Scientific Research Publishing

Journal of Cancer Therapy, 2024, 15, 362-380
https://www.scirp.org/journal/jct
ISSN Online: 2151-1942
ISSN Print: 2151-1934

Identification of MUC1 as a Novel Oncogene of *Fusobacterium nucleatum*-Associated Colorectal Cancer by a Combined Bioinformatics and Experimental Approach

Xinli Ye[1*], Shouru Zhang[1], Zhaoli Zhang[2], Jie Zou[3], Xiaojie Gao[1#]

[1]Department of Gastrointestinal Tumor Center, Chongqing University Cancer Hospital, Chongqing, China
[2]Nursing Department, Chongqing Cancer Hospital, Chongqing, China
[3]Department of Hepatobiliary Surgery, Children's Hospital of Chongqing Medical University, Chongqing, China
Email: [#]1548461644@qq.com

How to cite this paper: Ye, X.L., Zhang, S.R., Zou, J., Gao, X.J. and Zhang, Z.l. (2024) Identification of MUC1 as a Novel Oncogene of *Fusobacterium nucleatum*-Associated Colorectal Cancer by a Combined Bioinformatics and Experimental Approach. *Journal of Cancer Therapy*, **15**, 362-380.
https://doi.org/10.4236/jct.2024.1510034

Received: September 29, 2024
Accepted: October 27, 2024
Published: October 30, 2024

Abstract

Background: *Fusobacterium nucleatum* can cause opportunistic and chronic infections and has recently been shown to be involved in colorectal cancer. However, the specific mechanism by which *F. nucleatum* induces colorectal carcinoma remains unclear. **Methods:** We downloaded the GSE110223, GSE110224, GSE113513 and GSE122183 microarray datasets from the Gene Expression Omnibus (GEO) database. Identification of differentially expressed genes (DEGs) related to *F. nucleatum* in CRC by overlapping data sets was performed. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genome pathway (KEGG) analyses were used for enrichment analysis. Moreover, Cytoscape software constructed a protein-protein interaction (PPI) network of differentially expressed genes. Finally, western blot and RT-qPCR analysis identified the relative protein and mRNA expression of hub genes in the cell model. **Results:** In total, 118 DEGs in *F. nucleatum*-associated CRC were screened from nonoverlapping microarray data, among which 20 upregulated and 98 downregulated DEGs were identified. The 118 DEGs were significantly correlated with diverse functions and pathways. The hub gene MUC1 had higher centrality scores in the PPI network, and the top 5 closely interacting hub genes, SLC7A11, AGR2, KRT18, CARTPT and TSPYL5, were identified. **Conclusion:** Our evidence suggests that the identified DEGs associated with *F. nucleatum* enhance our comprehension of the molecular

*First author.
[#]Corresponding authors.

Mechanisms underlying the tumorigenesis and development of CRC and might be used as molecular targets and diagnostic biomarkers for the treatment of CRC.

Keywords

CRC, *F. nucleatum*, DEGs, MUC1

1. Introduction

Colorectal carcinoma (CRC) is a major and severe health problem worldwide. Over 1.8 million new colorectal carcinoma cases and 881,000 deaths were estimated to occur in 2018, accounting for approximately 1/10 of cancer morbidities and mortalities. In general, colorectal carcinoma ranked third in incidence and second in mortality [1]. Recently, microorganisms have been increasingly closely associated with the occurrence, development and metastasis of tumors. There are many risk factors involved in CRC. *Fusobacterium nucleatum*, a gram-negative anaerobic bacillus that may induce colorectal cancer, has received increased attention [2] [3]. According to recent findings, *F. nucleatum* can induce a proinflammatory microenvironment in the bowel through antiphlogistic mechanisms and immune responses, thereby promoting a microecology propitious for tumorigenesis and CRC progression [4] [5]. However, the molecular biology and pathogenesis of CRC attributed to *F. nucleatum* infection needs to be further elucidated.

The etiology of colorectal cancer has always been considered to be complex due to the many genetic alterations in critical genes and pathways [6]. With related studies in the field of genomics, high-throughput technology for the analysis and quantification of gene expression, such as microarrays, is increasingly utilized [7]. Genomic analysis with microarray technology has shown potential in clinical cancer research and in fundamental oncology [8]. One of the most promising techniques is bioinformatics analysis with microarrays, which is an up-to-date, open, systematic, and multidimensional method and can comprehensively analyze the changes in genetic material, including mRNA expression, during the occurrence and development of CRC. Therefore, we could help elucidate the process of tumor occurrence and find novel prognostic biomarkers and cancer therapeutic targets.

The epithelial cell molecule mucin 1 (MUC1) is a transmembrane glycoprotein aberrantly overexpressed in various cancers, including CRC [9]. Accumulated experimental evidence has shown that the expression of MUC1 is increased in CRC. Moreover, MUC1 could be a novel therapeutic target in chronic diseases, such as inflammation, cancer and autoimmune diseases [10]. Previous studies have reported that MUC1 is a novel T cell regulator and is expressed by activated human T cells. MUC1 prevents the activation of the intrinsic apoptotic pathway, thus assisting cancer cells in evading cell death [11]. In hypoxia, elevated reactive oxygen

species (ROS) levels can activate apoptotic pathways. However, MUC1 overexpression plays a key role in decreasing intracellular ROS levels. Furthermore, MUC1 impedes hypoxic cell death in colon cancer by facilitating decreases in intracellular ROS concentration [12]. It has been shown that MUC1 is related to enhanced tumor initiation and progression [13]. Recently, studies have shown that approximately 15% of malignancies are associated with chronic microbial infection worldwide [14]. Nevertheless, few studies have investigated the effect of aberrant MUC1 expression associated with intestinal anaerobes, especially *F. nucleatum*, in human CRC.

In this study, we analyzed microarray data by bioinformatics, screened the DEGs of colorectal cancer caused by *F. nucleatum* infection from multiple datasets, and carried out functional, pathway enrichment and PPI network analysis on the aforementioned DEGs. The expression of MUC1 and its closely interacting hub genes in *F. nucleatum*-infected CRC cell lines were identified by experiments. Our research was designed to screen oncological targets and demonstrate the potential carcinogenic process of colorectal cancer caused by *F. nucleatum* infection.

2. Materials and Methods

2.1. Microarray Data

The gene expression profiles of GSE110223, GSE110224, GSE113513 and GSE122183 were downloaded from the GEO database. GSE110223 was based on the GPL96 platform and contained 26 samples: 13 CRC samples and 13 normal colon samples. GSE110224 was registered on the GPL570 platform. A total of 34 samples were divided into two groups and included 17 cancer tissue specimens and 17 normal controls. The GSE113513 platform was GPL15207, comprising 14 colorectal tumor samples and the same number of normal samples. The GSE122183 dataset was based on the GPL17586 platform, and it contained 25 samples (15 colorectal cancers, 10 normal controls).

2.2. Processing of Microarray Data

All of the sample information was divided into a tumor group and a normal group. Fold change (FC) values and adjusted *P*-values were calculated based on t-tests in the LIMMA package in RStudio software. Adjusted *P*-values < 0.01 and |logFC| > 2 were defined as the cut-off criteria and were adopted for DEGs. Subsequently, the heat map and volcano plot were created with the pheatmap, ggplot2, ggthemes and Cairo packages in RStudio.

2.3. Functional and Pathway Enrichment Analyses

For functional annotations of the DEGs, we then submitted the differentially expressed gene symbols to DAVID to conduct the gene ontology (GO) analyses. KOBAS 3.0 was used to conduct the Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. A *P*-value < 0.05 was set as the statistically significant threshold.

2.4. Construction of the PPI Network and Hub Gene Identification

The STRING database can predict interactions of proteins encoded by differentially expressed genes [15]. In the STRING online database, we screened genes with interaction scores ≥0.4 and established the PPI network. The visualized results of the PPI network were output with Cytoscape software (version 3.6.1), and the CytoHubba plug-in ranked the DEGs according to the ranking rules of degree and then screened the hub gene. Finally, the top 6 genes were identified in CytoHubba.

2.5. Bacterial Strains and Cell Lines

F. nucleatum subsp. (ATCC 25586) was purchased from ATCC [2]. *F. nucleatum* was cultured overnight at 37˚C in an anaerobic environment (5% CO_2, 25% H_2N_2, and 70% N_2) using a Hypoxia Physiological Cell Culture Workstation (Ruskinn Invivo 2 400, Baker, UK). The human normal colorectal epithelial cell line NCW460 and colorectal carcinoma cell lines (HCT116, SW620 and SW480) were supplied by the cell bank of the Chinese Academy of Sciences (CAS). It was cultured in DMEM (Gibco, USA) supplemented with 10% fetal bovine serum (FBS, Gibco, USA) at 36.5˚C in a humidified atmosphere with 5% CO_2. The cells were seeded in 6-well plates and then infected with *F. nucleatum* (MOI 1:100) for 4, 12 or 24 h [16].

