considered significant. These analyses could identify the most enriched pathways in the DEGs. The most reliable biomarkers would be among them.

ceRNA Network Construction and Network Attribute

Salmena et al. (34) proposed the competing endogenous RNA (ceRNA) hypothesis. This hypothesis states that all types of RNA

TABLE 2 | Clinicopathologic characteristics of the different sets for patients with colorectal cancer.

Characteristics	COAD Cohort (n = 551)	READ Cohort (n = 186)	GSE39582 Cohort (n = 585)
Mean age (SD, range), years	67 (13, 31–90)	64 (12, 31–90)	67 (13, 54–80)
Gender (male/female) (percent)	284/262 (52/48)	99/87 (53/47)	323/263 (55/45)
Pathological stage (percent)			
I	86 (16)	35 (19)	37 (6)
II	219 (40)	55 (30)	266 (46)
III	152 (27)	56 (30)	209 (36)
IV	78 (14)	28 (15)	60 (10)
NA	16 (3)	11 (6)	13 (2)
Primary tumor location (percent)			
Rectum	0	186 (100)	NA
Left colon	218 (40)	0	NA
Right colon	309 (56)	0	NA
NA	24 (4)	0	NA
Mismatch repair protein (percent)			
pMMR	343 (62)	142 (76)	459 (79)
dMMR	66 (12)	9 (5)	77 (13)
NA	142 (26)	35 (19)	49 (8)
Median follow-up (SD, range), months	28 (26, 0–150)	24 (23, 0–131)	2 (1.3, 0–6.7)

dMMR, mismatch repair-deficient; MMR, mismatch repair; pMMR, proficient mismatch repair.
NA, Not Available.

transcripts communicate through miRNA response elements (MREs). The ceRNA network for glycosylation was constructed using the following steps: first, DELs, DECs, and 3′-untranslated region (UTR) sequences of DEGs were predicted as miRNA targets using Miranda (Miranda Energy <-10) and TargetScan (TargetScan score ≥50) tools. Only the associations of miRNA–mRNA, miRNA–lncRNA, and miRNA–circRNA presented in the two methods were used to construct the ceRNA network. We then selected all genes related to glycosylation in the KEGG pathway enrichment analysis. DEM nodes related to glycogenes combined with DEL and DEC nodes were chosen to construct a sub-ceRNA network. Finally, the ceRNA network was generated using Cytoscape 3.8.2. The mRNAs and miRNAs with p < 0.05, |log2 FC| ≥1 and lncRNAs and circRNAs with p < 0.01 and |log2 FC| ≥3 were retained in the ceRNA network. To identify the nature of the network, we used the poweRlaw package to fit the power law distribution of the node degree distribution and created a power law scatter plot.

miRNA-Related Heatmap

We calculated the degree of all nodes in the ceRNA network using Cytoscape and then selected the top 10 degrees of miRNAs and the mRNAs related to these miRNAs. Finally, we plotted a heatmap to show these molecules using the pheatmap R package.

Statistical Analysis

The paired-samples t-test was used for the significance analysis of DEGs, DELs, DECs, and DEMs. The likelihood ratio test was used for multiple Cox regression analysis and univariate Cox regression analysis. The log-rank test was used for K-M survival analysis. Statistical analyses were carried out using SPSS Statistics 19.0 (SPSS Inc., USA). Other analyses were carried out by R version 3.5.2 with the following packages: "edgeR", "pheatmap", "forestplot", "rms", "ggplot2", "survivalROC", and "survival".

All hypothesis testing was two-sided, and p < 0.05 was defined as statistically significant.

RESULTS

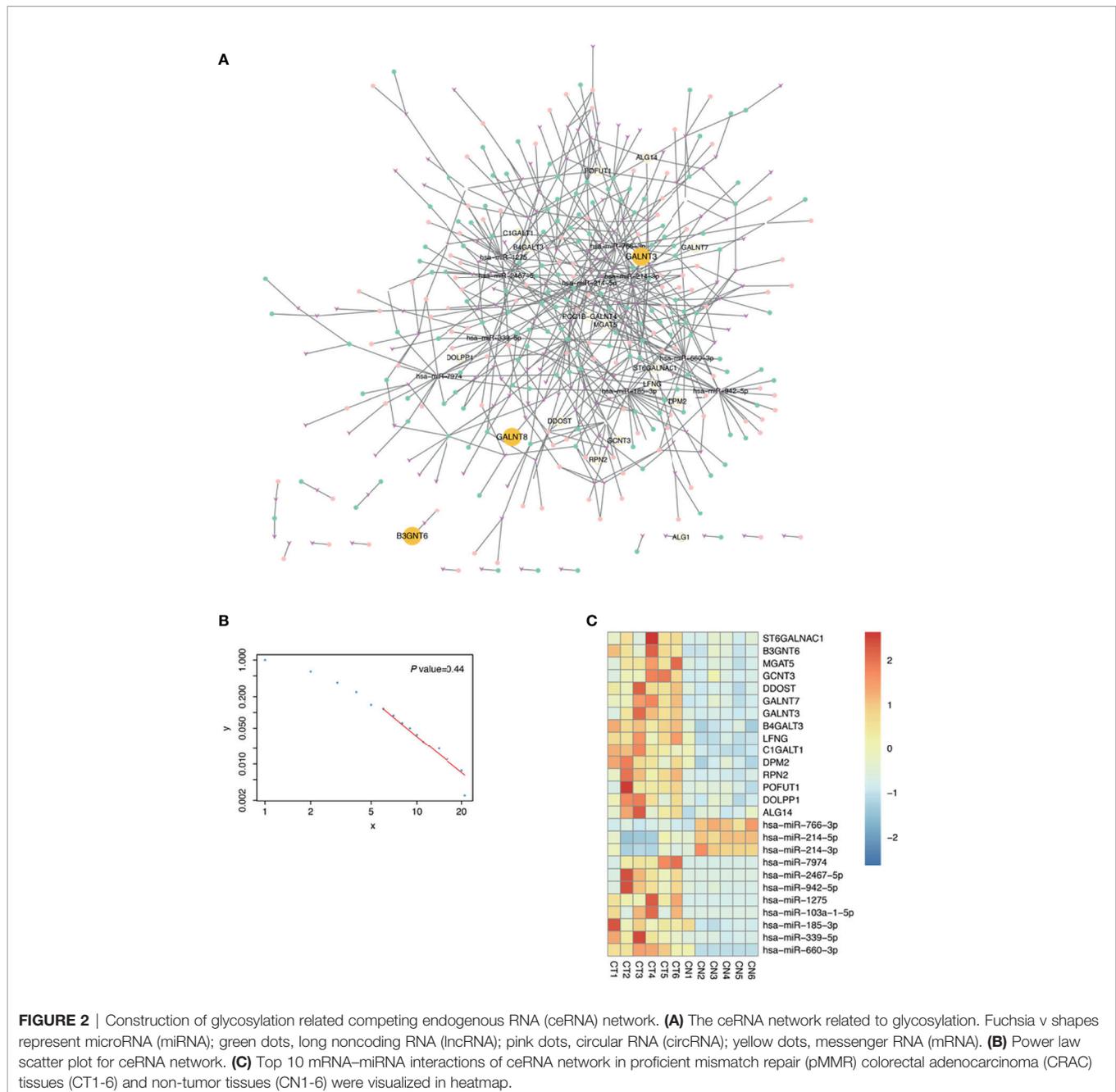## Glycogenes Are Enriched in Differentially Expressed mRNAs

