

Original Article

A co-regulatory network of *SPIB*, *AQP8*, and *GUCA2B* related to immune infiltration for early-stage colorectal cancer in silico and in vitro

Je-Ming Hu^{1,2,3}, Pei-Yao Liu⁴, Ying-Chuan Chen⁴, Wei-Zhi Lin⁵, Yu-Ching Chou⁶, Wen-Chiuan Tsai⁷, Chi-Ming Chu^{6,8,9,10,11}, Chia-Chao Wu¹², Yu-Tien Chang⁶

¹Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan; ²Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan; ³School of Medicine, National Defense Medical Center, Taipei, Taiwan; ⁴Department of Physiology and Biophysics, National Defense Medical Center, Taipei, Taiwan; ⁵AIOT Center, Tri-Service General Hospital, Taipei, Taiwan; ⁶School of Public Health, National Defense Medical Center, Taipei, Taiwan; ⁷Department of Pathology, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan; ⁸Division of Biostatistics and Informatics, Department of Epidemiology, School of Public Health, National Defense Medical Center, Taipei, Taiwan; ⁹Big Data Research Center, Fu-Jen Catholic University, New Taipei, Taiwan; ¹⁰Department of Public Health, China Medical University, Taichung, Taiwan; ¹¹Department of Healthcare Administration and Medical Informatics College of Health Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan; ¹²Division of Nephrology, Department of Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan

Received July 5, 2023; Accepted October 4, 2023; Epub November 15, 2023; Published November 30, 2023

Abstract: In early-stage colorectal cancer (CRC), *AQP8*, *GUCA2B*, and *SPIB* were important suppressor genes and frequently co-expressed. However, the underlying co-regulation effect remains unknown and need to be elucidated. We aimed to investigate the co-regulatory network of *AQP8*, *GUCA2B*, and *SPIB* in CRC using in vitro and in silico methods. Q-PCR, western blot, and immunohistochemistry were used to assess the co-regulatory network of the target genes in the HCT-116 cell line and fresh tumor tissues. Bioinformatical methods were used to validate the findings using the Cancer Genome Atlas COlon ADenocarcinoma and REctum ADenocarcinoma datasets, as well as large scale integrated data sets from Gene Expression Omnibus. In clinical CRC tissues, *SPIB*, *AQP8*, and *GUCA2B* were barely expressed compared to normal mucosa. When compared to 22 well-known genetic biomarkers, they are independent predictors of CRC identification with near 100% accuracy. In the co-regulatory network, they were co-upregulated at the mRNA and protein expression levels. *AQP8*, *GUCA2B* and *SPIB* were linked to immune cell infiltration and *GUCA2B* and *SPIB* were negatively associated with tumor purity. The co-regulatory network in miRNA-mRNA analysis was mediated by cancer-related microRNAs miR-182-5p and miR-27a-3. The functional analysis of the co-regulatory network's protein-protein interaction networks reveals three clusters and three major functions: complex interactions of transcription factors in mediating cytokine biology in T cells (*SPIB* cluster), guanylin, and Intestinal infectious diseases (*GUCA2B* cluster), and water channel activity balance (*AQP8* cluster). The co-regulatory network of *SPIB*, *AQP8*, and *GUCA2B* was confirmed. MiR-27a-3p and miR-182-5p were two possible mediators. The mechanisms of *SPIB*, *AQP8*, *GUCA2B*, miR-182-5p, and miR-27a-3p in CRC merit further investigation.

Keywords: Colorectal cancer, co-regulatory network, immune infiltration, microRNA, early stage

Introduction

Colorectal cancer (CRC) is a major public health problem. It is the third most common cancer and the fourth leading cause of cancer-related death worldwide [1]. Timely diagnosis and proper management can cure up to 80% of patients [2]; however, numerous cases are diagnosed

only when this tumor has advanced because CRC is asymptomatic in the early stage. Late diagnosis could lead to treatment challenges and reduced survival times.

Few diagnostic genetic biomarkers are currently being utilized in clinical practice [3, 4]. One reason is that the findings for diagnostic bio-

markers have lacked overlap among studies. These disparities among biomarkers are caused by differences in such aspects as tumor heterogeneity, dataset source, analysis platforms, and approaches [5]. In our previous study [5], the top candidate genes were highly associated. We speculated that their co-function might cause the discrepancy in genetic biomarkers from various studies in the same co-regulatory network.

Defects in multiple tumor suppressor genes are markedly associated with carcinogenesis and cancer [6-8]. The accumulation of genetic and epigenetic alterations is the driving force for CRC tumorigenesis [9]. In our previous study [5], we used a large number of integrated genome datasets and multiple bioinformatics approaches, including machine learning and traditional statistics, to determine *AQP8*, *GUCA2B*, and *SPIB* were repeatedly selected as top suppressor genes for CRC identification.

AQP8 has been discovered to be differentially expressed in a variety of cancers [10]. The upregulation of *AQP8* inhibited CRC cell proliferation in vivo [11]. *AQP8* inhibits colorectal cancer cell proliferation and metastasis by interfering with PI3K/AKT signaling and regulating *PCDH7* expression [10]. Uroguanylin (*GUCA2B*-encoded) is involved in the regulation of intestinal secretion [12]. Uroguanylin inhibits intestinal epithelial cell proliferation by upregulating nuclear transcription of cell cycle inhibitors (p21 and p27) and inhibiting proliferative transcription activated by Wnt/beta-catenin/tcf and AKT pathways [13]. *SPIB* functions as a tumour suppressor in colorectal cancer cells by activating the NFkB and JNK signalling pathways via *MAP4K1* [14]. Additionally, *AQP8*, *GUCA2B*, and *SPIB* were frequently co-expressed and confirmed as significant biomarkers to early CRC [15-18]. However, the underlying co-regulation effect remains unknown. Therefore, we aimed to investigate the co-regulatory network of *AQP8*, *GUCA2B*, and *SPIB* in CRC using in vitro and in silico methods.

Methods and materials

Datasets

RNA sequencing data were retrieved from The Cancer Genome Atlas (TCGA) colon adenocarcinoma and rectum adenocarcinoma datasets

(COADREAD) containing 20,531 probes and 434 samples. COADREAD consists of data on 380 primary tumors, two recurrent tumors, one metastatic tumor, and 51 normal solid tissues. Recurrent and metastatic tumors were excluded because of the small sample size. The gene expression profile was experimentally measured using an Illumina HiSeq 2000 RNA sequencing platform at TCGA Genome Characterization Center of the University of North Carolina. This dataset revealed gene-level transcription estimates as $\log_2(x + 1)$ -transformed RSEM normalized counts. We used the integrated mRNA dataset from GSE4045, GSE4107, GSE4183, GSE5851, GSE8671, GSE9348, GSE1096, GSE12630, GSE12945, GSE13067, GSE13294, GSE13471, GSE15960, GSE17538, GSE18105, and GSE14333 [5], mentioned as “integrated dataset” in the study. There were 88 cases of normal mucosa, 53 cases of adenoma, 521 cases of adenocarcinoma and 79 cases of metastatic tumors.

Statistics

COADREAD was used to examine the genetic predictors of CRC/NM classification and prognosis prediction (overall survival [OS] and relapse-free survival [RFS]) under the control of clinical and demographic characteristics. The classification and prognosis prediction analyses involved logistic regression and the Cox proportional hazard model, respectively. The significance threshold was set at $P < 0.05$. These analyses were performed using SPSS 20.0 and R software (<https://www.r-project.org/>). Accuracy, sensitivity, specificity, and area under the curve (AUC) of receiver operating characteristic (ROC) curves were calculated to evaluate the classification of CRC and NM.

T-distributed stochastic neighbor embedding and heatmap

T-SNE (t-Distributed Stochastic Neighbor Embedding) is a popular nonlinear dimensionality reduction technique that allows for high-dimensional data visualization and clustering. It preserves the inherent relationships between data points while projecting the data's local and global structure into a lower-dimensional space. T-SNE, unlike traditional methods such as PCA, excels at preserving complex relationships, making it an effective tool for uncovering hidden structures and visualizing data distributions. We use t-SNE to cluster the CRC data into

different tissue types. A heatmap was plotted to analyze the gene expression patterns using the R package “pheatmap”.

Receiver operating characteristic curve

ROC curves were plotted to evaluate the classification performance regarding CRCs of the genes of interest using COADREAD and R packages of “pROC” [19], “ggplot2”, and “tidyverse” [20]. The reference script is available at <https://stackoverflow.com/questions/66505014/how-to-add-auc-to-a-multiple-roc-graph-with-procs-ggroc>. AUC ranges from 0 to 1. The higher the AUC, the better the model.

Protein-protein interaction

The protein-protein interaction of *AQP8*, *GUCA2B* and *SPIB* were analyzed using STRING (<https://string-db.org/>). The networks were clustered using kmeans clustering. Function analysis was conducted using Gene Ontology (GO) knowledgebase.

Ethics statement

The handling of tissue samples and patient data in the present study was approved by the Tri-Service General Hospital Institutional Review Board in Taiwan (IRB; TSGHIRB approval number: 098-05-292). The board was organized and operated in accordance with the International Conference on Harmonization (ICH)/WHO GCP and applicable laws and regulations. Written informed consent was obtained from each patient, documented in each patient's record, and considered sufficient by the ethics committee. Tissue samples were registered as case numbers without names or personal identification numbers.

Experiment on clinical specimens

Specimens: The pairs of specimens from normal and tumor tissues were collected from five CRC patients with an average age of 66 ± 10 years old and BMIs ranging from 20 to 28. Sample collection was performed in surgical clinics, where the tumor and normal tissues were simultaneously resected. Adjacent normal tissue specimens were collected from an incision > 10 cm away from the carcinoma sites. All the specimens were immediately stored in liquid nitrogen. The resection procedure was

reviewed by the Department of Colorectal Surgery, Tri-Service General Hospital.

Hematoxylin-eosin (H&E) and immunohistochemical analysis: Neoplastic colon and non-neoplastic tissues around the tumor were collected from the waste from surgical resection. Tissue preparation and staining were processed following regular protocol. Formalin-fixed, paraffin-embedded tissues were sliced into 4 μm -thick sections and stained with H&E. For immunohistochemical (IHC) analysis, formalin-fixed, paraffin-embedded tissue sections with 4 μm -thickness were deparaffined and rehydrated in xylene. Antigens were retrieved by immersing sections in 10 mM sodium citrate at a pre-heat of 98°C and then heated for 20 min. Next, the sections were permeabilized with Triton X-100, blocked with bovine serum albumin, and stained with primary antibody [anti-AQP8 (Abclonal, A8539), anti-SPIB (Abclonal, A7451), and anti-GUCA2B (Abclonal, A8390)] for 16 h. Anti-SPIB, anti-GUCA2B, and anti-AQP8 were diluted following the manufacturer's protocol. Sections were stained with HRP-linked secondary antibodies (Optiview DAB IHC Detection kit, Ventana Roche, Arizona, USA) for an additional 4 h.