2.6. Quantitative Real-Time PCR Analysis

RT-qPCR was carried out to confirm the results of the microarray assay. In summary, total RNA was extracted by TRIzol reagent (Invitrogen, USA) and reverse transcribed to cDNA by a PrimeScript™ RT reagent kit (TaKaRa, Tokyo, Japan). TB Green (TaKaRa, China) amplified and quantified complementary DNA (cDNA) in the real-time reaction system, and the reaction conditions were as follows: 30 s at 95˚C and 40 cycles of 5 s at 95˚C and 30 s at 60˚C. The relative expression of total mRNA was calculated by the $2^{-\Delta\Delta Ct}$ method, and GAPDH was used as the experimental internal control. We screened the mRNA of 6 hub genes on the basis of their PPI networks for validation, and the relevant primer sequences.

2.7. Western Blot Analysis

We extracted total cellular protein from *F. nucleatum*-treated cell lines by protein extraction solution (Beyotime, China). Equivalent amounts of protein were denatured in 5× SDS loading buffer, separated by 12% SDS-PAGE, and then transferred to a polyvinylidene difluoride (PVDF) membrane by constant voltage or current. The membranes were blocked in TBST with 5% BSA for 2 h at room temperature. After incubation with the primary antibody and secondary antibodies overnight at 4˚C and 2 h at room temperature, positive bands were developed by the BeyoECL Plus kit (Beyotime). The antibodies against MUC1 (ab45167) and GAPDH (ab8245) were obtained from Abcam Biotechnology.

2.8. Statistical Analysis

All experimental data in this study are presented as the mean ± standard deviation (SD). Data were analyzed in Student's t-test or post hoc tests for one-way ANOVA (GraphPad Prism 6.01, USA). A P-value < 0.05 was considered statistically significant.

3. Results

3.1. Identification of the Aberrant DEGs in CRC Associated with *F. nucleatum*

After we standardized all the original gene expression data, the significant DEGs related to CRC (GSE110223 contains 82, GSE110224 contains 132 and GSE113513 contains 199) were screened. In addition, for the expression microarray GSE122183, a total of 139 DEGs in *F. nucleatum*-associated CRC were obtained. Given the small sample size of the microarray, the 4 datasets were overlapped to remove the DEGs in CRC associated with non-*F. nucleatum*. As a result, the nonoverlapping area of the GSE122183 dataset is considered to represent the DEGs of *F. nucleatum*-associated CRC, and 118 target genes in the nonoverlapping area were screened and visualized by a Venn diagram (**Figure 1A**), which consisted of 20 upregulated and 98 downregulated genes, as presented in **Figure 1B**. As shown in **Figure 1C** and **Figure 1D**, 118 DEGs were further identified by a heat map and volcano plot.

3.2. GO and KEGG Enrichment Analysis for DEGs in CRC Associated with *F. nucleatum*

To analyze the key functional classification of DEGs, we performed GO analysis by DAVID. GO annotated and classified gene function primarily through BP (biological process), CC (cellular component) and MF (molecular function). GO enrichment analysis identified the BPs of the DEGs with significant enrichment in chemical synaptic transmission and regulation of cell proliferation, etc. (**Table 1**). The enriched DEGs in CC were mainly involved in integral components of the membrane, extracellular space and plasma membrane (**Table 1**). The most significantly enriched MFs were protein homodimerization activity and receptor binding (**Table 1**). The comprehensive results of the gene ontology enrichment analysis of the above differentially expressed genes are presented in **Figure 2**. At present, for genomic data analysis, high throughput data mining and many other omics research, KEGG has been proven to be a useful research tool and is widely used [17]. DEGs were analyzed in KEGG by KOBAS 3.0. These DEGs were mainly enriched in the following pathways: axon guidance (hsa04360), phospholipase D signaling pathway (hsa04072), calcium signaling pathway (hsa04020), histidine metabolism (hsa00340) and serotonergic synapse (hsa04726). These results are presented in **Table 2**, and KEGG enrichment analyses are visualized in **Figure 3**.

3.3. Target Gene Screening by Construction of the PPI Network

To predict the associations of protein functions of the identified DEGs, we

submitted 118 target genes in STRING. Subsequently, PPI network analysis was performed using Cytoscape 3.8.0 software. The screening of the PPI network module in Cytoscape depends on the cytoHubba plug-in. Furthermore, the construction of a PPI network based on hub genes with the highest connective degree was performed (Figure 4A). The hub genes in the network were ranked by the degree method. Finally, Muc1, which is overexpressed in CRC, was screened from the network, while 5 genes interact with it: SLC7A11, AGR2, KRT18, CARTPT and TSPYL5 (Figure 4B). There were 4 upregulated genes (MUC1, SLC7A11, AGR2, and KRT18) and 2 downregulated genes (CARTPT and TSPYL5) in this module. They were identified by the heat map shown in Figure 4C.

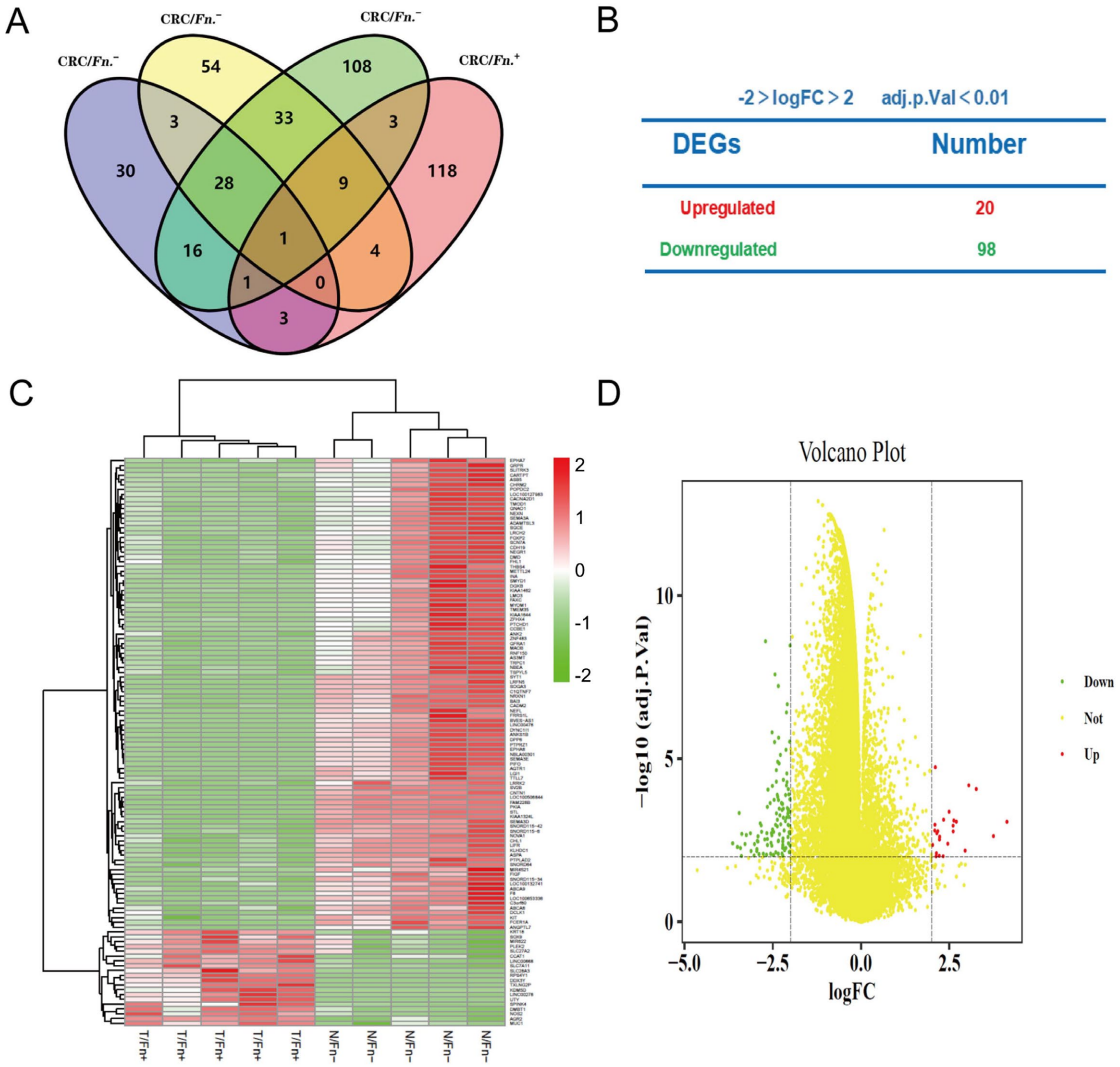

Figure 1. Identification and characterization of DEGs from the GSE110223, GSE110224, GSE113513 and GSE122183 datasets. (A) Venn diagram analysis of DEGs in CRC induced by *F. nucleatum* were obtained by overlapping 4 datasets. (B) The distribution of significant DEGs by *F. nucleatum*-induced CRC. (B) Heat map of the 118 DEGs; red indicates upregulation; conversely, green indicates downregulation, and gene expression with no difference is represented by white. (D) Volcano plot of DEGs by *F. nucleatum*-associated CRC. Red dots, significantly upregulated DEGs; green dots, DEGs; and yellow dots represent significantly upregulated, significantly downregulated and not significantly different DEGs. Significance was defined as adjusted *P*-value < 0.01 and |log FC| > 2.