We collected the tumor and normal tissues of six pMMR CRAC. The positive MMR status of all tumor tissues was confirmed using immunohistochemistry against the MMR proteins MLH1, MSH2, MSH6, and PMS2 (**Figure 1B**). We performed mRNA sequencing (mRNA-seq) for paired tumors and normal tissues. At the threshold of |log2 FC| ≥1 and p < 0.05, a total of 4,659 DEGs were identified in the pMMR CRAC tissues of the six patients studied, compared with adjacent non-tumor tissues, including 2,783 upregulated DEGs and 1,876 downregulated DEGs (**Figure S1A**). *MSH2* as one of the driver genes was identified by using OncoVar analysis (**Table S1**), indicating the significant role of MMR status for CRC patients. GO and KEGG pathway enrichment analysis showed that the glycosylation-related pathways were among the top 3 enriched GO molecular function terms (**Figure 1C**). The two most enriched terms were tubulin binding (35) and cell adhesion molecule binding (36), which are universal for cancers as previously reported. Moreover, the glycan synthesis pathways were also among the most enriched KEGG pathways (**Figure 1D**). These results showed that glycogenes are commonly dysregulated in the pMMR CRAC.

The miRNA–lncRNA–circRNA–mRNA ceRNA Network of Glycogenes

As numerous glycogenes were differentially expressed in the RNA-seq and considering the potential bias and errors of omics technology, identifying true signatures of pMMR CRAC

requires further mining of these data. The disorder of ceRNA crosstalk profoundly contributes to the CRC process (37). Glycogenes interact with and are regulated by the miRNA–lncRNA–circRNA ceRNA network and are thus stabilized in tumor cells. As for textbook knowledge, regulatory networks are more conserved than individual genes. Therefore, the DEGs that are highly robust in the regulatory network were more likely to be signatures. To construct such a network, we sequenced the miRNAs, lncRNAs, and circRNAs of the six patients. At the threshold of $|\log2 FC| \geq 1$ and $p < 0.05$, a total of 291 DEMs, 697 DELs, and 663 DECs were identified in pMMR CRAC tissues

compared with adjacent non-tumor tissues (**Figures S1B–D**). Using such criteria to identify nodes and edges, a ceRNA network for glycosylation pathways consisting of 18 DEGs, 152 DEMs, 153 DELs, and 115 DECs was constructed (**Figure 2A**), and the heatmap for DEGs, DEMs, DELs, and DECs is shown in **Figures S3A, D–F**, respectively. Most of the nodes are interconnected as a main graph, and the degree of the nodes follows the power law distribution (**Figure 2B**), suggesting that they form a typical biological scale-free network. This indicates that the glycogenes tend to be tightly connected by various noncoding RNAs. The top 10 mRNA–miRNA interactions in

FIGURE 2 | Construction of glycosylation related competing endogenous RNA (ceRNA) network. **(A)** The ceRNA network related to glycosylation. Fuchsia v shapes represent microRNA (miRNA); green dots, long noncoding RNA (lncRNA); pink dots, circular RNA (circRNA); yellow dots, messenger RNA (mRNA). **(B)** Power law scatter plot for ceRNA network. **(C)** Top 10 mRNA–miRNA interactions of ceRNA network in proficient mismatch repair (pMMR) colorectal adenocarcinoma (CRAC) tissues (CT1-6) and non-tumor tissues (CN1-6) were visualized in heatmap.

the ceRNA network showed 15 upregulated glycogenes and corresponding miRNAs (**Figure 2C**). Notably, these DEGs and DEMs were clearly valid in all six samples, indicating the robustness of the network-based screening strategy.

GlycoSig Model for the Prognosis of Proficient Mismatch Repair Colorectal Adenocarcinoma

Assuming that aberrant glycosylation is a hallmark of CRAC, we further investigated prognosis-related glycogenes for pMMR CRAC. We first performed univariate Cox regression analysis for all candidate genes related to glycosylation in KEGG pathway to evaluate the correlation between gene expression and overall survival (OS) through TCGA database for pMMR CRAC patients (n = 270). Seven glycogenes were statistically significant and thus considered variables for prognosis (**Table 3**), including five upregulated genes and two downregulated genes for TCGA dataset (**Figure 3A**). *B3GNT6*, *GALNT3*, and *GALNT8* belong to O-linked glycosylation processes, while *ALG8*, *STT3B*, *SRD5A3*, and *ALG6* belong to N-linked glycosylation processes. The seven-glycogene expression heatmap of in-house samples is shown in **Figure S3C**. Finally, we found that a signature based on glycogenes (GlycoSig) could be a prognostic factor for pMMR CRAC patients using multivariate Cox regression analysis (p < 0.0001; **Figure 3B**). Regression coefficients were calculated for the entire set, and the risk score of GlycoSig was identified for each patient using the following formula:

$$\begin{aligned} \text{GlycoSig risk score} \\ = (-4.4 \times ALG8) + (-1.3 \times SRD5A3) + (-0.14 \times ALG6) \\ + (-2.1 \times STT3B) + (0.11 \times GALNT3) \\ + (-0.71 \times B3GNT6) + (-0.013 \times GALNT8) \end{aligned}$$

In particular, *B3GNT6*, *GALNT3*, and *GALNT8* as members of GlycoSig were also included in the ceRNA network, suggesting that this biomarker set should be robust. *B3GNT6* [hazard ratio (HR) = 0.491, 95% CI 0.283–0.852, p = 0.011], *ALG8* (HR = 0.013, 95% CI 0.001–0.126, p<0.001), and *STT3B* (HR = 0.117, 95% CI 0.015–0.943, p = 0.044) might be associated with the development of poor prognosis of pMMR CRAC. The 270 patients were categorized into high- and low-risk groups using the median risk score of GlycoSig as the cutoff. Survival analysis was performed using the K-M method with a log-rank statistical test. Patients in the high-risk group had significantly worse OS than those in the low-risk group (median OS = 5.7 years, log-rank p < 0.0001; **Figure 3C**).