RNA quantification: Total RNA was extracted with TRIzol reagent (Invitrogen, MA, USA) and reverse transcribed into cDNA using a high-capacity cDNA reverse transcription kit (Applied Biosystems, MA, USA) according to the manufacturer's instructions. The cDNA samples were quantified using qPCR BIO SyGreen Blue Mix Lo-ROX (PCR Biosystems, London, UK) on an ABI QuantStudio 5 Real-Time PCR System according to the manufacturer's instructions. The primers were (forward) CGGTCATTGAGATGGGACGG and (reverse) AGGAGCATCACCA-GGTTGAGG for *AQP8*; (forward) AGCACACAGTCAGTCTACATCC and (reverse) CACACAGCTCACAGTCGTCG for *GUCA2B*; (forward) GGCAGGACTCGCAAGAAG and (reverse) TCTTGGCGT-AGTTTCGGAGG for *SPIB*; and (forward) TTCCACCATGGAGAAGGC and (reverse) GATGCATGGACTGTGGTC for *GAPDH*. Data were normalized to *GAPDH* expression levels in each sample using the $\Delta\Delta\text{Ct}$ method.

Co-regulatory network of target genes

Cell culture and transfection: HCT116 was purchased from the Bioresource Collection and

Research Center (Hsinchu, Taiwan) within 3 years and cultured in high-glucose DMEM (Gibco, TX, USA) following the standard protocol. The plasmids carrying the open reading frame for the candidate genes were synthesized by OriGene (MD, USA), amplified in *Escherichia coli* DH5 α , and extracted using a plasmid mini kit (Geneaid, Taipei City, Taiwan) following the manufacturer's instructions. Cells were seeded in six-well plates and transfected with 1 μ g of plasmid per well using Lipofectamine 2000 (Thermo Fisher Scientific, MA, USA) to overexpress the candidate genes according to the manufacturer's instructions. Serum-free medium was used to avoid quenching of the plasmids or transfection reagent by serum proteins. The medium was replaced with a regular culture medium after 6 h and refreshed daily. The transfected cells were incubated under standard culture conditions for 2 d.

Western blotting: Cells or colorectal cancer tissues were dissociated using RIPA lysis buffer supplemented with a proteinase inhibitor cocktail (Thermo Fisher Scientific) or a cComplete™, EDTA-free Protease Inhibitor Cocktail (Roche) according to the manufacturer's instructions. The proteins were quantified using Qubit, and the samples were analyzed by western blotting. Primary antibodies, including anti-AQP8 (Abclonal, A8539), anti-SPIB (Abclonal, A7451), and anti-GUCA2B (Abclonal, A8390), were used. For anti-AQP8 and anti-SPIB, the primary antibodies were diluted 1,000-fold with PBS supplemented with 0.1% Tween 20 (PBST), while for anti-GUCA2B, they were diluted 500-fold with TBS supplemented with 0.1% Tween 20 (TBST). The internal control antibody, anti-GAPDH, was diluted 10,000-fold with PBST for anti-AQP8 and anti-SPIB, and with TBST for anti-GUCA2B. Secondary antibodies, including goat anti-rabbit IgG-conjugated HRP and goat anti-mouse IgG-conjugated HRP, were diluted 2,000-fold with PBST for anti-AQP8 and anti-SPIB, and 10,000-fold with TBST for anti-GUCA2B. HRP-conjugated antibodies were then reacted with an ECL substrate (LF08-500, Visual Protein, Taipei City, Taiwan, or Immobilon Western Chemiluminescent HRP Substrate, Millipore), and the target proteins were detected using either the ChemiDoc Imaging System (Bio-Rad, CA, USA) or the Millipore GE Healthcare Life Sciences system (USA).

microRNA-transcription factor-mRNA functional analysis: Transcription factors and miRNAs influence the expression of target gene expression, and the influence of one regulator affects the impact of the other on the shared target gene expression in CRC [21]. Understanding the microRNA (miRNA)-transcription factor-mRNA regulatory network is essential for understanding its biological function [22]. The miRNet 2.0 website (freely available at <https://www.mirnet.ca>) was used to analyze the regulatory networks of the TGs, miRNAs, and Transcription factors. MiRNet's database was derived from miRBase, miRTarBase, TarBase, HMDD, and others. Human tissue-specific miRNA annotations are derived from the TSMiR and IMOTA databases, whereas human exosomal miRNA annotations are obtained from ExoCarta. TransmiR 2.0, ENCODE, JASPAR, and ChEA provide information on the interactions between miRNAs, TFs, and genes [23].

Results

Identification of colorectal cancers

Higher expression levels of *GUCA2B* and *SPIB* were significantly associated with lower stages using the COADREAD dataset. The expression levels of *AQP8* and *SPIB* were higher in women than in men (**Table 1**). No significant association was found with pathological T/N/M stage, lymphatic invasion, microsatellite stability status, presurgical therapy, or tumor site.

SPIB, *AQP8*, *GUCA2B* expressed significantly higher mRNA levels in normal mucosa compared with tumors using COAD dataset (**Figure 1A**). Their expression decreased from normal mucosa, adenoma, adenocarcinoma to metastasis in order using integrated dataset (**Figure 1B**). *AQP8*, *GUCA2B*, and *SPIB* mRNA expression could be used to differentiate normal mucosa, CRC and metastatic tumors using t-SNE (**Figure 1C**). In addition, *AQP8*, *GUCA2B*, and *SPIB* were independent genetic predictors for identifying CRC and NM under the adjustment for any of the clinical and demographic characteristics of tumor sites (colon/rectum), radiation therapy (yes/no), sex, microsatellite instability, lymphatic invasion (yes/no), pathological TNM, and stage (**Table S1**). The heatmap of *AQP8*, *GUCA2B*, and *SPIB* mRNA expression is shown in **Figure 1D**.

AQP8, GUCA2B, SPIB and CRC

Table 1. Association of target genes with clinical and demographic characteristics using the COAD-READ dataset and univariable linear regression

	AQP8 (Mn ± SD)	GUCA2B (Mn ± SD)	SPIB (Mn ± SD)
Gender	*		*
Female	5.73 (4.41)	3.41 (3.69)	5.02 (2.59)
Male	4.87 (4.07)	2.91 (3.29)	4.48 (2.48)
Tumor site			
Colon	5.27 (4.39)	3.16 (3.54)	4.82 (2.62)
Rectum	5.24 (3.75)	3.05 (3.27)	4.42 (2.28)
Stage		*	*
I-II	5.72 (4.83)	3.75 (4.14)	5.22 (2.83)
III-IV	5.05 (3.86)	2.82 (3.00)	4.41 (2.33)
Pathological T stage			
T1-2	5.22 (4.08)	2.88 (3.35)	4.79 (2.39)
T3-4	5.29 (4.31)	3.22 (3.53)	4.71 (2.58)
Pathological N stage			
N0	5.33 (4.22)	3.20 (3.50)	4.78 (2.49)
N1-2	5.13 (4.43)	2.98 (3.48)	4.53 (2.84)
Pathological M stage			
M0	5.20 (4.34)	3.12 (3.53)	4.78 (2.52)
M1	6.05 (4.14)	3.62 (3.61)	4.60 (2.83)
Lymphatic invasion			
No	5.15 (4.28)	3.01 (3.50)	4.70 (2.51)
Yes	5.65 (4.32)	3.44 (3.46)	4.95 (2.55)
Microsatellite stability status			
Microsatellite stability (MSS) and microsatellite instable-low (MSI-L)	5.35 (3.99)	3.11 (3.38)	4.63 (2.50)
Microsatellite instable-high (MSI-H)	4.26 (5.13)	2.96 (3.86)	5.27 (2.70)
Pre-surgical therapy			
No	5.19 (4.07)	3.09 (3.34)	4.81 (2.43)
Yes	5.58 (3.74)	3.53 (3.42)	4.09 (1.77)

*Significant variables are marked with an asterisk ($P < 0.05$). mRNA gene expression was log₂ base transformed. Mn, mean; SD, standard deviation.

We identified 22 well-known genes related to early-stage CRC [24-27], the phenotypic classification of nonpolyposis and polyposis [28], and biomarkers for early detection [29, 30] using literature review. The CRC classification performance of *SPIB*, *AQP8*, and *GUCA2B* was then compared to 22 well-known genes using COADREAD dataset. The heatmap is shown in **Figure 2A**. *SPIB*, *AQP8*, and *GUCA2B* expression was significantly associated with the CRC tissue types compared with other 22 genes.

The classification performance of *GUCA2B*, *SPIB*, and *AQP8* outperformed the 22 well-known genes with AUCs of 1.00, 0.99, and 0.98, respectively, using the TCGA COADREAD

dataset (**Figure 2B**). *GUCA2B* exhibited the best classification efficacy. The best genetic prediction model consisted of *SPIB* and *GUCA2B* in the logistic prediction model $\text{logit}(p/1-p) = 22.31 - 0.91 \cdot \log_2(\text{GUCA2B mRNA expression} + 1) - 1.80 \cdot \log_2(\text{SPIB mRNA expression} + 1)$ with the cut-off point of 0.5. The AUC of this genetic model was 0.9993 (sensitivity: 0.99, specificity: 1.00) (**Figure 2C**). The average accuracy of training and testing sets with 100 bootstrap replicates was 0.99 (**Figure 2D**).

Prognosis prediction

The prognostic prediction of CRC was evaluated in terms of RFS and OS using the COADREAD dataset and Cox proportional hazard regres-