Table 1. GO.

Category	Term/gene function	Gene Count	%	*P*-value	Gene symbol
BP	GO:0002027~regulation of heart rate	4	3.70	0.000633	ANK2, POPDC2, SEMA3A, DMD
BP	GO:0050919~negative chemotaxis	4	3.70	0.000692	SEMA3E, SEMA3D, SEMA3A, EPHA7
BP	GO:0048843~negative regulation of axon extension involved in axon guidance	3	2.77	0.007779	SEMA3E, EMA3D, SEMA3A
BP	GO:0071526~semaphorin-plexin signaling pathway	3	2.77	0.012349	SEMA3E, SEMA3D, SEMA3A
BP	GO:0042127~regulation of cell proliferation	5	4.62	0.015032	AGTR1, GRPR, KIT, NOS2, SOX9
BP	GO:0006936~muscle contraction	4	3.70	0.017568	GNAO1, SCN7A, MYOM1
BP	GO:0034613~cellular protein localization	3	2.77	0.020436	TMOD1
BP	GO:0001755~neural crest cell migration	3	2.77	0.022258	ANK2, DMD, LRRK2
BP	GO:0048812~neuron projection morphogenesis	3	2.77	0.025112	SEMA3E, SEMA3D, SEMA3A
BP	GO:0007268~chemical synaptic transmission	5	4.62	0.034794	DMD, LRRK2, NEFL, SYT1, CARTPT, NRXN1
BP	GO:0042297~vocal learning	2	1.85	0.03531	LRFN5, NOVA1
BP	GO:2000020~positive regulation of male gonad development	2	1.85	0.040253	NRXN1, FOXP2, SEMA3A, SOX9
BP	GO:0007411~axon guidance	4	3.70	0.048226	SEMA3A, NRXN1, LGI1, CHL1
CC	GO:0005886~plasma membrane	35	32.40	0.00158	SYT1, PTCHD1, CADM2, FHL1, NBEA, KIT, FRRS1L, AGTR1, DGKB, ANK2, DMD, GRPR, SV2B, SCN7A, LRFN5, DPP6, NEGR1, SLC28A3, FCER1A, TRPC1, CACNA2D1, GNAO1, F8, LIFR, NRXN1, SLC7A11, EPHA7, EPHA6, CHRM2, CDH19, CNTN1, GFRA1, SGCE, LRRK2, CHL1
CC	GO:0030425~dendrite	8	7.40	0.001633	EPHA7, GNAO1, CHRM2, SGCE, SEMA3A, TTLL7, LRRK2, CHL1
CC	GO:0005887~integral component of plasma membrane	16	14.81	0.005177	MUC1, FCER1A, TRPC1, PTPRZ1, CADM2, NRXN1, SLC7A11, AGTR1, EPHA7, EPHA6, CHRM2, GRPR, SGCE, DCLK1, SLC28A3
CC	GO:0036464~cytoplasmic ribonucleoprotein granule	3	2.77	0.007161	DYNC1I1, INA, DDX3Y
CC	GO:0005615~extracellular space	15	13.88	0.008195	INA, MUC1, SOGA3, F8, KIT, CCBE1, SEMA3E, SEMA3D, CARTPT, SEMA3A, LGI1, LRRK2, AGR2, DMBT1, THBS4
CC	GO:0030054~cell junction	8	7.40	0.009071	ANKS1B, SYT1, CHRM2, SV2B, NRXN1, LRRK2, LGI1, FRRS1L
CC	GO:0042383~sarcolemma	4	3.70	0.009398	ANK2, DMD, POPDC2, SGCE
CC	GO:0045202~synapse	5	4.62	0.013853	CHRM2, CADM2, DMD, LGI1, FRRS1L
CC	GO:0016021~integral component of mem-brane	37	34.25	0.019459	SYT1, ABCA9, PTCHD1, CADM2, UTY, C3ORF80, KIT, FRRS1L, ABCA6, AGTR1, SEMA3D, GRPR, SV2B, POPDC2, LRFN5, DPP6, FCER1A, MUC1, TRPC1, KIAA1324L, FAXC, SOGA3, PTPRZ1, MAOB, F8, LIFR, KIAA1644, NRXN1, SLC7A11, EPHA6, SLITRK3, RNF150, CDH19, SGCE, ASB5, SLC27A2, CHL1
CC	GO:0045211~postsynaptic membrane	5	4.62	0.022927	ANKS1B, EPHA7, ANK2, CHRM2, DMD
CC	GO:0030672~synaptic vesicle membrane	3	2.77	0.032117	SYT1, SV2B, LRRK2
CC	GO:0005883~neurofilament	2	1.85	0.040111	INA, NEFL
CC	GO:0005856~cytoskeleton	6	5.55	0.041366	ANK2, PLEK2, DMD, SGCE, NEXN, SLC7A11

Continued

CC	GO:0043209~myelin sheath	4	3.70	0.042846	INA, GNAO1, CNTN1, NEFL
MF	GO:0045499~chemorepellent activity	4	3.70	0.000237	EPHA7, SEMA3E, SEMA3D, SEMA3A
MF	GO:0038191~neuropilin binding	3	2.77	0.002022	SEMA3E, SEMA3D, SEMA3A
MF	GO:0030215~semaphorin receptor binding	3	2.77	0.00476	SEMA3E, SEMA3D, SEMA3A
MF	GO:0005200~structural constituent of cytoskeleton	4	3.70	0.013413	INA, ANK2, DMD, NEFL
MF	GO:0044325~ion channel binding	4	3.70	0.014412	TRPC1, ANK2, FHL1, LRRK2
MF	GO:0008307~structural constituent of muscle	3	2.77	0.015331	DMD, MYOM1, NEXN
MF	GO:0005102~receptor binding	6	5.55	0.021439	CADM2, GFRA1, NRXN1, NOS2, LGI1, SLC27A2
MF	GO:0048306~calcium-dependent protein binding	3	2.77	0.028108	SYT1, NRXN1, DMBT1
MF	GO:0002162~dystroglycan binding	2	1.85	0.039806	DMD, AGR2
MF	GO:0042803~protein homodimerization activity	8	7.40	0.045792	CADM2, MAOB, KIT, MYOM1, NOS2, LRRK2, AGR2, FOXP2

Table 2. KEGG.