## The Prognostic Power of the GlycoSig Biomarker Set

Clinicopathological factors of patients with pMMR CRAC are also known to have prognostic value. To compare the GlycoSig biomarker set to such clinicopathological factors, we performed univariate Cox regression analysis for clinicopathological factors, including age, sex, primary tumor location, and pathological stage, on TCGA data of 270 patients. Age less than 65 years (HR = 2.6, 95% CI 1.2–5.4, p = 0.012), stage I (HR = 0.11, 95% CI 0.02–0.59, p = 0.010), and stage II (HR = 0.24, 95% CI 0.077–0.76, p = 0.015) were significantly correlated with OS of pMMR CRAC patients (**Table 3**). The clinical information alone reached the areas under the ROC curve (AUC) of 0.727, 0.721, and 0.642 at 1, 3, and 5 years, respectively (**Figure 4A**). For the GlycoSig model only, the AUC reached 0.720, 0.741, and 0.714 at 1, 3, and 5 years, respectively (**Figure S2H**), indicating that the performance of GlycoSig was robust. Combining the clinical information with GlycoSig, the AUC reached 0.740, 0.752, and 0.733 at 1, 3, and 5 years, respectively (**Figure 4B**). The elevation of AUC, especially the remarkable increase in the long term,

TABLE 3 | Univariate Cox regression analysis for clinical features and candidate genes.

		Univariate analysis			
		HR	95% CI		P value
Clinical features	Sex	1.30	0.69	2.40	0.421
	Age less than 65 years	2.60	1.20	5.40	0.012*
	Primary tumor location				
	Rectum	1.30	0.53	3.00	0.598
	Right colon	1.40	0.64	2.90	0.426
	Pathologic stage				
	Stage I	0.11	0.02	0.59	0.010*
	Stage II	0.24	0.08	0.76	0.015*
	Stage III	0.56	0.19	1.70	0.301
	Stage IV	0.39	0.10	1.60	0.189
Candidate genes	ALG8	0.017	0.0017	0.17	<0.0001***
	SRD5A3	0.19	0.04	0.92	0.040*
	ALG6	0.03	0.0023	0.39	0.007**
	STT3B	0.07	0.0095	0.52	0.009**
	GALNT3	0.21	0.058	0.75	0.016*
	B3GNT6	0.48	0.32	0.72	<0.001***
	GALNT8	0.47	0.27	0.82	0.008**

HR, hazard ratio.
**P < 0.05, **P < 0.01, ***P < 0.001.*
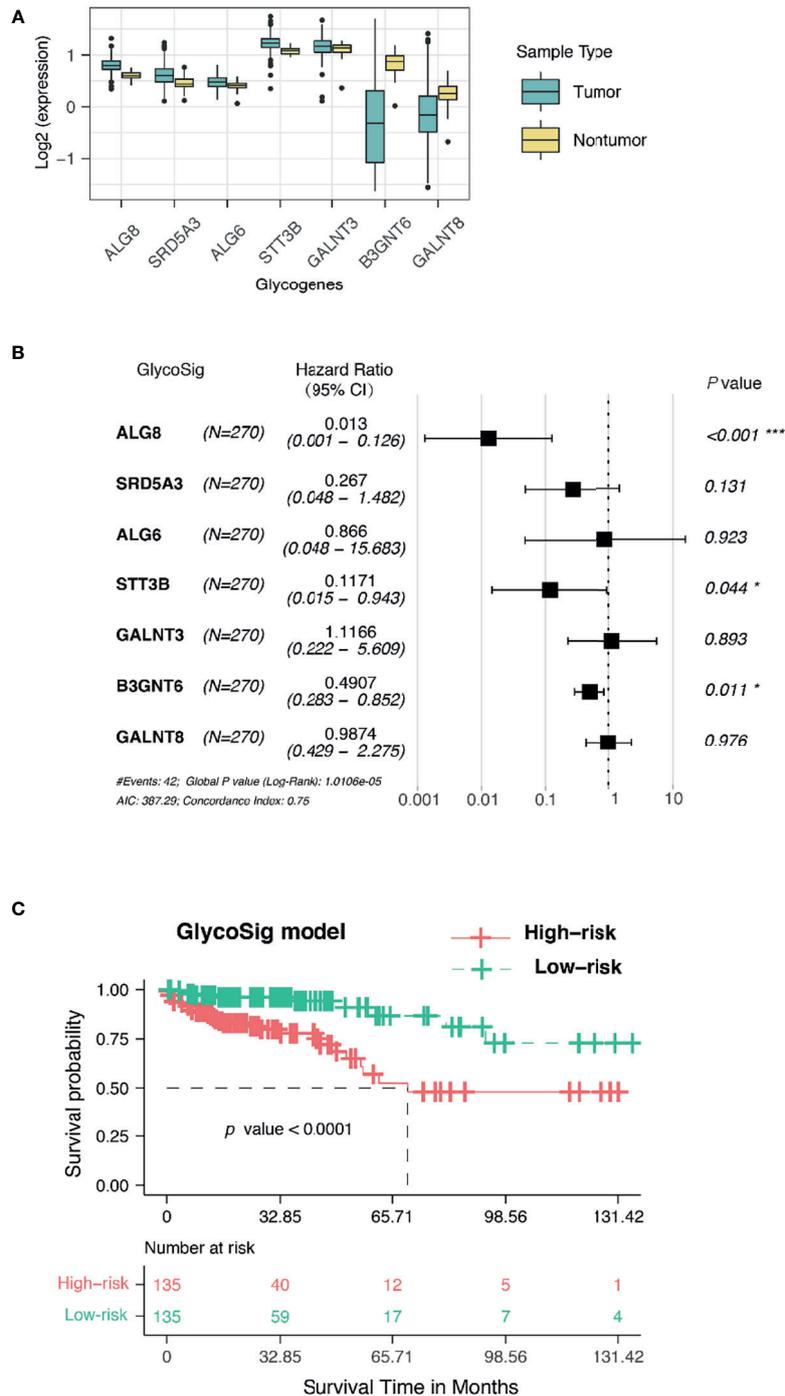
FIGURE 3 | GlycoSig model for the prognosis of proficient mismatch repair (pMMR) colorectal adenocarcinoma (CRAC) in The Cancer Genome Atlas (TCGA) dataset.
(A) The expression of seven glycogenes in tumor and non-tumor tissues from TCGA database. (B) Forest plots showed the multivariate Cox regression analysis for
GlycoSig; *P < 0.05; ***P < 0.001. (C) Kaplan–Meier survival curves of pMMR CRAC patients classified into high- and low-risk groups using the GlycoSig.