AQP8, GUCA2B, SPIB and CRC

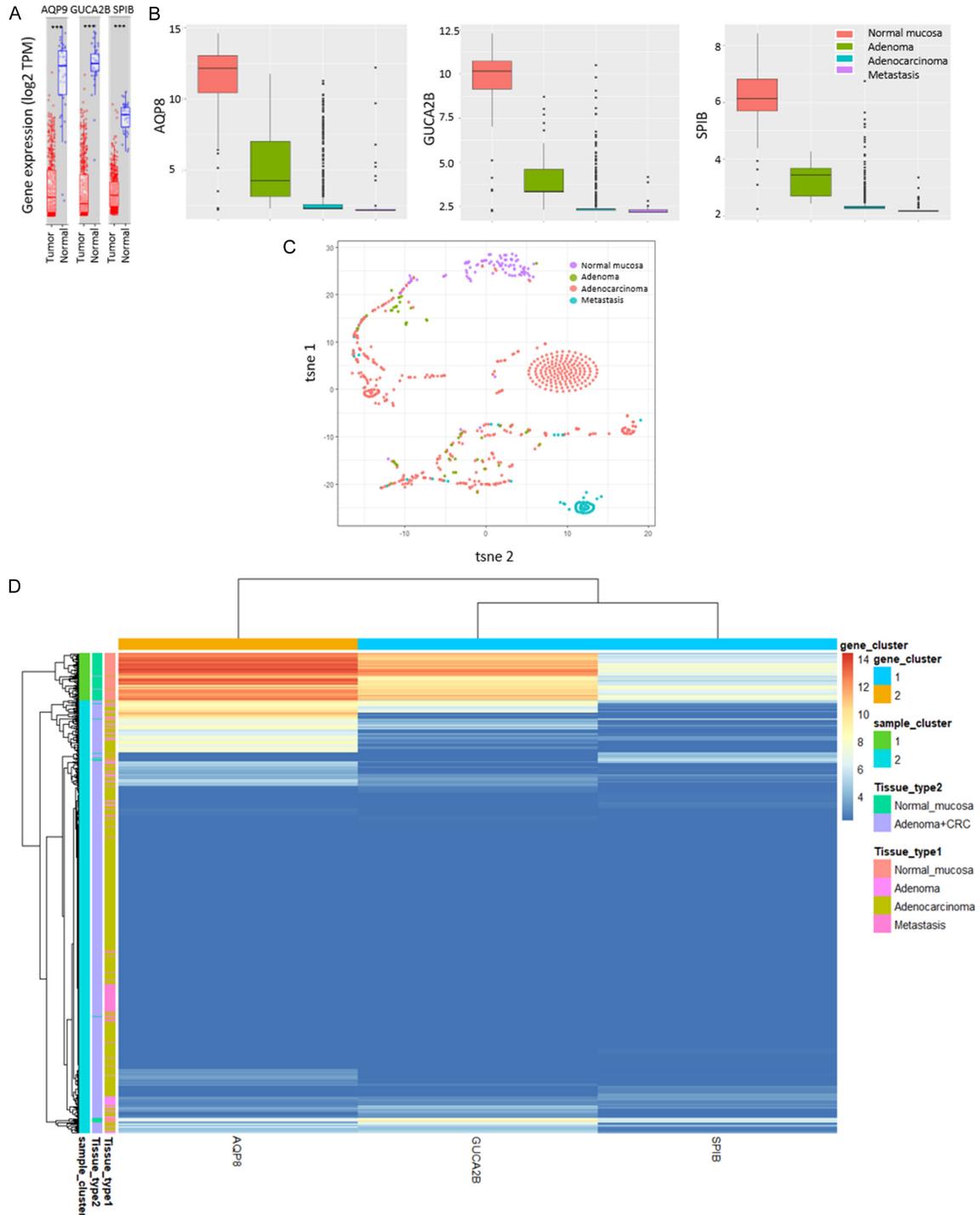


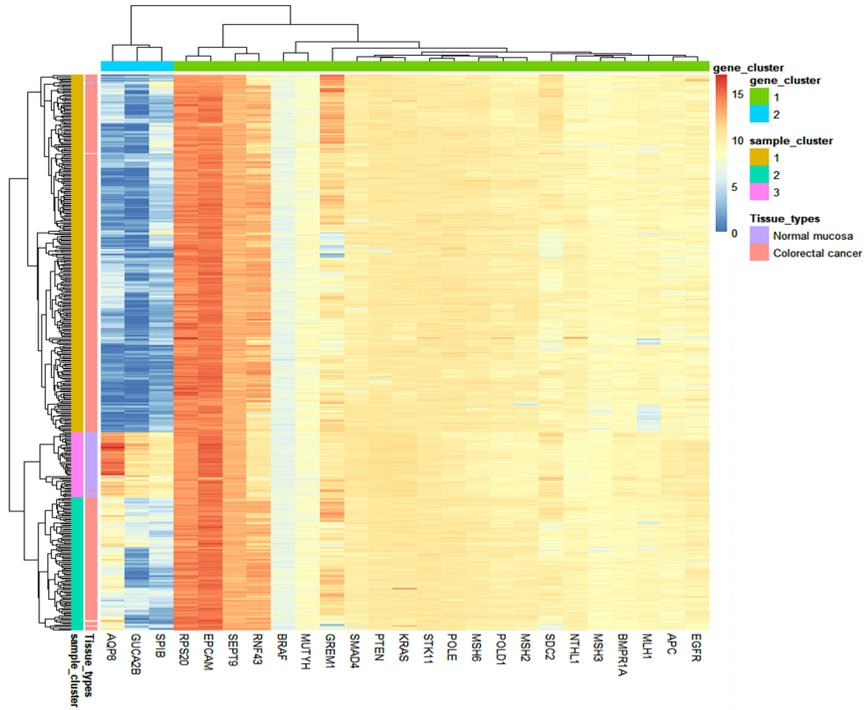
Figure 1. (A) *AQP8*, *GUCA2B*, and *SPIB* mRNA expression in colon cancer using the TCGA COAD dataset and the TIMER2.0 web tool (<https://cistrome.shinyapps.io/timer/>). (B) t-SNE plot visualizing assignment of tissue types using *AQP8*, *GUCA2B*, and *SPIB* gene expression. (D) Log2 mRNA expression of *AQP8*, *GUCA2B*, and *SPIB* in normal mucosa, adenoma, adenocarcinoma and colorectal cancer. The dataset of (B-D) was from integrated mRNA dataset (normal mucosa n = 88, adenoma n = 53, adenocarcinoma n = 521, metastatic tumors n = 79).

sion. *AQP8*, *GUCA2B* and *SPIB* were not statistically associated with OS and RFS in the univariate and multivariate models adjusted in terms of the stage ([Table S2](#)). However, the high

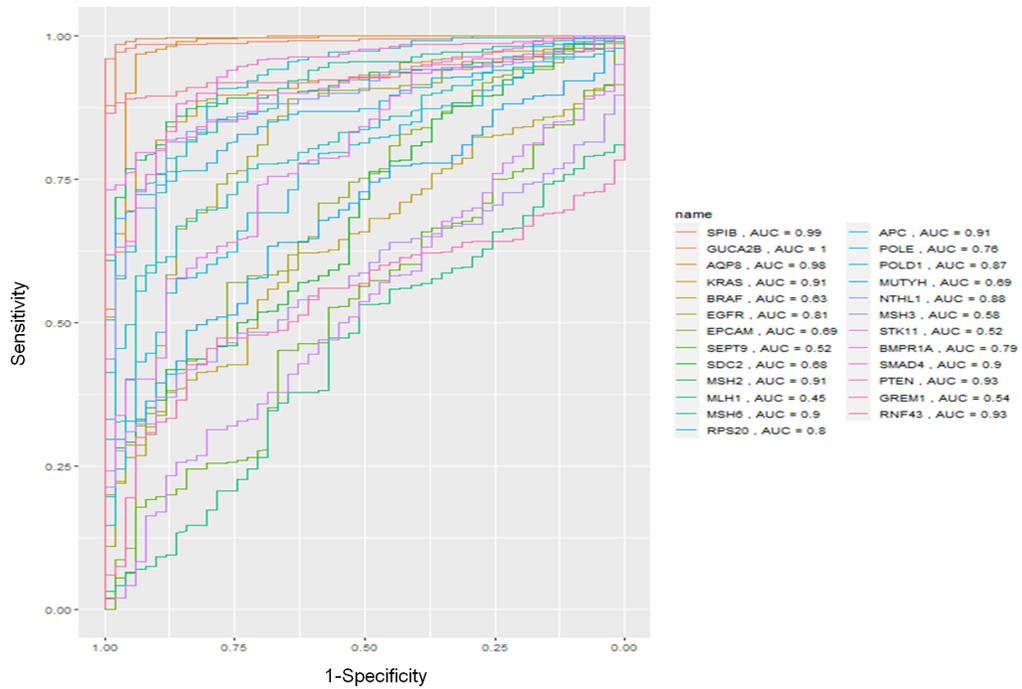
expression levels of *AQP8* (hazard ratio [HR] = 0.91, $P < 0.05$) and *GUCA2B* (HR = 0.87, $P < 0.05$) were associated with favorable RFS in the multivariable models adjusted regarding

AQP8, GUCA2B, SPIB and CRC

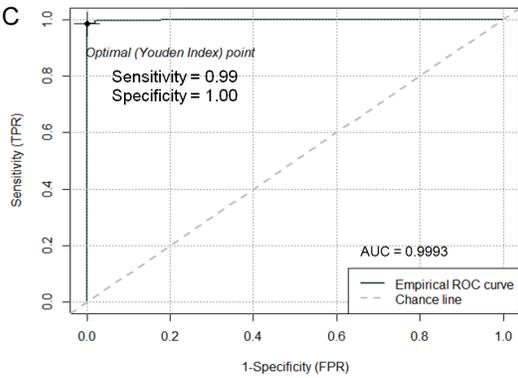
A



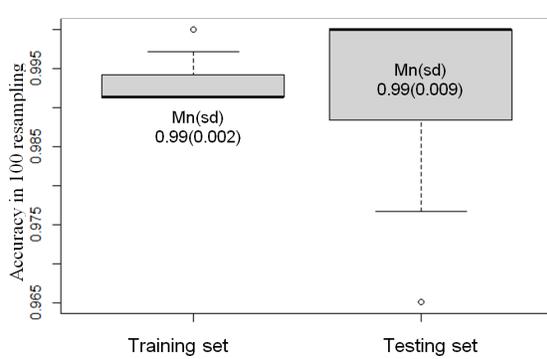
B



C



D



AQP8, GUCA2B, SPIB and CRC

Figure 2. A. Heatmap of the target genes (AQP8, GUCA2B, and SPIB) and known genes of the early stage, detection, and phenotypic classification of colorectal cancers. B. Receiver operating characteristic curves and area under the curves (AUC) of the target genes (AQP8, GUCA2B, and SPIB) and known genes for the early stage, detection, and phenotypic classification of colorectal cancers. C. We split the COADREAD dataset into a training set (80%) and a testing set (20%) and resampled 100 times to evaluate the prediction model of normal mucosa and primary CRCs using stepwise logistic regression. The best genetic prediction model was $\text{logit}(p/1-p) = 22.31 - 0.91 \cdot \log_2(\text{GUCA2B mRNA expression} + 1) - 1.80 \cdot \log_2(\text{SPIB mRNA expression} + 1)$ at the cut-off point of 0.5. The mean (Mn) and standard deviation (sd) of prediction accuracies are displayed in the boxplot. The average accuracies of the training and testing set were 0.99. D. The ROC curve of the genetic prediction model demonstrated perfect prediction efficacy with an AUC of 0.9993, sensitivity of 0.99, and specificity of 1.00. The dataset was from the COADREAD.