Pathway ID	Pathway term	Gene count	*P*-value	Gene symbol
hsa04360	Axon guidance	6	0.00000242	SEMA3A, SEMA3E, SEMA3D, EPHA7 EPHA6, TRPC1
hsa04072	Phospholipase D signaling pathway	4	0.000268	DGKB, AGTR1, FCER1A, KIT
hsa04020	Calcium signaling pathway	4	0.000711	AGTR1, GRPR, NOS2, CHRM2
hsa00340	Histidine metabolism	2	0.001176	MAOB, ASPA
hsa04726	Serotonergic synapse	3	0.001808	MAOB, TRPC1, GNAO1
hsa04926	Relaxin signaling pathway	3	0.00254	FIGF, GNAO1, NOS2
hsa04514	Cell adhesion molecules (CAMs)	3	0.003498	NEGR1, CNTN1, NRXN1
hsa04261	Adrenergic signaling in cardiomyocytes	3	0.003699	SCN7A, AGTR1, CACNA2D1
hsa02010	ABC transporters	2	0.004118	ABCA6, ABCA9
hsa00330	Arginine and proline metabolism	2	0.005018	MAOB, NOS2
hsa04151	PI3K-Akt signaling pathway	4	0.006142	FIGF, THBS4, CHRM2, KIT
hsa05034	Alcoholism	3	0.006191	MAOB, PKIA, GNAO1
hsa04015	Rap1 signaling pathway	3	0.009368	FIGF, GNAO1, KIT
hsa05412	Arrhythmogenic right ventricular cardiomyopathy	2	0.011254	CACNA2D1, DMD
hsa04146	Peroxisome	2	0.012939	NOS2, SLC27A2
hsa05132	Salmonella infection	2	0.012939	DYNC1I1, NOS2
hsa04512	ECM-receptor interaction	2	0.01382	THBS4, SV2B
hsa05410	Hypertrophic cardiomyopathy (HCM)	2	0.015033	CACNA2D1, DMD
hsa05414	Dilated cardiomyopathy (DCM)	2	0.016936	CACNA2D1, DMD
hsa04933	AGE-RAGE signaling pathway in diabetic complications	2	0.018259	FIGF, AGTR1
hsa04916	Melanogenesis	2	0.018596	GNAO1, KIT
hsa05142	Chagas disease (American trypanosomiasis)	2	0.019279	GNAO1, NOS2
hsa04725	Cholinergic synapse	2	0.022478	GNAO1, CHRM2

Continued

hsa04010	MAPK signaling pathway	3	0.022786	FIGF, CACNA2D1, KIT
hsa05145	Toxoplasmosis	2	0.022846	GNAO1, NOS2
hsa04724	Glutamatergic synapse	2	0.023217	GNAO1, TRPC1
hsa05200	Pathways in cancer	4	0.02339	FIGF, NOS2, AGTR1, KIT
hsa04728	Dopaminergic synapse	2	0.029887	MAOB, GNAO1
hsa04080	Neuroactive ligand-receptor interaction	3	0.032141	GRPR, AGTR1, CHRM2
hsa04371	Apelin signaling pathway	2	0.032402	NOS2, AGTR1
hsa04915	Estrogen signaling pathway	2	0.032829	GNAO1, KRT18
hsa00360	Phenylalanine metabolism	1	0.035737	MAOB
hsa04145	Phagosome	2	0.039036	DYNC1I1, THBS4
hsa04921	Oxytocin signaling pathway	2	0.039495	GNAO1, CACNA2D1
hsa00220	Arginine biosynthesis	1	0.043506	NOS2
hsa04630	Jak-STAT signaling pathway	2	0.04372	FHL1, LIFR
hsa04614	Renin-angiotensin system	1	0.047366	AGTR1

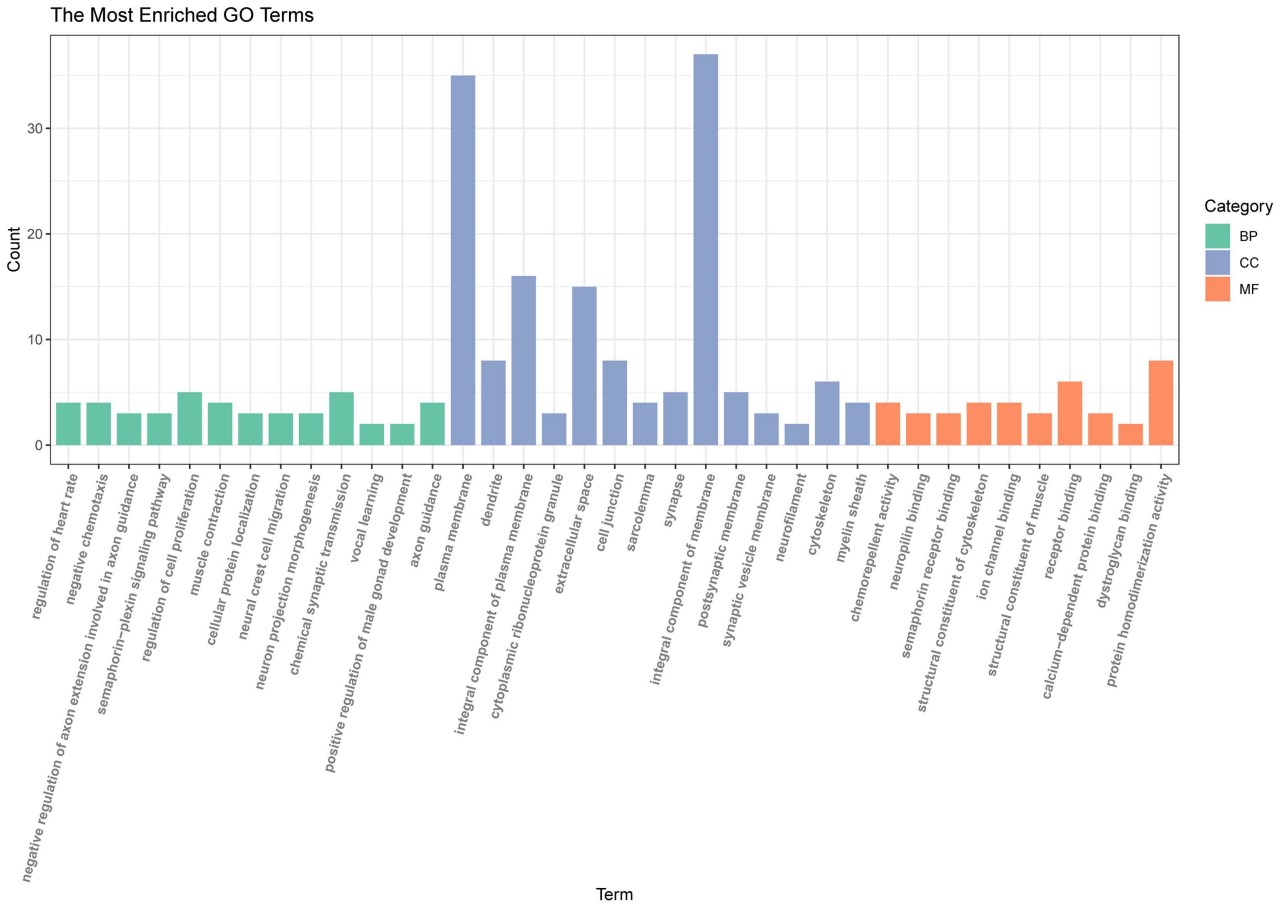

Figure 2. Gene Ontology (GO) annotation for differentially expressed genes in *F. nucleatum*-associated CRC. The X-axis represents gene function; the Y-axis represents the gene count. Three different colors of columns represent BP, CC, and MF.