showed the prognostic power and robustness of the GlycoSig biomarker set. When the patients were sorted according to the GlycoSig risk score, their clinical features were randomly distributed. As for the seven glycogenes in GlycoSig, *B3GNT6*,

ALG6, *ALG8*, and *SRD5A3* showed a slight correlation with the risk score, while the other three genes were visibly random (**Figure 4C**). In fact, these gene expression levels fluctuate considerably in general. The expression of single glycogenes
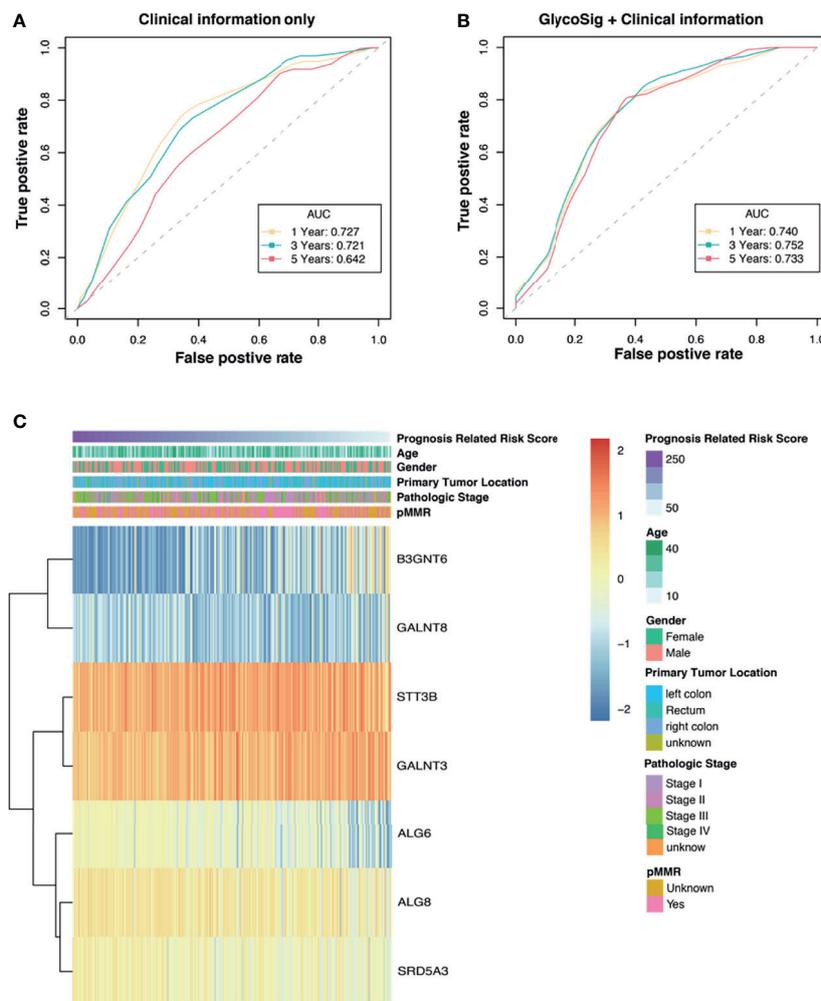
FIGURE 4 | The prognostic power of the GlycoSig biomarker set. Receiver operating characteristic (ROC) curve analysis of clinicopathologic information **(A)** and GlycoSig combined with clinical information **(B)**; Yellow, green, and red dotted lines represent GlycoSig risk score predicting 1-, 3-, and 5-year overall survival. **(C)** Heatmap of the GlycoSig combined with clinical features and messenger RNA (mRNA) expression of glycogenes in The Cancer Genome Atlas (TCGA) dataset.

could also provide prognostic value to a certain extent (**Figures S2A–G**), but not as significant as the biomarker set. These results also validated the necessity of the multi-gene GlycoSig in terms of sensitivity and robustness.

GlycoSig Validation and Nomogram Constructed

Many biomarkers can only be used in restricted cohorts and are not robust when tested in different populations. To test whether our GlycoSig is widely applicable, we downloaded data from 505 patients from the GEO database (accession number: GSE39582) (38), including the transcriptome profiles and clinical features. The patients in GSE39582 were from the French National Cartes d'Identité des Tumeurs (CIT) program. The risk scores were calculated for each patient. Interestingly, the survival analysis showed that patients with a high-risk score for GlycoSig had a shorter survival time than those with a low-risk score (p = 0.029;

Figure 5A). This validated the robustness of the GlycoSig biomarker set and its general applicability. We further constructed a simple-to-use nomogram that integrated the risk score of GlycoSig and clinicopathological features to predict the 1-, 3-, and 5-year survival probabilities of patients who had undergone surgical resection (**Figure 5B**). The points from each independent prognostic factor listed in the nomogram were summated. The total points were calculated by adding up the corresponding points of each individual covariate on the point scale.