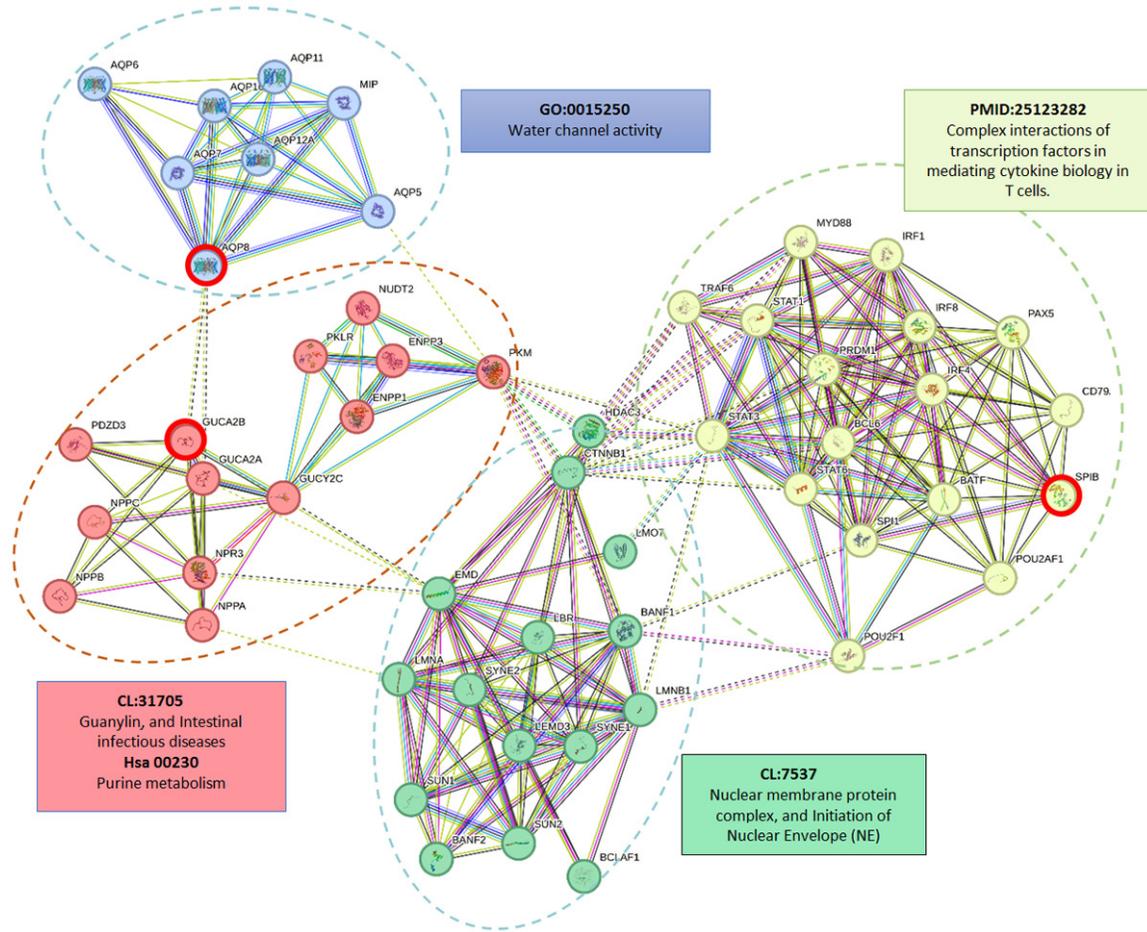


Figure 3. Protein-protein interaction using STRING. K-means clustering was used to divide the network into three clusters. The input genes are circled in red. The plot showed the annotations that were the most significant and insightful. The sources of pathway annotation are addressed by the prefixes PMID (Pubmed article), CL (String database), GO (Gene-ontology), and Has (KEGG).

pathological M stage. High *SPIB* expression was related to favorable OS in the multivariable model adjusted regarding presurgical therapy.

A co-regulatory network in vitro

We analyzed the protein-protein interaction of *AQP8*, *GUCA2B* and *SPIB* using STRING database. The result is shown in **Figure 3**. *GUCA2B*

and *AQP8* interacted directly but indirectly with *SPIB*. The interaction networks were clustered into several subnetworks for *AQP8*, *GUCA2B* and *SPIB*. The corresponding functional annotation are water channel activity for *AQP8*; guanylin and intestinal infectious diseases, and purine metabolism for *GUCA2B*; complex interactions of transcription factors in mediating cytokine biology in T cells for *SPIB*.

AQP8, GUCA2B, SPIB and CRC

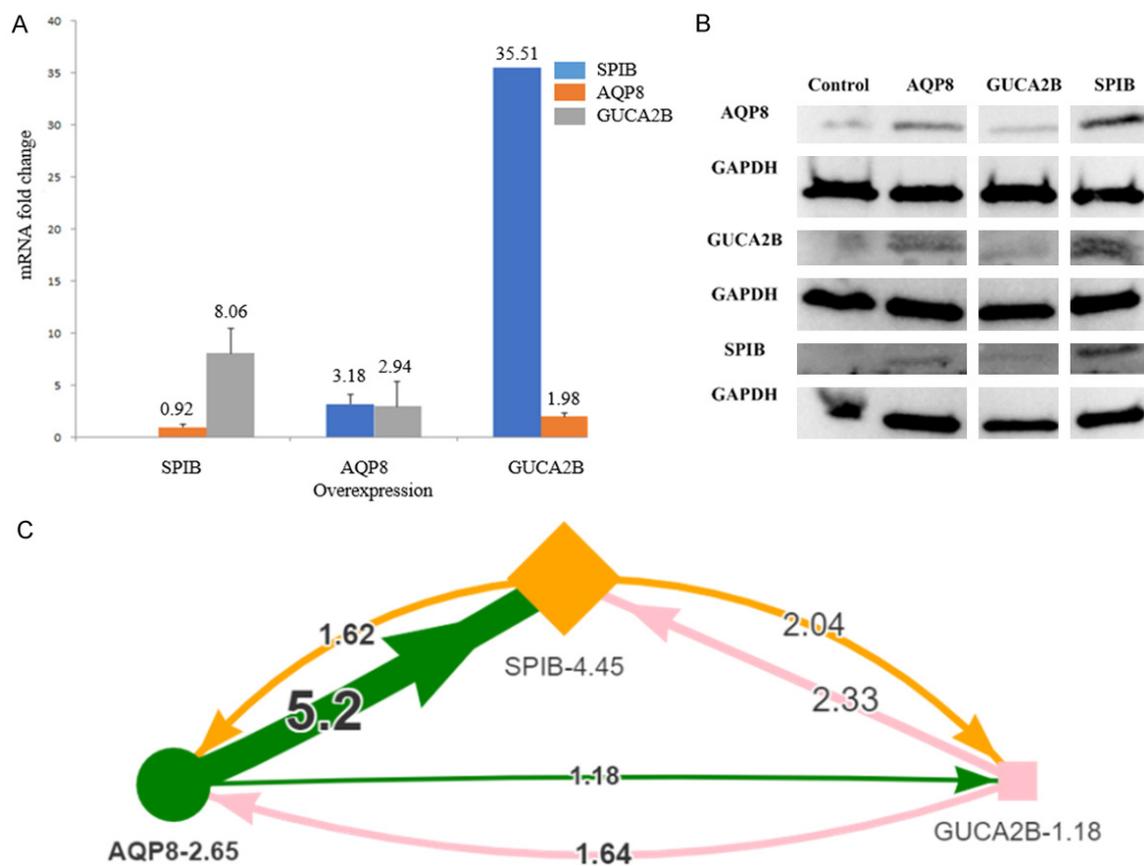


Figure 4. QPCR (A) and western blots (B) of the overexpression of one gene and the consequent expression of the other genes in HCT-116 cell lines. (C) The co-regulatory network of *AQP8*, *GUCA2B*, and *SPIB* was drawn using the quantitative data from the western blot. Figures shown in the plots are the fold changes for gene expression. The results of overexpressed *AQP8*, *GUCA2B*, and *SPIB* are marked in green, pink, and orange, respectively. The figure beside each node is the self-overexpression fold change. The baseline protein expression levels of *AQP8*, *GUCA2B*, and *SPIB* are 6,412, 12,279, and 6,884, respectively.

SPIB, *AQP8*, and *GUCA2B* gene expression were extremely low in CRC tumors. As a result, we used overexpression experiments to determine the co-regulatory network of *SPIB*, *AQP8*, and *GUCA2B*. In HCT116 cells, we overexpressed one and measured the levels of protein expression of the others. Although we overexpressed one gene, the expressions of other genes were simultaneously upregulated based on qPCR (Figure 4A) and western blot (Figure 4B). The overexpression of *SPIB* increased *GUCA2B* expression 8-fold, the overexpression of *AQP8* increased *SPIB* and *GUCA2B* expression by approximately 3-fold, and the overexpression of *GUCA2B* increased *AQP8* expression by 2-fold and *SPIB* expression by 36-fold at the mRNA levels (Figure 4A). The co-regulatory network was drawn using quantitative data from the western blot analysis (Figure 4C). These genes were also co-upregulated at the protein levels.

Transcription factors bind to DNA motifs and alter nearby transcription. Transcription factors were causally responsible for the observed transcriptional changes [31, 32]. *SPIB* is one of the ETS Transcription factors essential in regulating the immune system [33]. Its overexpression increased the protein levels of *AQP8* and *GUCA2B* by 1.62- and 2.04-fold, respectively. When *AQP8*/*GUCA2B* was overexpressed, *SPIB* expression increased by 5.2-/2.33-fold, respectively. The most significant change was apparent in the regulation of *AQP8* on *SPIB* (5.2-fold increase). *AQP8* and *GUCA2B* were positively co-regulated (Figure 4C).

Validation in clinical tissue samples

The expression levels of target proteins of *AQP8*, *GUCA2B*, and *SPIB* were extremely low in colonic adenocarcinoma tissues compared with those in non-neoplastic colon tissues

(**Figure 5A**). SPIB, AQP8, and GUCA2B mRNA expression levels in clinical samples from paired CRC tumors and adjacent NM were examined. AQP8, GUCA2B, and SPIB were barely expressed in tumor tissues compared with that seen in NM, except for GUCA2B in S67 and SPIB in S67/S68 (**Figure 5B**). In the western plot of fresh tissues, the protein expression of SPIB and AQP8 were relatively higher in tumors compared with adjacent NM. However, GUCA2B has the opposite finding (**Figure 5C**).

MiRNA-transcription factor-target gene regulatory network

miRNAs are non-coding RNAs that regulate the expression of TGs and play a role in the occurrence and development of cancers [34]. miRNAs have a strong potential for use as oncological biomarkers for CRC. Direct non-invasive detection of circulating miRNAs would provide information for the diagnosis, prognosis, and predictive treatment responses in CRC patients [35]. Therefore, we used miRNet 2.0 to plot the miRNA-TF-TG regulatory network (**Figure 6**). SPIB was connected with GUCA2B by miR-27a-3p, and GUCA2B was connected with AQP8 by miR-182-5p. The evidence suggests that miR-27a-3p and miR-182-5p may mediate the co-regulatory network of SPIB, GUCA2B, and AQP8.

Tumor microenvironment

We used the TIMER 2.0 web tool to assess the relationship between AQP8, GUCA2B, and SPIB and the tumor microenvironment; the results are shown in **Figure 7**. AQP8 and GUCA2B had similar association curve patterns with immune cell quantity. SPIB was found to be significantly associated with all immune markers, with tumor purity being negatively associated but the other immune cells being positively associated. SPIB and B cells were found to have a strong positive relationship.

Discussion

Aberrations in tumor suppressor genes play essential roles in carcinogenesis [6]. In this study, first, we validated that SPIB, AQP8, and GUCA2B were expressed at very low levels in CRC primary tumor tissues compared with that in adjacent normal mucosa. Second, we identified that SPIB, AQP8, and GUCA2B were in the

same co-regulatory network led by SPIB and mutually co-regulated *in vitro* at both mRNA and protein levels. Third, SPIB, AQP8, and GUCA2B were strong and independent genetic predictors for CRC identification, with considerable prediction efficacies AUCs close to 1. They all outperformed the well-known genetic biomarkers for early CRC included in this study. Fourth, SPIB, AQP8, and GUCA2B were associated with tumor microenvironment. Finally, the co-regulation of SPIB, AQP8, and GUCA2B may be moderated by miR-27a-3p and miR-182-5p. Taken together, in the tumorigenesis of CRC, we assumed that SPIB dysregulation resulted in complex interactions of transcription factors in mediating cytokine biology in T cells and transcription factor DNA binding dysregulation [36], which then cascaded into dysregulation of guanylin, and intestinal infectious diseases (related to GUCA2B) [12] and the balance of water channel activity (related to AQP8) in intestines.