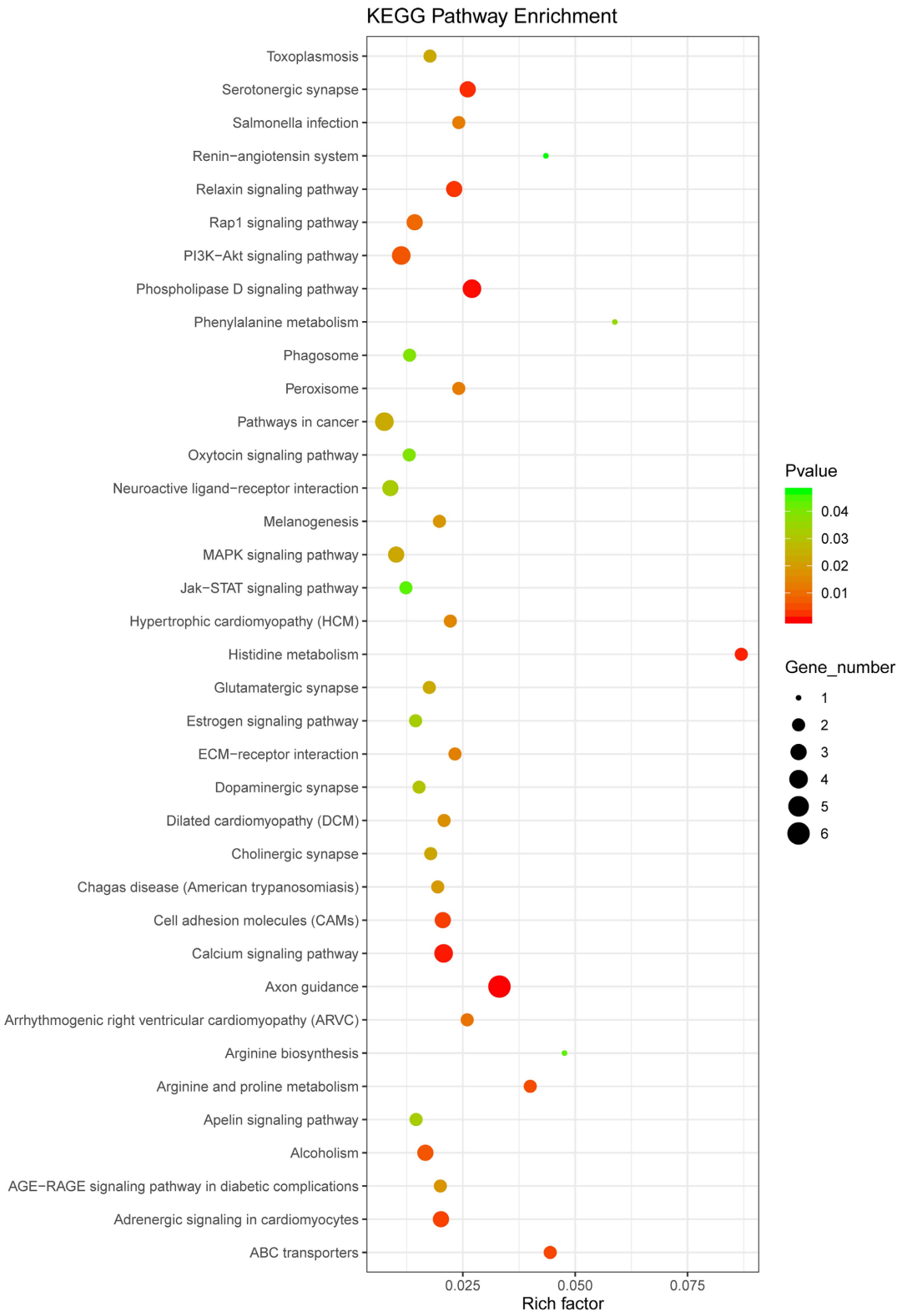

Figure 3. KEGG pathway enrichment for DEGS in *F. nucleatum*-associated CRC. The X-axis shows the rich factor; rich factor indicates the number of genes enriched in the target pathway and the proportion of all genes in the pathway; the Y-axis shows the significantly enriched pathways of DEGs. The size of the dot indicates the number of target genes in the pathway, and the color of the dot reflects the different *P*-value range ($P < 0.05$).

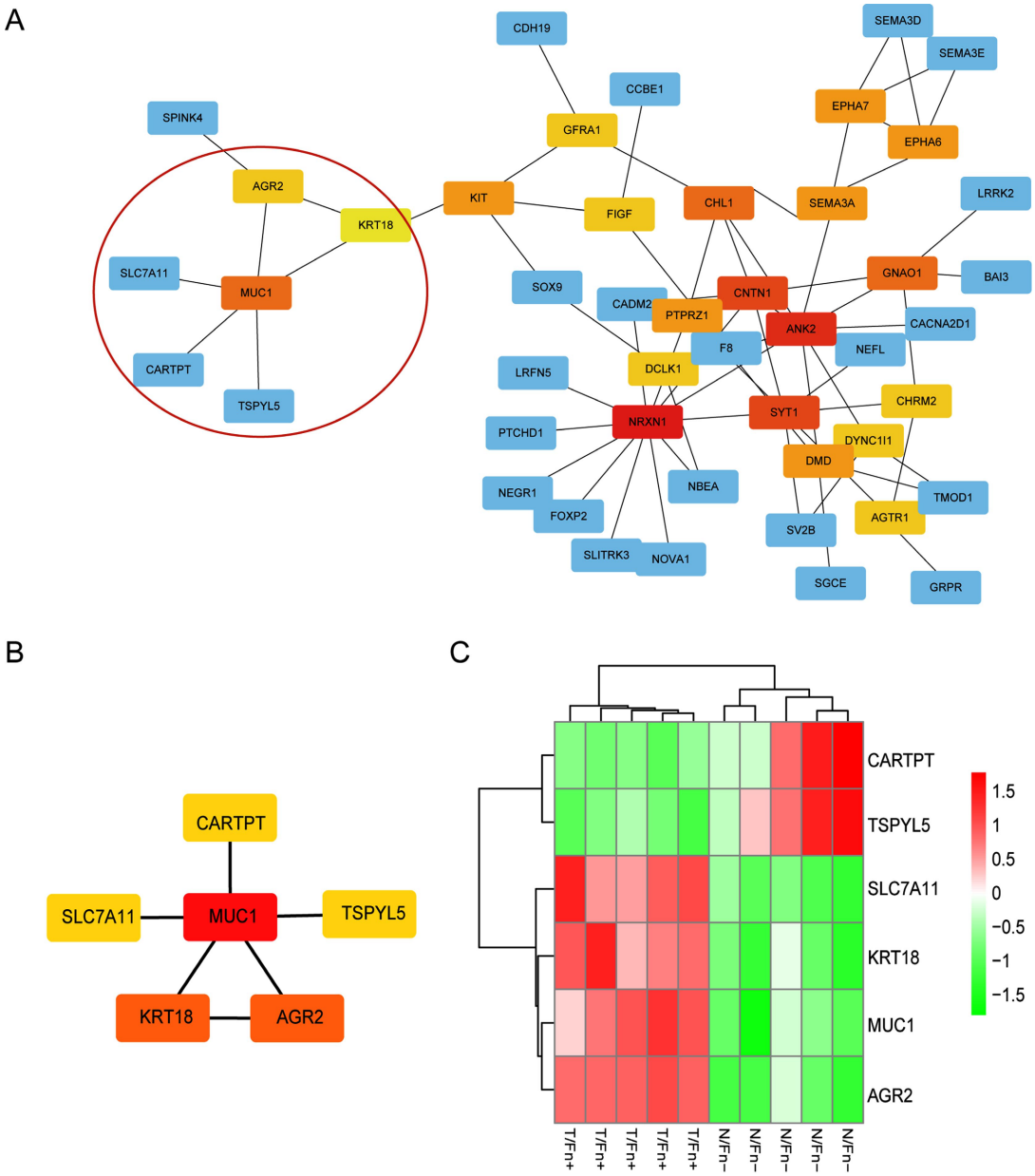

Figure 4. Screening of oncogenes in *F. nucleatum*-associated colorectal cancer through the construction of a PPI network. (A) Construction of the PPI network of differentially expressed genes in *F. nucleatum*-associated colorectal cancer. (B) PPI network of MUC1 and five other interacting hub genes. (C) Heat map of MUC1 and 5 hub genes interacting with it.

3.4. *F. nucleatum* Infection Was Significantly Associated with the Expression of Screened Hub Genes in Colorectal Normal/Cancer Cell Lines

To confirm the global effect of *F. nucleatum* infection *in vitro,* we cultured normal and tumor epithelial colorectal cell lines (NCW460, HCT116, SW620 and SW480) for 24 h with or without *F. nucleatum*. Then, the relative expression of MUC1, AGR2, SLC7A11, KRT18, CARTPT and TSPYL5 mRNA was measured by RT-qPCR. The measurements of the relative expression of mRNA were consistent

with the bioinformatics analysis. MUC1, SLC7A11, AGR2, and KRT18 were significantly upregulated in the CRC cell lines (HCT116, SW480 and SW620) infected with *F. nucleatum*. Moreover, CARTPT and TSPYL5 were significantly downregulated, similar to the results of the microarray assay, as shown in Figure 5A. Furthermore, we investigated the effect of *F. nucleatum* infection on the protein and mRNA expression of MUC1 in colorectal tumor or normal cell lines (NCW460, SW620, SW480 and HCT116). Real-time PCR analysis showed that *F. nucleatum* infection at different time intervals (4 h, 12 h, and 24 h) could increase the expression of MUC1 (Figure 6A). *F. nucleatum* infection caused differential expression of MUC1 protein at different time intervals in distinct cell lines, which were tested by western blots. Through the above experiments, we can conclude that the protein levels of MUC1 were significantly associated with *F. nucleatum* (Figure 6B).