DISCUSSION

CRAC is heterogeneous and characterized by different molecular and phenotypic characteristics (39). Although MMR testing can provide important information for clinical decision-making in
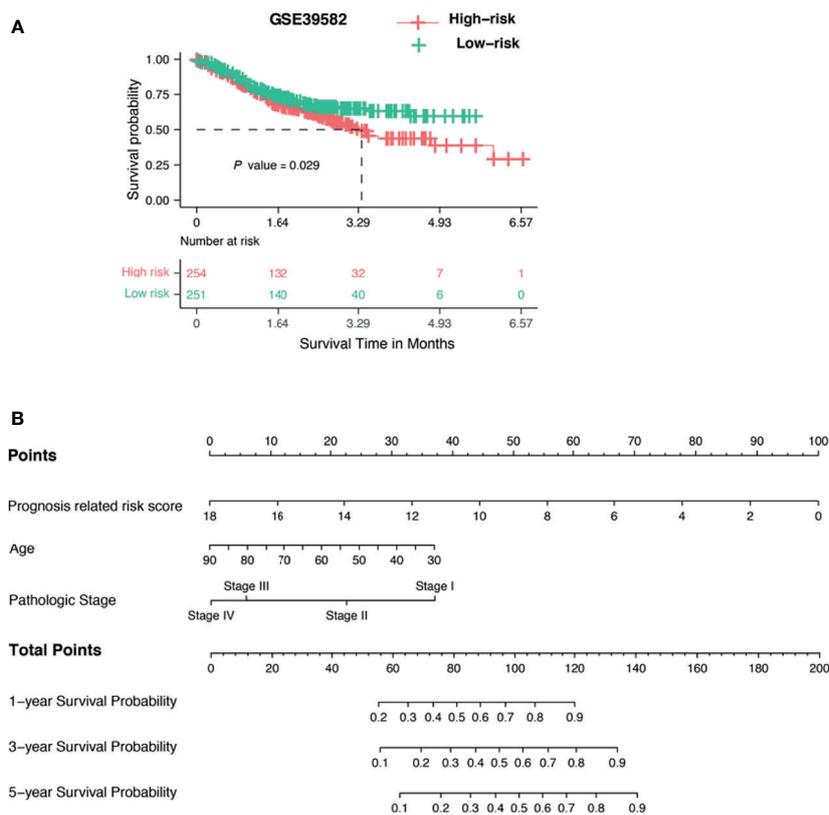
FIGURE 5 | Validation of GlycoSig and nomogram for proficient mismatch repair (pMMR) colorectal adenocarcinoma (CRAC). **(A)** Independent validation of GlycoSig in the GSE39582 set; Kaplan–Meier survival curves of patients classified into high- and low-risk groups using the GlycoSig for overall survival. **(B)** Nomograms to predict 1-, 3-, and 5-year survival probability in pMMR CRAC.

CRAC, pMMR CRAC patients with the same clinical features may have different prognostic and therapeutic responses. Previous research on prognostic biomarkers has generally found differentially expressed genes using next-generation sequencing or proteome mass spectrometry. However, this strategy does not produce robust biomarker sets generally due to cohort selection. A multicenter study is a possible way to avoid center-originated bias. However, methodological bias makes the normalization of multicenter datasets very difficult. Here, our strategy is to avoid blind screening by focusing on specific pathways and constructing the RNA regulatory network, which is more conserved than individual genes. We focused on aberrant glycosylation, which is a hallmark of CRC and has been shown to be altered during tumor development and progression (40). Indeed, our strategy found a signature of seven glycogenes based on the functional pathway and ceRNA network.

In particular, aberrant glycosylation has been confirmed to protect cancer cells from radiotherapeutic and drug-targeted treatments. Ionizing radiation treatment for CRC increases the sialylation of β1-integrin, which enhances CRC cell adhesion, migration, and invasion (41). The curative effect of anti-vascular endothelial growth factor (VEGF) therapy depends on glycosylation of the cancer cell surface, and the removal of α2-

6-linked sialic acid results in tumor resistance to anti-VEGF therapy (42). It was noted that specific glycans in tumor cells, such as Tn antigen, could serve as a novel immune checkpoint, offering new immunotherapeutic opportunities (43). A novel recombinant human chimeric immunoglobulin G (IgG)1 anti-Tn antibody, named Remab6, represents a new useful reagent to detect Tn-positive glycoproteins as a biomarker for human carcinomas and may also be a novel therapeutic agent for targeted cancer treatment (44). In addition, the monoclonal antibody CH129 that targets tumor-associated sialylated glycan demonstrated its potential for multimodal cancer therapy (45). Hence, targeting aberrant glycosylation as an immunotherapeutic strategy is a major field of research against CRC.

Three GlycoSig genes, *B3GNT6*, *GALNT3*, and *GALNT8*, have been noticed by other studies in CRC. Iwai et al. (46) demonstrated that the expression of *B3GNT6* is significantly decreased in CRC tissues and is a useful marker for distinguishing benign adenomas from premalignant lesions. Moreover, decreased expression of *B3GNT6* is related to epithelial–mesenchymal transition (EMT) and metastasis in CRC (46–48), which provides evidence for the prognostic value of *B3GNT6*. In addition, *GALNT3* has been shown to be regulated by the linc01296/miR-26a network in CRC (49). Tang et al. (50) found that lncRNA *GAU1* regulates the expression

of *GALNT8* in CRC. The overexpression of *GALNT8* significantly accelerates the cell cycle and thus promotes CRC cell lines and correlates with poor prognosis in CRC patients. Silencing of *GALNT8* suppresses the cancer cell proliferation and induced resistance against oxaliplatin in CRC cell lines. The results suggested that the *GALNT8* may play as a CRC prognosis marker and potential target against chemoresistance (50).

The other four genes of GlycoSig have not been investigated in the context of CRC; however, many studies have indicated their significance in other cancer types. Previous studies have shown that *ALG8* could be a variate of prognostic model for gastric adenocarcinoma (51) and frequently amplified hotspot on 11q14.1 (*ALG8*) associated with significantly worse prognosis for breast cancer (52). The *ALG6* single-nucleotide polymorphism (SNP) is a potential prognostic biomarker for cutaneous melanoma and is likely through modulating gene expression (53). Hsu et al. (54) demonstrate that EMT enriched programmed death-ligand 1 (PD-L1) in cancer stem-like cells by the EMT/β-catenin/STT3/PD-L1 signaling axis, in which EMT transcriptionally induced N-glycosyltransferase STT3 (including isoforms STT3A and STT3B) through β-catenin, and subsequent STT3-dependent PD-L1 N-glycosylation stabilized and upregulated PD-L1. Zhang et al. (55) found that *SRD5A3* was highly expressed in breast cancer tissues, and high *SRD5A3* expression was related to poorer prognosis. Also, Mai et al. (56) demonstrated that *SRD5A3* is upregulated in hepatocellular carcinoma (HCC) tissues, and higher *SRD5A3* level predicts poor OS, progression-free survival, relapse-free survival, and disease-specific survival in HCC patients. *SRD5A3* polymorphism may contribute to a genetic predisposition for prostate cancer (57). Metastatic prostate cancer expressed higher transcript levels for *SRD5A3* (58). The SRD5A3 enzyme converts testosterone to dihydrotestosterone (DHT). It is highly expressed in metastatic prostate cancer compared to benign and localized prostate cancer (59). Knockdown of *SRD5A3* expression in prostate cancer cells resulted in a significant decrease in DHT production and a drastic reduction in cell viability (60). These findings indicate that *SRD5A3* should be a promising molecular target for prostate cancer therapy.