AQP8, GUCA2B and SPIB were frequently co-expressed in the gene expression analysis of CRC and normal mucosa [5, 15, 16, 37]. The single-cell sequencing profile supported our assumed co-regulatory network of AQP8, GUCA2B, and SPIB. AQP8, GUCA2B, and SPIB were identified as significant genes for CRC identification by Zhang et al. [38]. Their findings of dysregulated pathways were enriched on cellular response to zinc ion, response to zinc ion, cellular response to cadmium ion, and digestion biological processes for down regulated transcripts in CRC epithelial cells. These were epithelial cell-specific functions that were disrupted in tumor tissues. These findings corroborated our findings on AQP8-protein networks involved in water channel activity. Furthermore, dysregulation of the receptor guanylyl cyclase signaling pathway (GUCA2B-protein networks) would result in intestine barrier breakdown, genomic instability, and abnormal metabolism [12].

The immune system's adaptive and innate arms can clearly work together to boost the anti-tumor response. Many studies had underpin the potential of harnessing the innate immune system and local immunological microenvironment to treat colorectal cancer [39]. SPIB was found to be significantly associated with tumor immune infiltration and immune checkpoint genes in over 35 tumors. In most

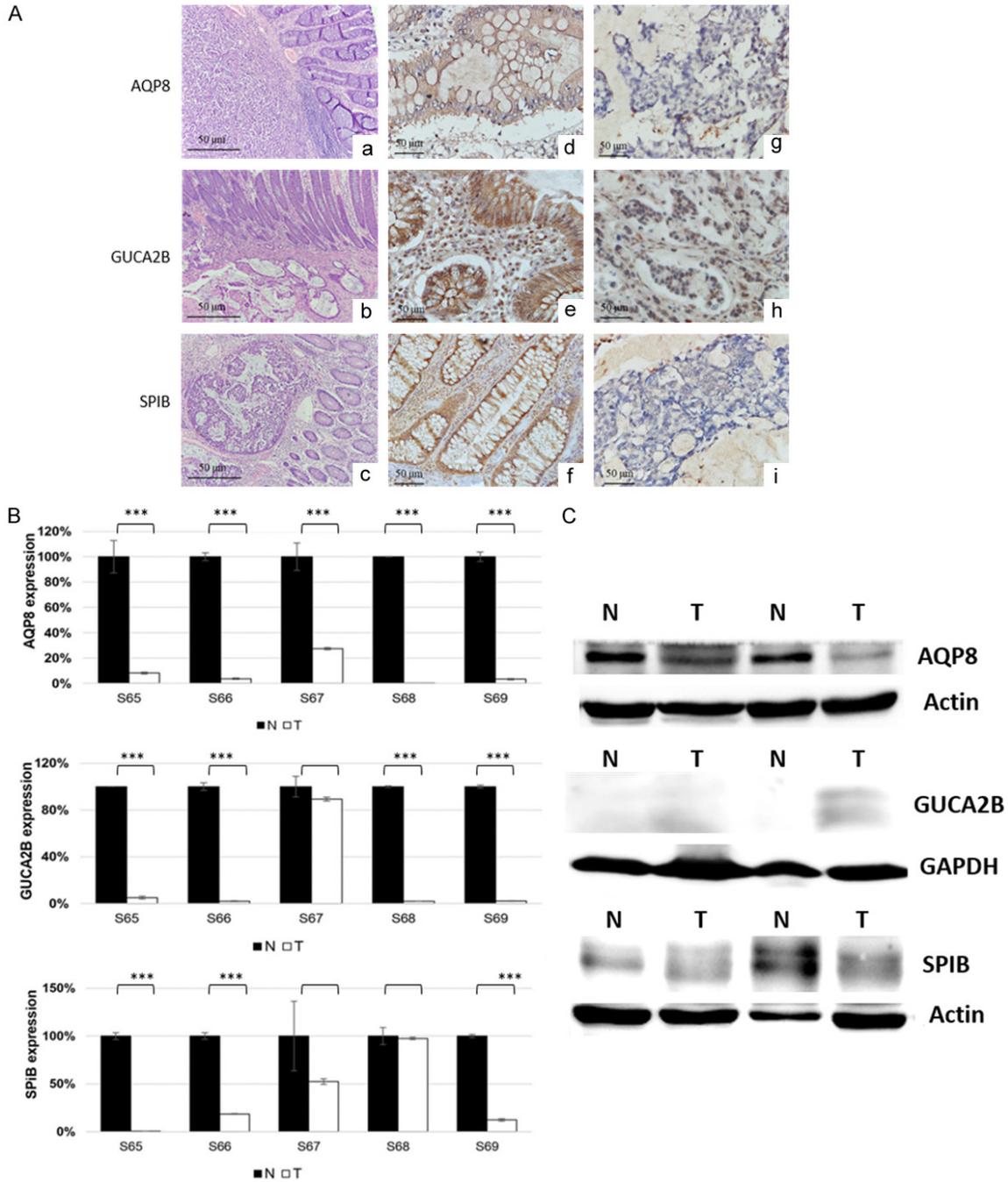


Figure 5. A. Hematoxylin and eosin (H & E) staining of neoplastic and non-neoplastic colon tissue (a-c, original magnification $\times 200$); immunohistochemical (IHC) analysis of AQP8 (d), GUCA2B (e), and SPIB (f) in non-neoplastic colon tissue; and IHC analysis of AQP8 (g), GUCA2B (h), and SPIB (i) in colonic adenocarcinoma tissue (original magnification $\times 400$). B. Relative mRNA expression levels of the genetic predictors in the clinical samples of colorectal cancer (T) and paired adjacent normal mucosa (N). AQP8, GUCA2B, and SPIB are suppressed in the tumor tissues. S65-69 is the ID number of the tissue samples. *** $P < 0.001$ using a signed test. C. The western plot of two paired fresh tissues from two patients of adjacent normal mucosa and tumors.

tumors, *SPIB* was found to be inversely related to tumor mutational burden and microsatellite instability. *SPIB* may be involved in NF-kappa B

and B-cell receptor signaling pathways [40]. Furthermore, *SPIB* is a member of the erythroblast transformation-specific transcription fac-

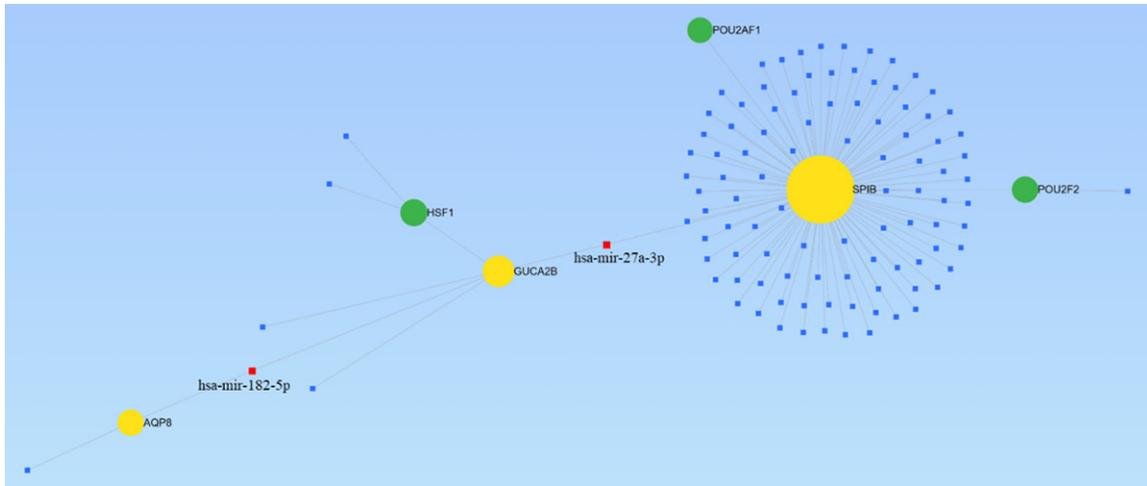


Figure 6. miRNA-transcription factor-target gene regulatory network using miRNet 2.0, a miRNA-centric network visual analytics platform (<https://www.mirnet.ca/>). Yellow circles indicate target genes, green circles transcription factors, squares miRNA, and red squares common miRNA between target genes.

tor family. It is a candidate master regulator of the differentiation of intestinal microfold cells, which initiate mucosal immune responses through the uptake and transcytosis of luminal antigens [41]. *SPIB* is also a tumor suppressor in CRC cells through the NF- κ B and JNK signaling pathways. In CRC, it inhibits cell proliferation, motility, and invasion, prevents angiogenesis, induces cell cycle arrest in the G2/M phase, and promotes cell apoptosis [42]. A number of AQP isoforms were found upregulated in inflammatory conditions and are considered essential for the migration and survival of immune cells. The downregulation of AQP3 and AQP8 was accompanied by an increase in intestinal inflammation and injury, suggesting that both AQP3 and AQP8 may be involved in the pathogenesis of inflammatory bowel disease [43, 44]. A model of 5-fluorouracil (5-FU)-induced diarrhea in mice showed increased pro-inflammatory cytokines (TNF- α , IL-1 β , IL-6, IL-17A and IL-22) correlating with decreased AQP4 and AQP8 mRNA throughout the entire colon compared to control mice [45]. miRNAs are non-coding RNAs that regulate the expression of target genes and play a role in the occurrence and development of cancers. The integrated analysis of miRNA and mRNA has facilitated the identification of potential biomarkers of CRC [34]. In addition, miRNAs are notable diagnostic biomarkers and therapeutic targets [46-53] that can be detected in blood [52, 54, 55] and are regulated by microRNA sponges, such as circular RNAs [56]. Moreover, a panel

combining serum CA19-9 and peripheral blood mononuclear cells miR-27a-3p level could have considerable clinical value in diagnosing pancreatic cancer [55]; miR-182-5p and miR-375-3p in blood plasma are better than prostate-specific antigens for discriminating prostate cancer from benign prostate hyperplasia [57].

MiR-27a-3p [51, 53, 56, 58-64] and miR-182-5p [48, 52, 54, 57, 65-76] have been studied extensively in cancers this decade. MiR-27a-3p was associated with the oncogenesis or progression of gastric, cervical, breast, non-small cell lung, and esophageal cancers. In terms of CRC, miR-27a-3p is a diagnostic, prognostic, and potential therapeutic biomarker [51, 64]. MiR-182-5p can discriminate prostate cancer [57], increase tamoxifen sensitivity in breast cancer [65], and promote glioblastoma angiogenesis [54, 77]. In CRC, miR-182-5p mediates cell proliferation, migration, and invasion via the Tiam1/Rac1/p38 MAPK axis [66], inhibits proliferation and metastasis by targeting MTDH [78], accelerates CRC progression via E2F4-induced AGAP2-AS1 expression upregulation [59], and inhibits tumorigenesis, angiogenesis, and lymphangiogenesis by directly downregulating VEGF-C expression [71]. Furthermore, miR-27a-3p has been found to promote immune evasion in breast cancer by increasing the expression of PD-L1, a protein associated with immune checkpoint inhibition. In the case of lung adenocarcinoma and obesity, miR-27a-3p has been shown to inhibit ICOS(+) T cell prolif-