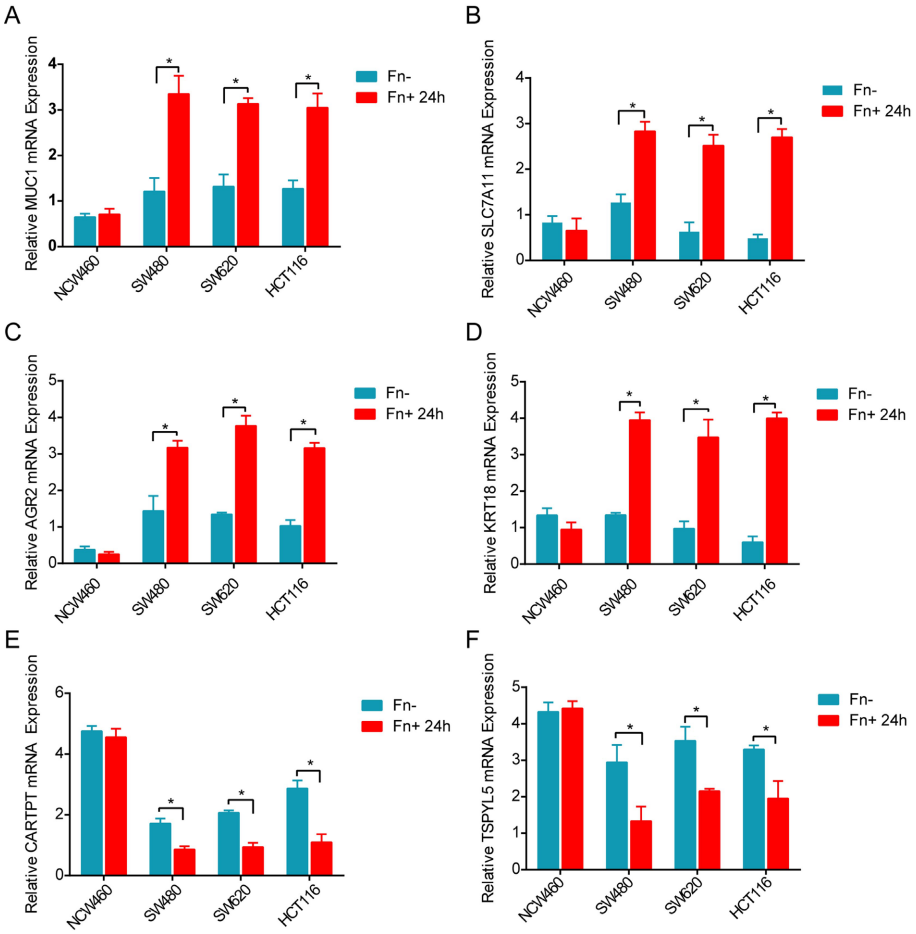

Figure 5. qRT-PCR analysis for *F. nucleatum* infection significantly induced the expression of screened hub genes. (A)-(D) Relative MUC1, SLC7A11, AGR2, and KRT18 mRNA expression in the CRC cell lines was significantly upregulated by *F. nucleatum* treatment. In addition, there was no significant change in the NCW460 cell model. (E)-(F) CARTPT and TSPYL5 are significantly downregulated by *F. nucleatum* infection in the treatment groups. Similarly, there was no significant change in the control groups. (All results were standardized by GAPDH. ***: $P < 0.001$, **: $P < 0.01$, *: $P < 0.05$.).

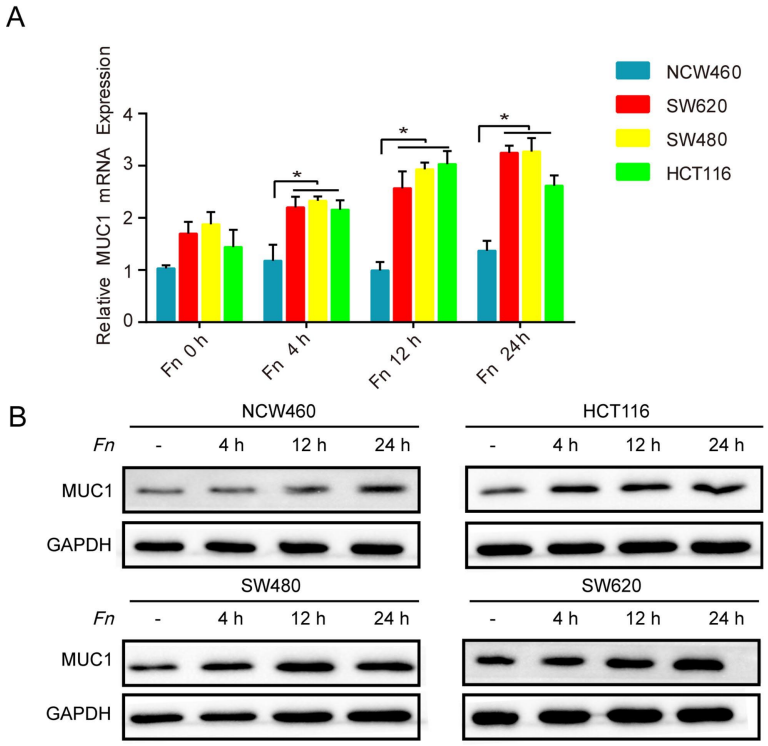

Figure 6. (A) Relative MUC1 mRNA expression was determined in the CRC cell lines with or without *F. nucleatum* for 4 h, 12 h and 24 h. (B) Representative western blots for the MUC1 and GAPDH proteins were extracted from the CRC cells infected with *F. nucleatum* for 0 h, 4 h, 12 h and 24 h.

4. Discussion

Colorectal carcinoma (CRC) is one of the most common cancers with a high morbidity/fatality rate worldwide [1]. CRC survival rates have been improved in recent decades, mainly due to screening. However, morbidity has increased among young patients [18]. Although this disease was previously attributed to an unhealthy diet, obesity, sedentary lifestyle and underlying genetic background, the cause of carcinogenesis is being reevaluated. The triggers for cancers are multifactorial, and accumulating evidence has shown that tissues continuously infected with microorganisms are more likely to become cancerous [19] [20]. Recent studies have revealed that *F. nucleatum* is closely associated with CRC and plays a key role in tumor initiation and progression through multiple biological mechanisms. *F. nucleatum* induced the cell surface protein Fad A to bind with E-cadherin in CRC epithelial cells to promote tumor cell growth by activating the β-catenin pathway and enhancing the expression of lymphoid enhancer factor (LEF)/T cytokines [21] [22]. In addition, E-cadherin activated the Wnt/β-catenin-Annexin A1 pathway by Fad A, thereby promoting bacterial invasion and tumor metastasis [23]. Human intestinal microbiota plays a vital role in various gastrointestinal diseases, especially in inflammatory bowel disease and CRC [24] [25], so the molecular mechanism between *F. nucleatum* and colorectal cancer is a current research focus [2]-[4].

Thousands of proteins and genes can be simultaneously examined by bioinformatics in a single experiment [26], and this technique is also increasingly used in clinical oncology research. The occurrence and development of CRC is a complex biological process, and bioinformatics has been a guide for discovering tumor markers through high-throughput computing, novel hypotheses and logical analytical strategies. These methods have become precise and efficient tools for studying the molecular mechanisms of colorectal cancer and are particularly necessary in our research. In this study, we used bioinformatics to screen hub genes differentially expressed in colorectal cancer resulting from *F. nucleatum* infection. The effect of *F. nucleatum* on differentially expressed genes in colorectal cancer was confirmed in cell line experiments. Through the identification of a reliable biomarker, it may become possible to detect patients during the primary stage of disease, thus significantly reducing the mortality and enhancing the survival of CRC. In this paper, the GSE110223, GSE110224, GSE113513 and GSE122183 datasets were extracted from GEO. A total of 118 DEGs of *F. nucleatum*-associated CRC were screened. Gene ontology functional analysis suggested that these DEGs were densely enriched in diverse GO terms, particularly chemical synaptic transmission, regulation of cell proliferation, integral component of membrane, extracellular space, protein homodimerization activity and receptor binding. In addition, several KEGG signaling pathways of DEGs were closely associated with CRC [27]-[30], including the PI3K−Akt, MAPK, Rap1 and JAK−STAT signaling pathways. Moreover, a PPI network with 6 hub genes was constructed (MUC1, AGR2, SLC7A11, KRT18, CARTPT and TSPYL5), and MUC1 was suggested as a key biotarget closely related to CRC. Finally, the results of cell lines by RT-qPCR and immunoblotting are consistent with the results of bioinformatics analysis.

Aberrantly glycosylated MUC1 as an oncogenic molecule has generated intense attention since it is overexpressed in most human epithelial cancers [31]. The interaction between some bacterial infections and MUC1 may be associated with the risk of developing cancer. MUC1 interacts directly with components of the NF-kB pathway during *Helicobacter pylori* infection, thus inhibiting tumor necrosis factor (TNF-α) inflammatory responses in gastric epithelial cells [32]. A previous study reported that highly invasive *F. nucleatum* evoked significantly greater MUC2 and tumor necrosis factor alpha (TNF-α) gene expression than minimally invasive strains in colonic epithelial cells [33]. In addition, since MUC1 is a proinflammatory factor following intestinal infection, it is an independent biomarker that can detect high expression in various stages of CRC, and it can be used in the early detection of CRC [34]. Therefore, we believe that MUC1, which is a valuable clinical tumor biomarker, is significantly overexpressed in CRC caused by *F. nucleatum*. The hub gene MUC1 was selected from the PPI network and verified *in vitro* using RT-qPCR and western blot experiments. In human CRC cell lines (SW480, SW620, HCT116), we found that the expression of MUC1 was significantly higher than that in the colonic epithelial cell line (NCW460). Similarly, MUC1 was significantly overexpressed in *F. nucleatum*-infected CRC

and intestinal epithelial cells compared to that in the noninfected control group. Thus, MUC1 was overexpressed in *F. nucleatum*-associated colorectal cancer, which indicates that it may be a novel oncogene and promising therapeutic biomarker for evaluating the prognosis of CRC.