Although most of the glycogenes are not well characterized in pMMR CRAC patients, they showed mechanistic and diagnostic significance in other cancers. As a textbook knowledge, pathways are more conserved than single proteins. Therefore, such information also indicates the potential mechanistic contribution of these glycogenes in CRAC. As we provided robust evidence for their connection to prognosis, further experimental studies on these prognosis-related glycogenes are worth conducting. Another limitation of this study is that the prognostic value of the GlysoSig was validated in two retrospective cohorts. Therefore, the applicability of the GlysoSig should be further verified in prospective pMMR CRAC cohorts. Also, their mechanism in pMMR CRAC is worth an investigation.

In conclusion, we screened and validated the seven-glycogene prognostic signature (GlycoSig) for pMMR CRAC patients using the pathway–network strategy. These findings indicate that the seven glycogenes should be a promising molecular target for pMMR CRAC therapy.

DATA AVAILABILITY STATEMENT

The raw data of RNA sequencing are available in the GSA for Human (https://bigd.big.ac.cn/gsub/) of BIG Sub at accession number HRA000868 and HRA000866. Other datasets generated during the current study are available from the corresponding authors upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Shenzhen People's Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

YD, DT, and LY designed the research. FZ, CL, ZZ, WC, and LL collected the data. YXL, DL, YC, and YPL performed the data analysis. YXL and GZ wrote the article (YXL drafted the initial article; GZ and YXL finalized the article). QL, MY, and JD critically reviewed the article. YD and DT had primary responsibility for the final content. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.727752/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590

2. Brenner H, Kloor M, Pox CP. Colorectal Cancer. *Lancet* (2014) 383 (9927):1490–502. doi: 10.1016/S0140-6736(13)61649-9

3. Hewish M, Lord CJ, Martin SA, Cunningham D, Ashworth A. Mismatch Repair Deficient Colorectal Cancer in the Era of Personalized Treatment. *Nat Rev Clin Oncol* (2010) 7(4):197–208. doi: 10.1038/nrclinonc.2010.18

4. Kishore C, Bhadra P. Current Advancements and Future Perspectives of Immunotherapy in Colorectal Cancer Research. *Eur J Pharmacol* (2021) 893:173819. doi: 10.1016/j.ejphar.2020.173819

5. Gryfe R, Kim H, Hsieh ET, Aronson MD, Holowaty EJ, Bull SB, et al. Tumor Microsatellite Instability and Clinical Outcome in Young Patients With Colorectal Cancer. *N Engl J Med* (2000) 342(2):69–77. doi: 10.1056/NEJM200001133420201

6. Feng W, Huang W, Chen J, Qiao C, Liu D, Ji X, et al. CXCL12-Mediated HOXB5 Overexpression Facilitates Colorectal Cancer Metastasis Through Transactivating CXCR4 and ITGB3. *Theranostics* (2021) 11(6):2612–33. doi: 10.7150/thno.52199

7. Dong Y, Zhang Y, Kang W, Wang G, Chen H, Higashimori A, et al. VSTM2A Suppresses Colorectal Cancer and Antagonizes Wnt Signaling Receptor LRP6. *Theranostics* (2019) 9(22):6517–31. doi: 10.7150/thno.34989

8. He D, Ma Z, Fang C, Ding J, Yang W, Chen P, et al. Pseudophosphatase STYX Promotes Tumor Growth and Metastasis by Inhibiting FBXW7 Function in Colorectal Cancer. *Cancer Lett* (2019) 454:53–65. doi: 10.1016/j.canlet.2019.04.014

9. Chu Y, Liu Z, Liu J, Yu L, Zhang D, Pei F. Characterization of lncRNA-Perturbed TLR-Signaling Network Identifies Novel lncRNA Prognostic Biomarkers in Colorectal Cancer. *Front Cell Dev Biol* (2020) 8:503. doi: 10.3389/fcell.2020.00503

10. Xiao Y-S, Tong H-Z, Yuan X-H, Xiong C-H, Xu X-Y, Zeng Y-F. CircFADS2: A Potential Prognostic Biomarker of Colorectal Cancer. *Exp Biol Med* (2020) 245(14):1233–41. doi: 10.1177/1535370220929965

11. Yuan Y, Liu W, Zhang Y, Zhang Y, Sun S. CircRNA Circ_0026344 as a Prognostic Biomarker Suppresses Colorectal Cancer Progression *via* microRNA-21 and microRNA-31. *Biochem Biophys Res Commun* (2018) 503(2):870–5. doi: 10.1016/j.bbrc.2018.06.089

12. Xing L, Xia M, Jiao X, Fan L. Hsa_circ_0004831 Serves as a Blood-Based Prognostic Biomarker for Colorectal Cancer and Its Potentially circRNA-miRNA-mRNA Regulatory Network Construction. *Cancer Cell Int* (2020) 20 (1):557. doi: 10.1186/s12935-020-01651-8

13. Imaoka H, Toiyama Y, Fujikawa H, Hiro J, Saigusa S, Tanaka K, et al. Circulating microRNA-1290 as a Novel Diagnostic and Prognostic Biomarker in Human Colorectal Cancer. *Ann Oncol* (2016) 27(10):1879–86. doi: 10.1093/annonc/mdw279

14. Yang G, Zhang Y, Yang J. A Five-microRNA Signature as Prognostic Biomarker in Colorectal Cancer by Bioinformatics Analysis. *Front Oncol* (2019) 9:1207. doi: 10.3389/fonc.2019.01207

15. Karve TM, Cheema AK. Small Changes Huge Impact: The Role of Protein Posttranslational Modifications in Cellular Homeostasis and Disease. *J Amino Acids* (2011) 2011:207691. doi: 10.4061/2011/207691

16. Almeida A, Kolarich D. The Promise of Protein Glycosylation for Personalised Medicine. *Biochim Biophys Acta Gen Subj* (2016) 1860 (8):1583–95. doi: 10.1016/j.bbagen.2016.03.012

17. Dall'Olio F, Trinchera M. Epigenetic Bases of Aberrant Glycosylation in Cancer. *Int J Mol Sci* (2017) 18(5):998. doi: 10.3390/ijms18050998