AQP8, GUCA2B, SPIB and CRC

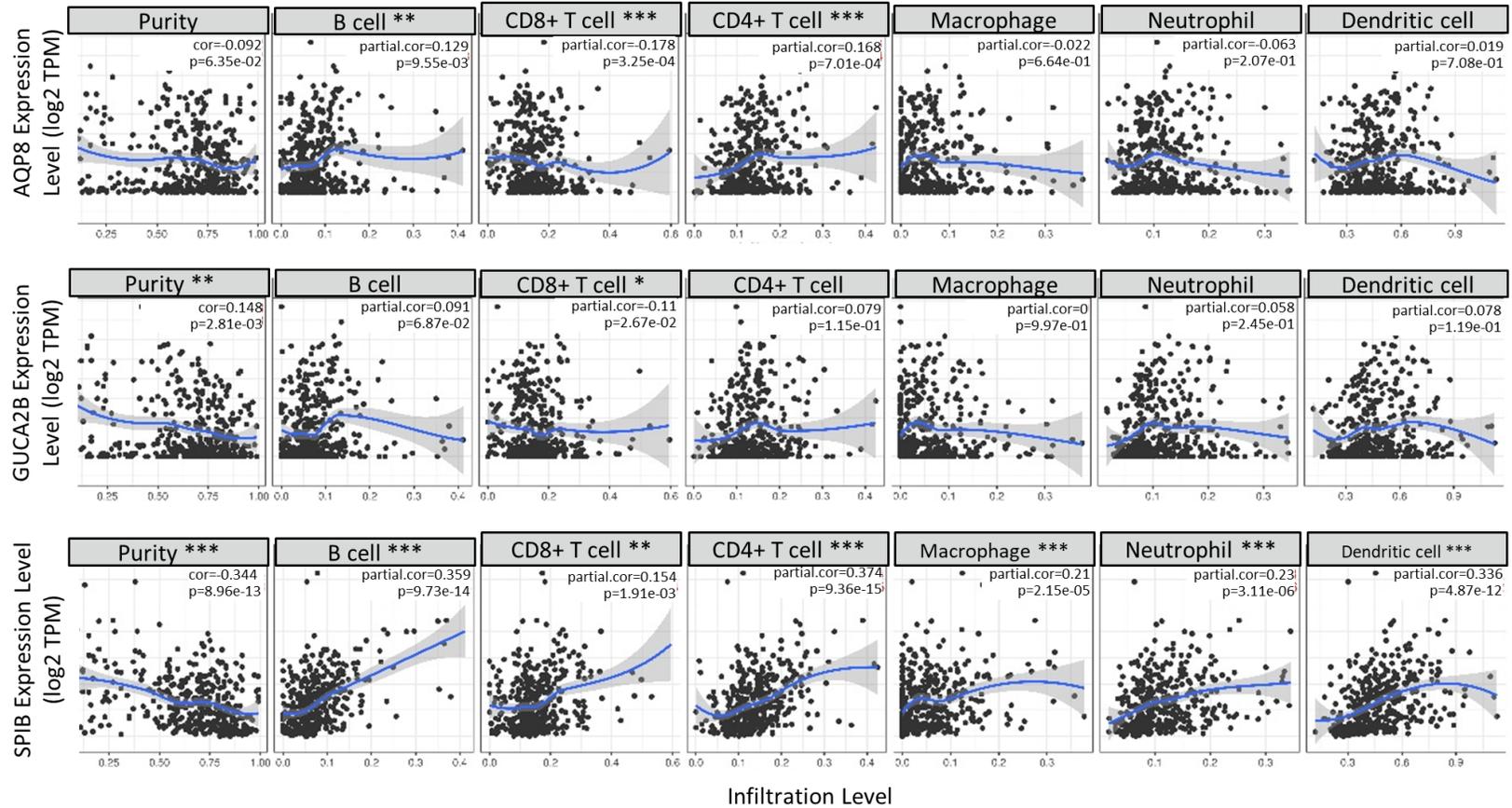


Figure 7. MRNA expression of AQP8, GUCA2B and SPIB with immune infiltration markers using TCGA COAD dataset and TIMER 2.0. Tumor purity is defined as the proportion of cancer cells in tumor tissue that reflects tumor microenvironmental characteristics.

eration and interferon-gamma secretion, which could explain why immunotherapy is more effective in obese patients [62, 79]. This finding highlights the role of miR-27a-3p in immune response modulation and shed light on potential therapeutic targets for cancer treatment.

GUCA2B is a physiological regulator of intestinal fluid and electrolyte transport. It is a secreted protein specific to colon tissues and plays a role as a tumor suppressor gene in CRC [80] in blood and urine [81]. Guanylate cyclase activator 2A (*GUCA2A*) and *GUCA2B* are endogenous hormones that bind to and activate the transmembrane receptor *GUCY2C* to mediate and orchestrate intestinal homeostatic mechanisms [13]. Therefore, *GUCA2B* is a new paradigm for CRC prevention via hormone replacement therapy involving synthetic hormone analogs [13]. However, in this study, the protein expression patterns of *GUCA2B* in fresh tissues were found heterogeneous in the study, contradicting to low expression in tumor tissues. It suggests that *GUCA2B* may have multiple roles in different parts of the intestine and at different stages of cancer.

AQP8 is a water channel transporter expressed primarily at the apical surface of enterocytes facing the lumen of the normal colonic mucosa [82]. It is a marker of normal proliferating colonic epithelial cells [83]. It is mainly expressed in paraneoplastic normal tissues and is barely expressed in colorectal carcinoma cells [84]. Its downregulation serves as an early driver of CRC tumorigenesis and persists until tumor formation [82, 84]. *AQP8* restrains CRC cell proliferation, migration, and invasion capacities [11] by downregulating PI3K/AKT signaling [10].

The divergence between mRNA and protein expression is a notable observation, signifying that the coordinated control of *AQP8*, *GUCA2B*, and *SPIB* is governed by complex factors. In the context of cancer, this disparity is a multifaceted phenomenon shaped by various factors including post-transcriptional processes, translation machinery alterations, and protein degradation disruptions. N6-methyladenosine (m6A) modification, driven by *METTL3*, boosts mRNA translation and has oncogenic implications, potentially contributing to the mRNA-protein disparity [85]. Dysregulation of m6A modification, particularly in the 3' untranslated region, is linked to cancer. Abnormal expression of translation factors like eIF4A and RNA

helicases can also affect protein synthesis, impacting the mRNA-protein relationship [86]. Dysfunctions in protein degradation pathways, including proteasome and autophagy, play roles in cancer and can worsen the disparity [87]. Further research is essential to fully grasp this phenomenon's mechanisms and its implications for cancer. Interestingly, genes with differentially expressed mRNA in an ovarian cancer xenograft model exhibit stronger correlations between mRNA and protein levels, underlining the biological significance of mRNA changes [88].

There are some limitations in this study, the gene expression of *AQP8*, *GUCA2B* and *SPIB* were extremely low and barely detected that make gene knock-down of target genes experiments hard to operate. In future work, knock-out experiments are demanded to get gain more evidence. The relatively modest clinical sample size may limit the generalizability of our findings to broader CRC patients. We discovered the inconsistent mRNA and protein expression patterns among target genes inconsistency that may implicate valuable regulation information. More patients' samples were essential to figure out the co-regulatory pattern of three genes. However, due to abovementioned limitations, we did the best to validate our hypothesis with multiple public datasets, analysis methods and in vitro experiments.

Conclusion

SPIB, *AQP8*, and *GUCA2B* are powerful and independent predictors of early CRC detection. They have been shown to function in a co-regulatory network. The major functions were transcription regulation, water channel balance, guanylin regulation, and intestinal infectious diseases. MiR-27a-3p and miR-182-5p are two possible mediators. The mechanisms of *SPIB*, *AQP8*, *GUCA2B*, miR-182-5p, and miR-27a-3p in CRC merit further investigation.

Acknowledgements

This work was supported by grants from the Tri-Service General Hospital (No. TSGH-E-110211).

Written informed consent was obtained from all patients.

Disclosure of conflict of interest

None.

Address correspondence to: Dr. Yu-Tien Chang, School of Public Health, National Defense Medical Center, No. 161, Section 6, Minquan East Road, Neihu District, Taipei 11490, Taiwan. Tel: +886-2-87923100 Ext. 18454; E-mail: greengarden720-925@gmail.com

References

- [1] Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi RE and Corcione F. Worldwide burden of colorectal cancer: a review. *Updates Surg* 2016; 68: 7-11.
- [2] Hamm A, Prenen H, Van Delm W, Di Matteo M, Wenes M, Delamarre E, Schmidt T, Weitz J, Sarmiento R, Dezi A, Gasparini G, Rothe F, Schmitz R, D'Hoore A, Iserentant H, Hendlisz A and Mazzone M. Tumour-educated circulating monocytes are powerful candidate biomarkers for diagnosis and disease follow-up of colorectal cancer. *Gut* 2016; 65: 990-1000.
- [3] Sisik A, Kaya M, Bas G, Basak F and Alimoglu O. CEA and CA 19-9 are still valuable markers for the prognosis of colorectal and gastric cancer patients. *Asian Pac J Cancer Prev* 2013; 14: 4289-4294.
- [4] Duffy MJ, van Dalen A, Haglund C, Hansson L, Klapdor R, Lamerz R, Nilsson O, Sturgeon C and Topolcan O. Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. *Eur J Cancer* 2003; 39: 718-727.
- [5] Chu CM, Yao CT, Chang YT, Chou HL, Chou YC, Chen KH, Terng HJ, Huang CS, Lee CC, Su SL, Liu YC, Lin FG, Wetter T and Chang CW. Gene expression profiling of colorectal tumors and normal mucosa by microarrays meta-analysis using prediction analysis of microarray, artificial neural network, classification, and regression trees. *Dis Markers* 2014; 2014: 634123.
- [6] Joyce C, Rayi A and Kasi A. Tumor-suppressor genes. *StatPearls*. Treasure Island (FL): 2021.
- [7] Zhang L and Shay JW. Multiple roles of APC and its therapeutic implications in colorectal cancer. *J Natl Cancer Inst* 2017; 109: djw332.
- [8] Qi L and Ding Y. Screening of tumor suppressor genes in metastatic colorectal cancer. *Biomed Res Int* 2017; 2017: 2769140.
- [9] Armaghany T, Wilson JD, Chu Q and Mills G. Genetic alterations in colorectal cancer. *Gastrointest Cancer Res* 2012; 5: 19-27.
- [10] Wu Q, Yang ZF, Wang KJ, Feng XY, Lv ZJ, Li Y and Jian ZX. AQP8 inhibits colorectal cancer growth and metastasis by down-regulating PI3K/AKT signaling and PCDH7 expression. *Am J Cancer Res* 2018; 8: 266-279.
- [11] Zhang H, Du WB, Guo XM, Wang LK, Cheng JM and Wei LJ. Identification of the AQP8-miR-92a network associated with the aggressive traits of colorectal cancer. *Biochem Biophys Res Commun* 2020; 527: 218-225.
- [12] Nomiri S, Hoshyar R, Chamani E, Rezaei Z, Salmani F, Larki P, Tavakoli T, Gholipour F, Tabrizi NJ, Derakhshani A, Santarpia M, Franchina T, Brunetti O, Silvestris N and Safarpour H. Prediction and validation of GUCA2B as the hub-gene in colorectal cancer based on co-expression network analysis: In-silico and in-vivo study. *Biomed Pharmacother* 2022; 147: 112691.
- [13] Pattison AM, Merlino DJ, Blomain ES and Waldman SA. Guanylyl cyclase C signaling axis and colon cancer prevention. *World J Gastroenterol* 2016; 22: 8070-8077.
- [14] Zhao X, Li L, Yuan S, Zhang Q, Jiang X and Luo T. SPIB acts as a tumor suppressor by activating the NFkB and JNK signaling pathways through MAP4K1 in colorectal cancer cells. *Cell Signal* 2021; 88: 110148.
- [15] Xu H, Ma Y, Zhang J, Gu J, Jing X, Lu S, Fu S and Huo J. Identification and verification of core genes in colorectal cancer. *Biomed Res Int* 2020; 2020: 8082697.
- [16] Laczmanska I, Sasiadek M and Laczmanski L. The comparison between molecular tumour profiling in microdissected and surgical tissue samples. *Anticancer Res* 2018; 38: 1415-1418.
- [17] Kanke M, Kennedy Ng MM, Connelly S, Singh M, Schaner M, Shanahan MT, Wolber EA, Beasley C, Lian G, Jain A, Long MD, Barnes EL, Herfarth HH, Isaacs KL, Hansen JJ, Kapadia M, Guillem JG, Feschotte C, Furey TS, Sheikh SZ and Sethupathy P. Single-cell analysis reveals unexpected cellular changes and transposon expression signatures in the colonic epithelium of treatment-naive adult Crohn's disease patients. *Cell Mol Gastroenterol Hepatol* 2022; 13: 1717-1740.
- [18] Burclaff J, Bliton RJ, Breau KA, Ok MT, Gomez-Martinez I, Ranek JS, Bhatt AP, Purvis JE, Woosley JT and Magness ST. A proximal-to-distal survey of healthy adult human small intestine and colon epithelium by single-cell transcriptomics. *Cell Mol Gastroenterol Hepatol* 2022; 13: 1554-1589.
- [19] Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC and Muller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011; 12: 77.
- [20] Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, Francois R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Muller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K and Yutani H. Welcome to the tidyverse. *J Open Source Softw* 2019; 4: 1686.

