A comprehensive analysis of the microbiota composition and gene expression in colorectal cancer

Qian Zhang

The First Affiliated Hospital of Zhengzhou University, The First People's Hospital of Zhengzhou

Huan Zhao

The First Affiliated Hospital of Zhengzhou University

Dedong Wu

The First People's Hospital of Zhengzhou

Dayong Cao

The First People's Hospital of Zhengzhou

Wang Ma (✉ m_w010203@126.com)

The First Affiliated Hospital of Zhengzhou University

Research article

Keywords: colorectal cancer, gut microflora, gene expression, pathways enrichment, survival analysis

Posted Date: October 30th, 2019

DOI: <https://doi.org/10.21203/rs.2.16589/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Subject: The dysbiosis of gut microbiota is pivotal in colorectal carcinogenesis. However, the synergy between an altered gut microbiota composition and differential gene expression of specific genes in colorectal cancer (CRC) remains elusive.

Method: The gut microbiota dataset with number SRP158779, which contained 19 CRC samples and 19 normal samples, was downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database. The 16S rRNA gene sequences from this dataset were clustered into operational taxonomic units (OTUs); thereafter, the OTUs that were differentially enriched in CRC were identified and classified, followed by prediction of their functions. Additionally, RNA sequencing data from CRC samples was obtained from The Cancer Genome Atlas project (TCGA), and the differentially expressed genes (DEGs) and enriched pathways were identified. Finally, similar pathways that were significantly enriched in both differential OTUs and DEGs were screened. Key genes related to these pathways were executed the prognosis analysis.

Results: The presence of Proteobacteria and Fusobacteria increased considerably in CRC samples; conversely, the abundance of Firmicute and Spirochaetes decreased markedly. In particular, the genera *Fusobacterium*, *Catenibacterium*, and *Shewanella* were detectable in tumor samples. Moreover, 246 DEGs were identified between tumor and normal tissues. Both DEGs and microbiota were involved in bile secretion and steroid hormone biosynthesis pathways. Finally, CYP3A4 and ABCG2 expression in CRC was related to the prognostic outcomes of CRC patients.

Conclusion: Identifying the complicated interplay between gut microbiota and the DEGs could help in further understanding the pathogenesis of CRC, and these findings would enable better diagnosis and treatment of CRC patients.

Keywords: colorectal cancer, gut microflora, gene expression, pathways enrichment, survival analysis

Background

Colorectal cancer (CRC) is one of the primary causes of mortality and morbidity worldwide, thus representing a major public health issue [1]. Although heritable genetic mutations are closely linked to some types of CRC [2], increasing evidences indicate diet is regarded as a notable risk factor of CRC [3, 4]. Intake of extra red meat and animal fat might increase the risk of CRC [5]. It is obvious that diet can modulate the composition of gut microbiota. Different members of the intestinal microbiota can work together to regulate the host immune and metabolic systems, subsequently producing carcinogenic or anticancer substances [6, 7]. Lately, evidence for the role of intestinal microbiota in health and disease has been increasing [8, 9]. Flemer *et al.* proposed that the disharmony of intestinal microbiota might influence the pathogenesis of CRC [10]. Hence, applying strategies to manage the gut microbiota composition to promote recovery of a favorable microbiota community may be therapeutically feasible in patients with CRC.

Additionally, it has been proposed that an altered gut microbiome may affect the development of intestinal diseases through interaction with the innate immune system and other host factors [11]. Huang and colleagues demonstrated that the possible pathogenic flora of colitis-associated cancer connected with the C-X-C motif receptor 2 (CXCR2) signaling axis during cancer progression [12]. Similarly, enteric microbiota dysbiosis and genetic abnormalities led to disruption of the intestinal barrier, thus triggering early kidney injury in mice [13]. Moreover, Thompson *et al.* identified that *Haemophilus influenza* was significantly connected to the genes that were enriched in the G2M checkpoint and E2F transcription pathways in breast cancer [14]. These findings emphasize that the significance of the metabolic functions, that these enrichment of specific pathways can influence the development of cancer through altered gene expression and microbiota composition. However, pathways that are involved in the altered microbiota composition and differential gene expression in CRC have not been identified. Furthermore, specific genes that can disrupt the gut microbial composition and ultimately cause CRC are not well recognized.

The objective of this study was to characterize the relationship between gut microbiota and CRC mRNA expression profiles. An integrated approach of 16S rRNA gene sequence and transcriptome sequence profiling from a public database was performed. The intestinal microbiota dataset (SRP158779) was downloaded from the Sequence Read Archive (SRA) database, and the RNA sequencing data was obtained from The Cancer Genome Atlas project (TCGA). In addition, we analyzed the significantly altered gut microbiota composition and several differentially expressed genes (DEGs) in CRC sample datasets. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed in relation to the differentially enriched microbiota OTUs and gene expression, and the co-enrichment pathways were screened. Finally, survival analysis was further performed in the DEGs that were enriched in the co-enrichment pathways.