18. Gebert J, Kloor M, Lee J, Lohr M, Andre S, Wagner R, et al. Colonic Carcinogenesis Along Different Genetic Routes: Glycophenotyping of Tumor Cases Separated by Microsatellite Instability/Stability. *Histochem Cell Biol* (2012) 138(2):339–50. doi: 10.1007/s00418-012-0957-9

19. Matsumoto T, Okayama H, Nakajima S, Saito K, Nakano H, Endo E, et al. Tn Antigen Expression Defines an Immune Cold Subset of Mismatch-Repair Deficient Colorectal Cancer. *Int J Mol Sci* (2020) 21(23):9081. doi: 10.3390/ijms21239081

20. Okayama H, Noda M, Kono K. Aberrant Glycosylation in Colorectal Cancer With Genomic and Epigenomic Alterations. *Oncotarget* (2018) 9(86):35609–10. doi: 10.18632/oncotarget.26292

21. Noda M, Okayama H, Tachibana K, Sakamoto W, Saito K, Min AKT, et al. Glycosyltransferase Gene Expression Identifies a Poor Prognostic Colorectal Cancer Subtype Associated With Mismatch Repair Deficiency and Incomplete Glycan Synthesis. *Clin Cancer Res* (2018) 24(18):4468–81. doi: 10.1158/1078-0432.CCR-17-3533

22. Zhou F, Tang D, Xu Y, He H, Wu Y, Lin L, et al. Identification of microRNAs and Their Endonucleolytic Cleavaged Target mRNAs in Colorectal Cancer. *BMC Cancer* (2020) 20(1):242. doi: 10.1186/s12885-020-06717-4

23. Schuierer S, Carbone W, Knehr J, Petitjean V, Fernandez A, Sultan M, et al. A Comprehensive Assessment of RNA-Seq Protocols for Degraded and Low-Quantity Samples. *BMC Genomics* (2017) 18(1):442. doi: 10.1186/s12864-017-3827-y

24. Kechin A, Boyarskikh U, Kel A, Filipenko M. Cutprimers: A New Tool for Accurate Cutting of Primers From Reads of Targeted Next Generation Sequencing. *J Comput Biol* (2017) 24(11):1138–43. doi: 10.1089/cmb.2017.0096

25. Beekman R, Chapaprieta V, Russinol N, Vilarrasa-Blasi R, Verdaguer-Dot N, Martens JHA, et al. The Reference Epigenome and Regulatory Chromatin Landscape of Chronic Lymphocytic Leukemia. *Nat Med* (2018) 24(6):868–80. doi: 10.1038/s41591-018-0028-4

26. Kim D, Langmead B, Salzberg SL. HISAT: A Fast Spliced Aligner With Low Memory Requirements. *Nat Methods* (2015) 12(4):357–60. doi: 10.1038/nmeth.3317

27. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie Enables Improved Reconstruction of a Transcriptome From RNA-Seq Reads. *Nat Biotechnol* (2015) 33(3):290–5. doi: 10.1038/nbt.3122

28. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript Assembly and Quantification by RNA-Seq Reveals Unannotated Transcripts and Isoform Switching During Cell Differentiation. *Nat Biotechnol* (2010) 28(5):511–5. doi: 10.1038/nbt.1621

29. Wang T, Ruan S, Zhao X, Shi X, Teng H, Zhong J, et al. OncoVar: An Integrated Database and Analysis Platform for Oncogenic Driver Variants in Cancers. *Nucleic Acids Res* (2021) 49(D1):D1289–301. doi: 10.1093/nar/gkaa1033

30. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: Assess the Protein-Coding Potential of Transcripts Using Sequence Features and Support Vector Machine. *Nucleic Acids Res* (2007) 35(Web Server issue): W345–9. doi: 10.1093/nar/gkm391

31. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing Sequence Intrinsic Composition to Classify Protein-Coding and Long non-Coding Transcripts. *Nucleic Acids Res* (2013) 41(17):e166. doi: 10.1093/nar/gkt646

32. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: Accurate Alignment of Transcriptomes in the Presence of Insertions, Deletions and Gene Fusions. *Genome Biol* (2013) 14(4):R36. doi: 10.1186/gb-2013-14-4-r36

33. Zhang XO, Dong R, Zhang Y, Zhang JL, Luo Z, Zhang J, et al. Diverse Alternative Back-Splicing and Alternative Splicing Landscape of Circular RNAs. *Genome Res* (2016) 26(9):1277–87. doi: 10.1101/gr.202895.115

34. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language? *Cell* (2011) 146(3):353–8. doi: 10.1016/j.cell.2011.07.014

35. Kavallaris M. Microtubules and Resistance to Tubulin-Binding Agents. *Nat Rev Cancer* (2010) 10(3):194–204. doi: 10.1038/nrc2803

36. Carpenter G, Red Brewer M. EpCAM: Another Surface-to-Nucleus Missile. *Cancer Cell* (2009) 15(3):165–6. doi: 10.1016/j.ccr.2009.02.005

37. Qi X, Lin Y, Chen J, Shen B. The Landscape of Emerging ceRNA Crosstalks in Colorectal Cancer: Systems Biological Perspectives and Translational Applications. *Clin Trans Med* (2020) 10(4):e153. doi: 10.1002/ctm2.153

38. Marisa L, de Reynies A, Duval A, Selves J, Gaub MP, Vescovo L, et al. Gene Expression Classification of Colon Cancer Into Molecular Subtypes: Characterization, Validation, and Prognostic Value. *PLoS Med* (2013) 10(5): e1001453. doi: 10.1371/journal.pmed.1001453

39. Fanelli GN, Dal Pozzo CA, Depetris I, Schirripa M, Brignola S, Biason P, et al. The Heterogeneous Clinical and Pathological Landscapes of Metastatic Braf-Mutated Colorectal Cancer. *Cancer Cell Int* (2020) 20:30. doi: 10.1186/s12935-020-1117-2

40. Mereiter S, Balmana M, Gomes J, Magalhaes A, Reis CA. Glycomic Approaches for the Discovery of Targets in Gastrointestinal Cancer. *Front Oncol* (2016) 6:55. doi: 10.3389/fonc.2016.00055

41. Lee M, Lee HJ, Seo WD, Park KH, Lee YS. Sialylation of Integrin Beta1 Is Involved in Radiation-Induced Adhesion and Migration in Human Colon Cancer Cells. *Int J Radiat Oncol Biol Phys* (2010) 76(5):1528–36. doi: 10.1016/j.ijrobp.2009.11.022