- [21] Mullany LE, Herrick JS, Wolff RK, Stevens JR, Samowitz W and Slattery ML. MicroRNA-transcription factor interactions and their combined effect on target gene expression in colon cancer cases. *Genes Chromosomes Cancer* 2018; 57: 192-202.
- [22] Martinez-Sanchez A and Murphy CL. MicroRNA target identification-experimental approaches. *Biology (Basel)* 2013; 2: 189-205.
- [23] Chang L, Zhou G, Soufan O and Xia J. miRNet 2.0: network-based visual analytics for miRNA functional analysis and systems biology. *Nucleic Acids Res* 2020; 48: W244-W251.
- [24] Watson R, Liu TC and Ruzinova MB. High frequency of KRAS mutation in early onset colorectal adenocarcinoma: implications for pathogenesis. *Hum Pathol* 2016; 56: 163-170.
- [25] Fernandez-Rozadilla C, Brea-Fernandez A, Bessa X, Alvarez-Urturi C, Abuli A, Clofent J, Paya A; EPICOLON Consortium, Jover R, Xicola R, Llor X, Andreu M, Castells A, Carracedo A, Castellvibel S and Ruiz-Ponte C. BMPR1A mutations in early-onset colorectal cancer with mismatch repair proficiency. *Clin Genet* 2013; 84: 94-96.
- [26] Yantiss RK, Goodarzi M, Zhou XK, Rennert H, Pirog EC, Banner BF and Chen YT. Clinical, pathologic, and molecular features of early-onset colorectal carcinoma. *Am J Surg Pathol* 2009; 33: 572-582.
- [27] Ngeow J, Heald B, Rybicki LA, Orloff MS, Chen JL, Liu X, Yerian L, Willis J, Lehtonen HJ, Lehtonen R, Mester JL, Moline J, Burke CA, Church J, Aaltonen LA and Eng C. Prevalence of germline PTEN, BMPR1A, SMAD4, STK11, and ENG mutations in patients with moderate-load colorectal polyps. *Gastroenterology* 2013; 144: 1402-1409, 1409, e1-5.
- [28] Valle L, Vilar E, Tavtigian SV and Stoffel EM. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. *J Pathol* 2019; 247: 574-588.
- [29] Han YD, Oh TJ, Chung TH, Jang HW, Kim YN, An S and Kim NK. Early detection of colorectal cancer based on presence of methylated syndecan-2 (SDC2) in stool DNA. *Clin Epigenetics* 2019; 11: 51.
- [30] Nicolazzo C, Raimondi C, Francescangeli F, Ceccarelli S, Trenta P, Magri V, Marchese C, Zeuner A, Gradilone A and Gazzaniga P. EpCAM-expressing circulating tumor cells in colorectal cancer. *Int J Biol Markers* 2017; 32: e415-e420.
- [31] Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR and Weirauch MT. The human transcription factors. *Cell* 2018; 172: 650-665.
- [32] Rubin JD, Stanley JT, Sigauke RF, Levandowski CB, Maas ZL, Westfall J, Taatjes DJ and Dowell RD. Transcription factor enrichment analysis (TFEA) quantifies the activity of multiple transcription factors from a single experiment. *Commun Biol* 2021; 4: 661.
- [33] Gallant S and Gilkeson G. ETS transcription factors and regulation of immunity. *Arch Immunol Ther Exp (Warsz)* 2006; 54: 149-163.
- [34] Zhu J, Xu Y, Liu S, Qiao L, Sun J and Zhao Q. MicroRNAs associated with colon cancer: new potential prognostic markers and targets for therapy. *Front Bioeng Biotechnol* 2020; 8: 176.
- [35] Chen B, Xia Z, Deng YN, Yang Y, Zhang P, Zhu H, Xu N and Liang S. Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis. *Open Biol* 2019; 9: 180212.
- [36] Li P, Spolski R, Liao W and Leonard WJ. Complex interactions of transcription factors in mediating cytokine biology in T cells. *Immunol Rev* 2014; 261: 141-156.
- [37] Shangkuang WC, Lin HC, Chang YT, Jian CE, Fan HC, Chen KH, Liu YF, Hsu HM, Chou HL, Yao CT, Chu CM, Su SL and Chang CW. Risk analysis of colorectal cancer incidence by gene expression analysis. *PeerJ* 2017; 5: e3003.
- [38] Zhang GL, Pan LL, Huang T and Wang JH. The transcriptome difference between colorectal tumor and normal tissues revealed by single-cell sequencing. *J Cancer* 2019; 10: 5883-5890.
- [39] Kather JN and Halama N. Harnessing the innate immune system and local immunological microenvironment to treat colorectal cancer. *Br J Cancer* 2019; 120: 871-882.
- [40] Ding M, Li Q, Tan X, Zhang L, Tan J and Zheng L. Comprehensive pan-cancer analysis reveals the prognostic value and immunological role of SPIB. *Aging (Albany NY)* 2022; 14: 6338-6357.
- [41] Kanaya T, Hase K, Takahashi D, Fukuda S, Hoshino K, Sasaki I, Hemmi H, Knoop KA, Kumar N, Sato M, Katsuno T, Yokosuka O, Toyooka K, Nakai K, Sakamoto A, Kitahara Y, Jinnohara T, McSorley SJ, Kaisho T, Williams IR and Ohno H. The Ets transcription factor Spi-B is essential for the differentiation of intestinal microfold cells. *Nat Immunol* 2012; 13: 729-736.
- [42] Zhao X, Li L, Yuan S, Zhang Q, Jiang X and Luo T. SPIB acts as a tumor suppressor by activating the NFkB and JNK signaling pathways through MAP4K1 in colorectal cancer cells. *Cell Signal* 2021; 88: 110148.
- [43] Zhao G, Li J, Wang J, Shen X and Sun J. Aquaporin 3 and 8 are down-regulated in TNBS-induced rat colitis. *Biochem Biophys Res Commun* 2014; 443: 161-166.
- [44] Sakai H, Sagara A, Matsumoto K, Hasegawa S, Sato K, Nishizaki M, Shoji T, Horie S, Nakagawa T, Tokuyama S and Narita M. 5-Fluorouracil induces diarrhea with changes in the expression of inflammatory cytokines and aquapo-

- rins in mouse intestines. *PLoS One* 2013; 8: e54788.
- [45] da Silva IV and Soveral G. Aquaporins in immune cells and inflammation: new targets for drug development. *Int J Mol Sci* 2021; 22: 1845.
- [46] Li YK, Hsu HM, Lin MC, Chang CW, Chu CM, Chang YJ, Yu JC, Chen CT, Jian CE, Sun CA, Chen KH, Kuo MH, Cheng CS, Chang YT, Wu YS, Wu HY, Yang YT, Lin C, Lin HC, Hu JM and Chang YT. Genetic co-expression networks contribute to creating predictive model and exploring novel biomarkers for the prognosis of breast cancer. *Sci Rep* 2021; 11: 7268.
- [47] Lin MY, Chang YC, Wang SY, Yang MH, Chang CH, Hsiao M, Kitsis RN and Lee YJ. OncomiR miR-182-5p Enhances Radiosensitivity by Inhibiting the radiation-induced antioxidant effect through SESN2 in head and neck cancer. *Antioxidants (Basel)* 2021; 10: 1808.
- [48] Lu JT, Tan CC, Wu XR, He R, Zhang X, Wang QS, Li XQ, Zhang R and Feng YM. FOXF2 deficiency accelerates the visceral metastasis of basal-like breast cancer by unrestrictedly increasing TGF-beta and miR-182-5p. *Cell Death Differ* 2020; 27: 2973-2987.
- [49] Ma D, Zhu Y, Zhang X, Zhang J, Chen W, Chen X, Qian Y, Zhao Y, Hu T, Yao Z, Zhao W, Zhang Y and Liu F. Long non-coding RNA RUNDC3A-AS1 promotes lung metastasis of thyroid cancer via targeting the miR-182-5p/ADAM9. *Front Cell Dev Biol* 2021; 9: 650004.
- [50] Seo HA, Moeng S, Sim S, Kuh HJ, Choi SY and Park JK. MicroRNA-based combinatorial cancer therapy: effects of MicroRNAs on the efficacy of anti-cancer therapies. *Cells* 2019; 9: 29.
- [51] Su C, Huang DP, Liu JW, Liu WY and Cao YO. miR-27a-3p regulates proliferation and apoptosis of colon cancer cells by potentially targeting BTG1. *Oncol Lett* 2019; 18: 2825-2834.
- [52] Wang W, Hu W, Wang Y, An Y, Song L, Shang P and Yue Z. Long non-coding RNA UCA1 promotes malignant phenotypes of renal cancer cells by modulating the miR-182-5p/DLL4 axis as a ceRNA. *Mol Cancer* 2020; 19: 18.
- [53] Wang Z, Wang W, Zhao W, Wang Z, Yang J, Wang W, Teng P, Su X, Li D, Zhang X, Wang H and Hao M. Folate inhibits miR-27a-3p expression during cervical carcinoma progression and oncogenic activity in human cervical cancer cells. *Biomed Pharmacother* 2020; 122: 109654.
- [54] Li J, Yuan H, Xu H, Zhao H and Xiong N. Hypoxic cancer-secreted exosomal miR-182-5p promotes glioblastoma angiogenesis by targeting Kruppel-like factor 2 and 4. *Mol Cancer Res* 2020; 18: 1218-1231.
- [55] Wang WS, Liu LX, Li GP, Chen Y, Li CY, Jin DY and Wang XL. Combined serum CA19-9 and miR-27a-3p in peripheral blood mononuclear cells to diagnose pancreatic cancer. *Cancer Prev Res (Phila)* 2013; 6: 331-338.
- [56] Zheng F, Wang M, Li Y, Huang C, Tao D, Xie F, Zhang H, Sun J, Zhang C, Gu C, Wang Z and Jiang G. CircNR3C1 inhibits proliferation of bladder cancer cells by sponging miR-27a-3p and downregulating cyclin D1 expression. *Cancer Lett* 2019; 460: 139-151.
- [57] Abramovic I, Vrhovec B, Skara L, Vrtaric A, Nikolac Gabaj N, Kulis T, Stimac G, Ljiljak D, Ruzic B, Kastelan Z, Kruslin B, Bulic-Jakus F, Ulamec M, Katusic-Bojanac A and Sincic N. MiR-182-5p and miR-375-3p have higher performance than PSA in discriminating prostate cancer from benign prostate hyperplasia. *Cancers (Basel)* 2021; 13: 2068.
- [58] Lu X, Kang N, Ling X, Pan M, Du W and Gao S. MiR-27a-3p promotes non-small cell lung cancer through SLC7A11-mediated-ferroptosis. *Front Oncol* 2021; 11: 759346.
- [59] Li W, Xin X, Li X, Geng J and Sun Y. Exosomes secreted by M2 macrophages promote cancer stemness of hepatocellular carcinoma via the miR-27a-3p/TXNIP pathways. *Int Immunopharmacol* 2021; 101: 107585.
- [60] Guo X and Li M. LINC01089 is a tumor-suppressive lncRNA in gastric cancer and it regulates miR-27a-3p/TET1 axis. *Cancer Cell Int* 2020; 20: 507.
- [61] Li S, Han Y, Liang X and Zhao M. LINC01089 inhibits the progression of cervical cancer via inhibiting miR-27a-3p and increasing BTG2. *J Gene Med* 2021; 23: e3280.
- [62] Yao X, Tu Y, Xu Y, Guo Y, Yao F and Zhang X. Endoplasmic reticulum stress-induced exosomal miR-27a-3p promotes immune escape in breast cancer via regulating PD-L1 expression in macrophages. *J Cell Mol Med* 2020; 24: 9560-9573.
- [63] Wang Z, Wang W, Zhao W, Wang Z, Yang J, Wang W, Teng P, Su X, Li D, Zhang X, Wang H and Hao M. Corrigendum to "Folate inhibits miR-27a-3p expression during cervical carcinoma progression and oncogenic activity in human cervical cancer cells" [Biomed. Pharmacother. 122 (2020) 109654]. *Biomed Pharmacother* 2020; 125: 109981.
- [64] Liang J, Tang J, Shi H, Li H, Zhen T, Duan J, Kang L, Zhang F, Dong Y and Han A. miR-27a-3p targeting RXRalpha promotes colorectal cancer progression by activating Wnt/beta-catenin pathway. *Oncotarget* 2017; 8: 82991-83008.
- [65] Sang Y, Chen B, Song X, Li Y, Liang Y, Han D, Zhang N, Zhang H, Liu Y, Chen T, Li C, Wang L, Zhao W and Yang Q. circRNA_0025202 regulates tamoxifen sensitivity and tumor progression via regulating the miR-182-5p/FOXO3a axis in breast cancer. *Mol Ther* 2021; 29: 3525-3527.