We also screened five hub genes that interact with MUC1. Current studies have confirmed that AGR2, KRT18, TSPYL5 and SLC7A11 play a role in the initiation, chemotherapy resistance, metastasis, invasion and survival of CRC. Anterior gradient 2 (AGR2) protein upregulates Wnt11 to activate JNK and CaMKII to promote the migration of CRC cells [35]. One result showed that metformin-dependent activation of the AMPK signaling pathway was interfered with by the presence of AGR2. In colorectal cancer cells whose expression of AGR2 was interrupted, metformin enhanced the antitumor activity of 5-FU and oxaliplatin [36]. Moreover, in CRC patients, keratin 18 (KRT18) overexpression acted as an independent predictor and has been found to be detrimental for overall survival [37]. Testis-specific protein Y-encoded-like 5 (TSPYL5) is a tumor suppressor gene. Huang demonstrated that miR-19-5p plays a crucial role in the development and progression of CRC by targeting TSPYL5 [38]. Solute carrier family 7 member 11 (SLC7A11) is a crucial antioxidant component and a glutaminolysis-related gene. Ju found that downregulated SLC7A11 could inhibit tumor cell growth due to regulation by increased lactate production, resulting in a decreased tumor microenvironmental pH under the induction of low-dose paclitaxel (PTX) [39]. Although CARTPT has rarely been shown to play a role as a biomarker in the biological process of CRC, CARTPT has strong associations with other kinds of cancer. For example, CARTPT was associated with the risk of breast cancer [40]. According to the results of bioinformatics analysis, it is not difficult to predict that the abnormal expression of CARTPT caused by *F. nucleatum* may affect the tumorigenesis and biological behavior of colorectal cancer. Therefore, more studies will be explored and verified the role of Muc1 and the interaction with the aforementioned 5 DEGs in colorectal tumorigenesis.

In this study, we analyzed the potential association between *F. nucleatum* infection and colorectal carcinogenesis by bioinformatics. Although CRC patients infected with *F. nucleatum* are common, evidence showing that *F. nucleatum* is has a causative function in CRC is lacking. The causative role of *F. nucleatum* in CRC remains controversial [41]. In particular, *F. nucleatum* interacts with oncogenic bacteria in the intestinal microecology, and strategies for intestinal carcinogenesis are extremely important. To assess these roles, Tjalsma presented a bacterial 'driver–passenger' model to clarify how intestinal bacteria can directly or indirectly mediate CRC development [42]. Moreover, fusobacteria, an important 'bridging organism', cluster microorganisms together for interaction [43]. Many pathways that were screened from GO and KEGG are correlated with the pathogenesis of interactions between microbial species and interactions between intestinal microorganisms and hosts. Previous studies have shown that *P. gingivalis is* favorable for the survival of *F. nucleatum* and prompts them to invade deeper

periodontal tissues, which is due to infection inactivating the PI3K/AKT pathway [44]. The above conclusions indicate that the study of the correlation between *F. nucleatum* and CRC is important.

5. Conclusion

In conclusion, MUC1 was confirmed to be a novel oncogene of *F. nucleatum*-associated CRC by bioinformatics and experimental analyses. However, the role and mechanism of MUC1 in CRC caused by *F. nucleatum* remain to be confirmed by further discussion.

Acknowledgements

We thank the Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau), No. 2022MSXM002

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA*: *A Cancer Journal for Clinicians*, **68**, 394-424. https://doi.org/10.3322/caac.21492

[2] Castellarin, M., Warren, R.L., Freeman, J.D., Dreolini, L., Krzywinski, M., Strauss, J., *et al.* (2011) *Fusobacterium nucleatum* Infection Is Prevalent in Human Colorectal Carcinoma. *Genome Research*, **22**, 299-306. https://doi.org/10.1101/gr.126516.111

[3] Brennan, C.A. and Garrett, W.S. (2018) *Fusobacterium nucleatum*—Symbiont, Opportunist and Oncobacterium. *Nature Reviews Microbiology*, **17**, 156-166. https://doi.org/10.1038/s41579-018-0129-6

[4] Kostic, A.D., Chun, E., Robertson, L., Glickman, J.N., Gallini, C.A., Michaud, M., *et al.* (2013) *Fusobacterium nucleatum* Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host & Microbe*, **14**, 207-215. https://doi.org/10.1016/j.chom.2013.07.007

[5] Shang, F. and Liu, H. (2018) *Fusobacterium nucleatum* and Colorectal Cancer: A Review. *World Journal of Gastrointestinal Oncology*, **10**, 71-81. https://doi.org/10.4251/wjgo.v10.i3.71

[6] Brody, H. (2015) Colorectal Cancer. *Nature*, **521**, S1. https://doi.org/10.1038/521s1a

[7] Liang, B., Li, C. and Zhao, J. (2016) Identification of Key Pathways and Genes in Colorectal Cancer Using Bioinformatics Analysis. *Medical Oncology*, **33**, Article No. 111. https://doi.org/10.1007/s12032-016-0829-6

[8] Nannini, M., Pantaleo, M.A., Maleddu, A., Astolfi, A., Formica, S. and Biasco, G. (2009) Gene Expression Profiling in Colorectal Cancer Using Microarray Technologies: Results and Perspectives. *Cancer Treatment Reviews*, **35**, 201-209. https://doi.org/10.1016/j.ctrv.2008.10.006

[9] Guo, M., You, C. and Dou, J. (2018) Role of Transmembrane Glycoprotein Mucin 1

(MUC1) in Various Types of Colorectal Cancer and Therapies: Current Research Status and Updates. *Biomedicine & Pharmacotherapy*, **107**, 1318-1325. https://doi.org/10.1016/j.biopha.2018.08.109

[10] Agrawal, B., Gupta, N. and Konowalchuk, J.D. (2018) MUC1 Mucin: A Putative Regulatory (Checkpoint) Molecule of T Cells. *Frontiers in Immunology*, **9**, Article 2391. https://doi.org/10.3389/fimmu.2018.02391

[11] Yin, L., Li, Y., Ren, J., Kuwahara, H. and Kufe, D. (2003) Human MUC1 Carcinoma Antigen Regulates Intracellular Oxidant Levels and the Apoptotic Response to Oxidative Stress. *Journal of Biological Chemistry*, **278**, 35458-35464. https://doi.org/10.1074/jbc.m301987200

[12] Nath, S. and Mukherjee, P. (2014) MUC1: A Multifaceted Oncoprotein with a Key Role in Cancer Progression. *Trends in Molecular Medicine*, **20**, 332-342. https://doi.org/10.1016/j.molmed.2014.02.007

[13] Schroeder, J.A., Masri, A.A., Adriance, M.C., Tessier, J.C., Kotlarczyk, K.L., Thompson, M.C., *et al.* (2004) MUC1 Overexpression Results in Mammary Gland Tumorigenesis and Prolonged Alveolar Differentiation. *Oncogene*, **23**, 5739-5747. https://doi.org/10.1038/sj.onc.1207713

[14] Rakoff-Nahoum, S. (2006) Why Cancer and Inflammation? *Yale Journal of Biology and Medicine*, **79**, 123-130.