42. Croci DO, Cerliani JP, Dalotto-Moreno T, Mendez-Huergo SP, Mascanfroni ID, Dergan-Dylon S, et al. Glycosylation-Dependent Lectin-Receptor Interactions Preserve Angiogenesis in Anti-VEGF Refractory Tumors. *Cell* (2014) 156(4):744–58. doi: 10.1016/j.cell.2014.01.043

43. RodrIguez E, Schetters STT, van Kooyk Y. The Tumour Glyco-Code as a Novel Immune Checkpoint for Immunotherapy. *Nat Rev Immunol* (2018) 18 (3):204–11. doi: 10.1038/nri.2018.3

44. Matsumoto Y, Kudelka MR, Hanes MS, Lehoux S, Dutta S, Jones MB, et al. Identification of Tn Antigen O-GalNAc-Expressing Glycoproteins in Human Carcinomas Using Novel Anti-Tn Recombinant Antibodies. *Glycobiology* (2020) 30(5):282–300. doi: 10.1093/glycob/cwz095

45. Tivadar ST, McIntosh RS, Chua JX, Moss R, Parsons T, Zaitoun AM, et al. Monoclonal Antibody Targeting Sialyl-Di-Lewis(a)-Containing Internalizing and Noninternalizing Glycoproteins With Cancer Immunotherapy Development Potential. *Mol Cancer Ther* (2020) 19(3):790–801. doi: 10.1158/1535-7163.MCT-19-0221

46. Iwai T, Kudo T, Kawamoto R, Kubota T, Togayachi A, Hiruma T, et al. Core 3 Synthase Is Down-Regulated in Colon Carcinoma and Profoundly Suppresses the Metastatic Potential of Carcinoma Cells. *Proc Natl Acad Sci U S A* (2005) 102(12):4572–7. doi: 10.1073/pnas.0407983102

47. An G, Wei B, Xia B, McDaniel JM, Ju T, Cummings RD, et al. Increased Susceptibility to Colitis and Colorectal Tumors in Mice Lacking Core 3-Derived O-Glycans. *J Exp Med* (2007) 204(6):1417–29. doi: 10.1084/jem.20061929

48. Ye J, Wei X, Shang Y, Pan Q, Yang M, Tian Y, et al. Core 3 Mucin-Type O-Glycan Restoration in Colorectal Cancer Cells Promotes MUC1/p53/miR-200c-Dependent Epithelial Identity. *Oncogene* (2017) 36(46):6391–407. doi: 10.1038/onc.2017.241

49. Liu B, Pan S, Xiao Y, Liu Q, Xu J, Jia L. LINC01296/miR-26a/GALNT3 Axis Contributes to Colorectal Cancer Progression by Regulating O-Glycosylated MUC1 *via* PI3K/AKT Pathway. *J Exp Clin Cancer Res* (2018) 37:316. doi: 10.1186/s13046-018-0994-x

50. Tang X, Ruan H, Dong L, Li S, Wu Z, Guan M. lncRNA GAU1 Induces GALNT8 Overexpression and Potentiates Colorectal Cancer Progression. *Gastroenterol Res Pract* (2021) 2021:5960821. doi: 10.1155/2021/5960821

51. Zhou W, Li J, Lu X, Liu F, An T, Xiao X, et al. Derivation and Validation of a Prognostic Model for Cancer Dependency Genes Based on CRISPR-Cas9 in Gastric Adenocarcinoma. *Front Oncol* (2021) 11:617289. doi: 10.3389/fonc.2021.617289

52. Chin SF, Teschendorff AE, Marioni JC, Wang Y, Barbosa-Morais NL, Thorne NP, et al. High-Resolution aCGH and Expression Profiling Identifies a Novel Genomic Subtype of ER Negative Breast Cancer. *Genome Biol* (2007) 8(10): R215. doi: 10.1186/gb-2007-8-10-r215

53. Zhou B, Zhao YC, Liu H, Luo S, Amos CI, Lee JE, et al. Novel Genetic Variants of ALG6 and GALNTL4 of the Glycosylation Pathway Predict Cutaneous Melanoma-Specific Survival. *Cancers (Basel)* (2020) 12(2):288. doi: 10.3390/cancers12020288

54. Hsu JM, Xia W, Hsu YH, Chan LC, Yu WH, Cha JH, et al. STT3-Dependent PD-L1 Accumulation on Cancer Stem Cells Promotes Immune Evasion. *Nat Commun* (2018) 9(1):1908. doi: 10.1038/s41467-018-04313-6

55. Zhang YP, Na WT, Dai XQ, Li RF, Wang JX, Gao T, et al. Over-Expression of SRD5A3 and Its Prognostic Significance in Breast Cancer. *World J Surg Oncol* (2021) 19(1):260. doi: 10.1186/s12957-021-02377-1

56. Mai Q, Sheng D, Chen C, Gou Q, Chen M, Huang X, et al. Steroid 5 Alpha-Reductase 3 (SRD5A3) Promotes Tumor Growth and Predicts Poor Survival of Human Hepatocellular Carcinoma (HCC). *Aging (Albany NY)* (2020) 12 (24):25395–411. doi: 10.18632/aging.104142

57. Park JM, Song KH, Lim JS, Kim JW, Sul CK. Is the Expression of Androgen Receptor Protein Associated With the Length of AC Repeats in the Type III 5-α Reductase Gene in Prostate Cancer Patients? *Korean J Urol* (2013) 54 (6):404–8. doi: 10.4111/kju.2013.54.6.404

58. Mitsiades N, Sung CC, Schultz N, Danila DC, He B, Eedunuri VK, et al. Distinct Patterns of Dysregulated Expression of Enzymes Involved in Androgen Synthesis and Metabolism in Metastatic Prostate Cancer Tumors. *Cancer Res* (2012) 72(23):6142–52. doi: 10.1158/0008-5472.can-12-1335

59. Ateeq B, Vellaichamy A, Tomlins SA, Wang R, Cao Q, Lonigro RJ, et al. Role of Dutasteride in Pre-Clinical ETS Fusion-Positive Prostate Cancer Models. *Prostate* (2012) 72(14):1542–9. doi: 10.1002/pros.22509

60. Uemura M, Tamura K, Chung S, Honma S, Okuyama A, Nakamura Y, et al. Novel 5 Alpha-Steroid Reductase (SRD5A3, Type-3) Is Overexpressed in Hormone-Refractory Prostate Cancer. *Cancer Sci* (2008) 99(1):81–6. doi: 10.1111/j.1349-7006.2007.00656.x