- [66] Li F, Zhou YD, Liu J, Cai JD, Liao ZM and Chen GQ. RBP-J promotes cell growth and metastasis through regulating miR-182-5p-mediated Tiam1/Rac1/p38 MAPK axis in colorectal cancer. *Cell Signal* 2021; 87: 110103.
- [67] Lu C, Zhao Y, Wang J, Shi W, Dong F, Xin Y, Zhao X and Liu C. Breast cancer cell-derived extracellular vesicles transfer miR-182-5p and promote breast carcinogenesis via the CMTM7/EGFR/AKT axis. *Mol Med* 2021; 27: 78.
- [68] Zheng J, Wu D, Wang L, Qu F, Cheng D and Liu X. miR-182-5p regulates cell growth of liver cancer via targeting RCAN1. *Gastroenterol Res Pract* 2021; 2021: 6691305.
- [69] Soheilifar MH, Vaseghi H, Seif F, Ariana M, Ghorbanifar S, Habibi N, Papari Barjasteh F and Pornour M. Concomitant overexpression of miR-182-5p and miR-182-3p raises the possibility of IL-17-producing Treg formation in breast cancer by targeting CD3d, ITK, FOXO1, and NFATs: a meta-analysis and experimental study. *Cancer Sci* 2021; 112: 589-603.
- [70] Fu S, Zhao N, Jing G, Yang X, Liu J, Zhen D and Tang X. Matrine induces papillary thyroid cancer cell apoptosis in vitro and suppresses tumor growth in vivo by downregulating miR-182-5p. *Biomed Pharmacother* 2020; 128: 110327.
- [71] Yan S, Wang H, Chen X, Liang C, Shang W, Wang L, Li J and Xu D. MiR-182-5p inhibits colon cancer tumorigenesis, angiogenesis, and lymphangiogenesis by directly downregulating VEGF-C. *Cancer Lett* 2020; 488: 18-26.
- [72] Li X, Yang B, Ren H, Xiao T, Zhang L, Li L, Li M, Wang X, Zhou H and Zhang W. Hsa_circ_0002483 inhibited the progression and enhanced the Taxol sensitivity of non-small cell lung cancer by targeting miR-182-5p. *Cell Death Dis* 2019; 10: 953.
- [73] Sang Y, Chen B, Song X, Li Y, Liang Y, Han D, Zhang N, Zhang H, Liu Y, Chen T, Li C, Wang L, Zhao W and Yang Q. circRNA_0025202 regulates tamoxifen sensitivity and tumor progression via regulating the miR-182-5p/FOXO3a axis in breast cancer. *Mol Ther* 2019; 27: 1638-1652.
- [74] Wang F, Wu D, Xu Z, Chen J, Zhang J, Li X, Chen S, He F, Xu J, Su L, Luo D, Zhang S and Wang W. miR-182-5p affects human bladder cancer cell proliferation, migration and invasion through regulating Cofilin 1. *Cancer Cell Int* 2019; 19: 42.
- [75] Xie F, Li Y, Wang M, Huang C, Tao D, Zheng F, Zhang H, Zeng F, Xiao X and Jiang G. Circular RNA BCRC-3 suppresses bladder cancer proliferation through miR-182-5p/p27 axis. *Mol Cancer* 2018; 17: 144.
- [76] Wang A, Jin C, Li H, Qin Q and Li L. LncRNA ADAMTS9-AS2 regulates ovarian cancer progression by targeting miR-182-5p/FOXF2 signaling pathway. *Int J Biol Macromol* 2018; 120: 1705-1713.
- [77] Wu X, Chen H, Wu M, Peng S and Zhang L. Downregulation of miR-182-5p inhibits the proliferation and invasion of triple-negative breast cancer cells through regulating TLR4/NF-kappaB pathway activity by targeting FBXW7. *Ann Transl Med* 2020; 8: 995.
- [78] Jin Y, Zhang ZL, Huang Y, Zhang KN and Xiong B. MiR-182-5p inhibited proliferation and metastasis of colorectal cancer by targeting MTDH. *Eur Rev Med Pharmacol Sci* 2019; 23: 1494-1501.
- [79] Fan X, Wang J, Qin T, Zhang Y, Liu W, Jiang K and Huang D. Exosome miR-27a-3p secreted from adipocytes targets ICOS to promote anti-tumor immunity in lung adenocarcinoma. *Thorac Cancer* 2020; 11: 1453-1464.
- [80] Feodorova Y, Tashkova D, Koev I, Todorov A, Kostov G, Simitchiev K, Belovejdov V, Dimov R and Sarafian V. Novel insights into transcriptional dysregulation in colorectal cancer. *Neoplasma* 2018; 65: 415-424.
- [81] Tsukahara H, Sekine K, Uchiyama M, Miura M, Nakazato M, Date Y, Tsunezawa W, Kotsuji F, Nishida K, Hiraoka M and Mayumi M. Uroguanylin level in umbilical cord blood. *Pediatr Int* 2001; 43: 267-269.
- [82] Hong Y, Liew SC, Thean LF, Tang CL and Cheah PY. Human colorectal cancer initiation is bidirectional, and cell growth, metabolic genes and transporter genes are early drivers of tumorigenesis. *Cancer Lett* 2018; 431: 213-218.
- [83] Fischer H, Stenling R, Rubio C and Lindblom A. Differential expression of aquaporin 8 in human colonic epithelial cells and colorectal tumors. *BMC Physiol* 2001; 1: 1.
- [84] Wang W, Li Q, Yang T, Bai G, Li D, Li Q and Sun H. Expression of AQP5 and AQP8 in human colorectal carcinoma and their clinical significance. *World J Surg Oncol* 2012; 10: 242.
- [85] Choe J, Lin S, Zhang W, Liu Q, Wang L, Ramirez-Moya J, Du P, Kim W, Tang S, Sliz P, Santisteban P, George RE, Richards WG, Wong KK, Locker N, Slack FJ and Gregory RI. mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. *Nature* 2018; 561: 556-560.
- [86] Heerma van Voss MR, van Diest PJ and Raman V. Targeting RNA helicases in cancer: the translation trap. *Biochim Biophys Acta Rev Cancer* 2017; 1868: 510-520.
- [87] Edinger AL and Thompson CB. Defective autophagy leads to cancer. *Cancer Cell* 2003; 4: 422-424.
- [88] Koussounadis A, Langdon SP, Um IH, Harrison DJ and Smith VA. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci Rep* 2015; 5: 10775.

AQP8, GUCA2B, SPIB and CRC

Table S1. Genetic predictors of colorectal cancer and normal mucosa classification in univariable and multivariable logistic regression models using the COADREAD dataset

Gene	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
	OR		OR		OR		OR		OR		OR		OR		OR		OR		OR	
AQP8	0.37	***	0.36	***	0.41	***	0.36	***	0.38	***	0.37	***	0.31	***	0.37	***	0.36	***	0.29	***
GUCA2B	0.20	***	0.20	***	0.24	***	0.19	***	0.20	***	0.18	***	0.03	*	0.20	***	0.16	***	0.08	***
SPIB	0.14	***	0.14	***	0.12	***	0.12	***	0.13	***	0.11	***	0.13	***	0.10	***	0.12	***	0.13	***

Model 1 is univariable logistic regression. Models 2-10 are multivariable logistic regression models under the control of one covariate (2. tumor primary site [colon/rectum], 3. radiation therapy [yes, no], 4. gender, 5. MSI [MSS/MSI], 6. lymphatic invasion [yes/no], 7. pathological M stage [0 and 1], 8. pathological N stage [0, 1, and 2], 9. pathological T stage [1, 2, 3, and 4] or 10. pathological stage [1, 2, 3, and 4]) in each model. ***P < 0.001 and *P < 0.05.

Table S2. Multivariable Cox proportional hazard regression of the genetic predictors of relapse-free survival and overall survival

Gene	RFS (n = 380)		OS (n = 380)	
	HR	P	HR	P
AQP8 ^a	0.91	*	0.97	0.43
GUCA2B ^a	0.87	*	0.94	0.32
SPIB ^b	0.93	0.32	0.85	*

^aUnder the adjustment of pathological M stage. ^bUnder the adjustment of presurgical therapy. RFS: relapse-free survival; OS: overall survival; *P < 0.05.