[15] Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., *et al.* (2014) STRING V10: Protein-Protein Interaction Networks, Integrated over the Tree of Life. *Nucleic Acids Research*, **43**, D447-D452. https://doi.org/10.1093/nar/gku1003

[16] Mendes, R.T., Nguyen, D., Stephens, D., Pamuk, F., Fernandes, D., Van Dyke, T.E., *et al.* (2016) Endothelial Cell Response to *Fusobacterium nucleatum*. *Infection and Immunity*, **84**, 2141-2148. https://doi.org/10.1128/iai.01305-15

[17] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2016) KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Research*, **45**, D353-D361. https://doi.org/10.1093/nar/gkw1092

[18] The Lancet Oncology, (2017) Colorectal Cancer: A Disease of the Young? *The Lancet Oncology*, **18**, 413. https://doi.org/10.1016/s1470-2045(17)30202-4

[19] Dalton-Griffin, L. and Kellam, P. (2009) Infectious Causes of Cancer and Their Detection. *Journal of Biology*, **8**, Article No. 67. https://doi.org/10.1186/jbiol168

[20] zur Hausen, H. (2009) The Search for Infectious Causes of Human Cancers: Where and Why. *Virology*, **392**, 1-10. https://doi.org/10.1016/j.virol.2009.06.001

[21] Rubinstein, M.R., Wang, X., Liu, W., Hao, Y., Cai, G. and Han, Y.W. (2013) *Fusobacterium nucleatum* Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via Its FadA Adhesin. *Cell Host & Microbe*, **14**, 195-206. https://doi.org/10.1016/j.chom.2013.07.012

[22] Chen, Y., Peng, Y., Yu, J., Chen, T., Wu, Y., Shi, L., *et al.* (2017) Invasive *Fusobacterium nucleatum* Activates β-Catenin Signaling in Colorectal Cancer via a TLR4/P-PAK1 Cascade. *Oncotarget*, **8**, 31802-31814. https://doi.org/10.18632/oncotarget.15992

[23] Rubinstein, M.R., Baik, J.E., Lagana, S.M., Han, R.P., Raab, W.J., Sahoo, D., *et al.* (2019) *Fusobacterium nucleatum* Promotes Colorectal Cancer by Inducing Wnt/β-Catenin Modulator Annexin A1. *EMBO Reports*, **20**, e47638. https://doi.org/10.15252/embr.201847638

[24] Tilg, H., Adolph, T.E., Gerner, R.R. and Moschen, A.R. (2018) The Intestinal

Microbiota in Colorectal Cancer. *Cancer Cell*, **33**, 954-964. https://doi.org/10.1016/j.ccell.2018.03.004

[25] Arthur, J.C., Perez-Chanona, E., Mühlbauer, M., Tomkovich, S., Uronis, J.M., Fan, T., *et al.* (2012) Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science*, **338**, 120-123. https://doi.org/10.1126/science.1224820

[26] Kulasingam, V. and Diamandis, E.P. (2008) Strategies for Discovering Novel Cancer Biomarkers through Utilization of Emerging Technologies. *Nature Clinical Practice Oncology*, **5**, 588-599. https://doi.org/10.1038/ncponc1187

[27] Bahrami, A., Khazaei, M., Hasanzadeh, M., ShahidSales, S., Joudi Mashhad, M., Farazestanian, M., *et al.* (2017) Therapeutic Potential of Targeting PI3K/AKT Pathway in Treatment of Colorectal Cancer: Rational and Progress. *Journal of Cellular Biochemistry*, **119**, 2460-2469. https://doi.org/10.1002/jcb.25950

[28] Uddin, S., P.Bavi, P., Hussain, A.R., Alsbeih, G., Al-Sanea, N., AbdulJabbar, A., *et al.* (2009) Leptin Receptor Expression in Middle Eastern Colorectal Cancer and Its Potential Clinical Implication. *Carcinogenesis*, **30**, 1832-1840. https://doi.org/10.1093/carcin/bgp145

[29] Uddin, S., Hussain, A.R., Khan, O.S. and Al-Kuraya, K.S. (2013) Role of Dysregulated Expression of Leptin and Leptin Receptors in Colorectal Carcinogenesis. *Tumor Biology*, **35**, 871-879. https://doi.org/10.1007/s13277-013-1166-4

[30] Tang, Y., Zhao, Y., Song, X., Song, X., Niu, L. and Xie, L. (2019) Tumor-Derived Exosomal miRNA-320d as a Biomarker for Metastatic Colorectal Cancer. *Journal of Clinical Laboratory Analysis*, **33**, e23004. https://doi.org/10.1002/jcla.23004

[31] Lau, S.K., Weiss, L.M. and Chu, P.G. (2004) Differential Expression of MUC1, MUC2, and MUC5AC in Carcinomas of Various Sites an Immunohistochemical Study. *American Journal of Clinical Pathology*, **122**, 61-69. https://doi.org/10.1309/9r66-73qe-c06d-86y4

[32] Sheng, Y.H., Triyana, S., Wang, R., Das, I., Gerloff, K., Florin, T.H., *et al.* (2013) MUC1 and MUC13 Differentially Regulate Epithelial Inflammation in Response to Inflammatory and Infectious Stimuli. *Mucosal Immunology*, **6**, 557-568. https://doi.org/10.1038/mi.2012.98

[33] Dharmani, P., Strauss, J., Ambrose, C., Allen-Vercoe, E. and Chadee, K. (2011) *Fusobacterium nucleatum* Infection of Colonic Cells Stimulates MUC2 Mucin and Tumor Necrosis Factor *α*. *Infection and Immunity*, **79**, 2597-2607. https://doi.org/10.1128/iai.05118-11

[34] Al-Khayal, K., Abdulla, M., Al-Obaid, O., Zubaidi, A., Vaali-Mohammed, M., Alsheikh, A., *et al.* (2016) Differential Expression of Mucins in Middle Eastern Patients with Colorectal Cancer. *Oncology Letters*, **12**, 393-400. https://doi.org/10.3892/ol.2016.4672

[35] Tian, S., Hu, J., Tao, K., Wang, J., Chu, Y., Li, J., *et al.* (2018) Secreted AGR2 Promotes Invasion of Colorectal Cancer Cells via Wnt11-Mediated Non-Canonical Wnt Signaling. *Experimental Cell Research*, **364**, 198-207. https://doi.org/10.1016/j.yexcr.2018.02.004

[36] Martisova, A., Sommerova, L., Kuricova, K., Podhorec, J., Vojtesek, B., Kankova, K., *et al.* (2019) AGR2 Silencing Contributes to Metformin-Dependent Sensitization of Colorectal Cancer Cells to Chemotherapy. *Oncology Letters*, **18**, 4964-4973. https://doi.org/10.3892/ol.2019.10800

[37] Zhang, J., Hu, S. and Li, Y. (2019) KRT18 Is Correlated with the Malignant Status and Acts as an Oncogene in Colorectal Cancer. *Bioscience Reports*, **39**, BSR20190884. https://doi.org/10.1042/bsr20190884

[38] Huang, C. and Luo, H. (2018) *miR-19-5p* Enhances Tumorigenesis in Human Colorectal Cancer Cells by Targeting *TSPYL5*. *DNA and Cell Biology*, **37**, 23-30. https://doi.org/10.1089/dna.2017.3804

[39] Ju, H., Lu, Y., Chen, D., Tian, T., Mo, H., Wei, X., *et al.* (2016) Redox Regulation of Stem-Like Cells Though the CD44v-xCT Axis in Colorectal Cancer: Mechanisms and Therapeutic Implications. *Theranostics*, **6**, 1160-1175. https://doi.org/10.7150/thno.14848

[40] Slattery, M.L., Lundgreen, A., Hines, L., Wolff, R.K., Torres-Mejia, G., Baumgartner, K.N., *et al.* (2015) Energy Homeostasis Genes and Breast Cancer Risk: The Influence of Ancestry, Body Size, and Menopausal Status, the Breast Cancer Health Disparities Study. *Cancer Epidemiology*, **39**, 1113-1122. https://doi.org/10.1016/j.canep.2015.08.012

[41] Zhang, S., Cai, S. and Ma, Y. (2018) Association between *Fusobacterium nucleatum* and Colorectal Cancer: Progress and Future Directions. *Journal of Cancer*, **9**, 1652-1659. https://doi.org/10.7150/jca.24048

[42] Tjalsma, H., Boleij, A., Marchesi, J.R. and Dutilh, B.E. (2012) A Bacterial Driver-Passenger Model for Colorectal Cancer: Beyond the Usual Suspects. *Nature Reviews Microbiology*, **10**, 575-582. https://doi.org/10.1038/nrmicro2819

[43] Kaplan, C.W., Lux, R., Haake, S.K. and Shi, W. (2008) The *Fusobacterium nucleatum* Outer Membrane Protein Radd Is an Arginine-Inhibitable Adhesin Required for Inter-Species Adherence and the Structured Architecture of Multispecies Biofilm. *Molecular Microbiology*, **71**, 35-47. https://doi.org/10.1111/j.1365-2958.2008.06503.x

[44] Jung, Y., Jun, H. and Choi, B. (2017) *Porphyromonas gingivalis* Suppresses Invasion of *Fusobacterium nucleatum* into Gingival Epithelial Cells. *Journal of Oral Microbiology*, **9**, Article ID: 1320193. https://doi.org/10.1080/20002297.2017.1320193