Methods

Data resource

The raw intestinal microbiota data with the number SRP158779 (<http://www.ncbi.nlm.nih.gov/sra/SRP158779>), was retrieved from the NCBI SRA public database. This dataset contained 38 samples, including 19 CRC tumors samples and 19 paired nonneoplastic tissues patients. Sequencing was performed using Illumina HiSeq 2000 platform by single-ended sequencing. In addition to the intestinal microbiota data, raw RNA sequence data of 371 CRC tumors and 51 normal samples were retrieved from The Cancer Genome Atlas (TCGA).

Acquisition of the OTU table

The SRA raw data was converted to fastq format using fastq-dump software (parameter: split-3). We removed the low quality sequencing reads before subsequent analysis. Quantitative Insights Into

Microbial Ecology (QIIME) (version 1.4.0) [15] software was used to analyze the subsequent high quality sequences. Primarily, starting from the two-terminal sequence combination, the amplified primers were excised and chimera sequences were removed. Additionally, the clean sequences were divided for diversity and taxonomic composition based on the Greengenes database (release 13.5, <http://greengenes.secondgenome.com/>) [16]. Sequences were clustered into operational taxonomic units (OTUs) using UCLUST (version 1.2.22q, <http://www.drive5.com/usearch/>) [17], and 97% was defined as threshold. Thereafter, the sequence with the highest abundance was selected as the representative sequence. Ultimately, based on the number of sequences included in each OTU, the OTU table abundance in each sample was constructed.

Alpha and beta diversity analysis

Abundance and diversity of microbial communities could be reflected by alpha diversity. The species diversity was calculated using the Shannon, Simpson, Chao1, and PD_whole_tree indices. Concretely, the Shannon and Simpson indices were used to represent the community diversity, and chao 1 was used to indicate the community richness, whereas PD_whole_tree was used to symbolize the phylogenetic diversity. All the statistical analysis was based on the R phyloseq package [18]. Additionally, the rarefaction curve could explain the rationality of the amount of sequencing data. Beta diversity analysis was used to examine the similarity of community structure between different samples. Beta diversity was calculated by the QIIME software and cluster analysis was conducted by principal component analysis (PCA); thereafter, RGL package in R software was applied to visualize the results.

Differential OTU analysis and classification

The OTU data were normalized by using trimmed mean of M values method from edgeR package. Subsequently, analysis of the significantly different OTUs were performed by F-test of edgeR. All this analysis was based on R software. P value < 0.05 was considered as statistically significant. Additionally, taxonomic assignments of OTUs that reached 97% similarity level were performed using RDP classifier (<https://sourceforge.net/projects/rdp-classifier/>) [19] by comparing with the Greengene database (Release 13.5, <http://greengenes.secondgenome.com/>) [20].

Network analysis of microbiome

Based on the differential OTUs results, correlation coefficient matrix between OTU was calculated by igraph (version 1.2.2) and psych (version 1.8.4) in R software, then the screening threshold of the significant interaction was p value < 0.05 and $|r| > 0.6$. The network was visualized by Cytoscape (version 2.8).

KEGG pathway prediction of differential OTUs

Phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) is a computational approach to predict the function of bacteria by comparing the 16S rRNA gene sequencing data with the microbial reference genome database (Greengenes), which can predict metabolic function of bacteria and archaea. The main procedures are displayed as following: 1) According to the full-length sequence of 16S rRNA gene to predict genetic functional profiles of common ancestors. 2) The functional profiles of other untested species in the Greengenes database were deduced, and then the genetic function prediction spectrum of archaea and bacteria domain was constructed. 3) The composition of the sequenced bacteria was “mapped” into the KEGG database to predict the metabolic function of the microbiota. EdgeR was used to identify the bacteria associated pathways, and p value < 0.05 was considered as statistically significant.

DEG and Enriched pathway analyses

The raw reads of the TCGA dataset were transformed on the base (count + 1) logarithm for further analysis. Subsequently, the data were normalized and analyzed by edgeR, and genes with $|\log FC| > 1.5$ and FDR < 0.05 were selected as significantly differentially expressed in CRC tumor. The pathway enrichment analysis was carried out by clusterProfiler [21] of R package. Thereafter, pathways with the count ≥ 2 and p value < 0.05 were defined as the pathways significantly enriched in the CRC tumor.

Integration of amplicon and transcriptome

To explore the relationship between differential OTUs and DEGs, the functional prediction of differential OTUs and DEGs was integrated. The same or similar pathways that were shared between the two sets of data were selected, and the obtained pathways indicated that both transcriptome genes and intestinal microbiota affected the occurrence of CRC.

Survival analysis

The prognosis outcomes including overall survival (OS) and survival status of CRC patients were obtained from TCGA, and the survival analysis of 371 patients with CRC were performed. The Kaplan–Meier survival curves of these two groups (high expression and low expression) were plotted, and the relationship between gene expression and survival was analyzed by the log-rank tests. $P < 0.05$ was set as the cut-off criteria for statistical significance.

Results

Rarefaction curve and diversity analysis

Rarefaction curves of the number of OTUs at different sequencing depths were obtained (Figure 1A). The curves of each sample tend to flatten, suggesting that reasonable amount of sequencing data was available. The three-dimensional PCA plot showed a difference in the composition of the intestinal microbiota community between CRC and normal samples (Figure 1B). Furthermore, the alpha diversity results revealed that there was a complex bacterial community that was not significantly different from the normal samples in the tumor samples (Figure 1C).

Taxonomic composition

A total of 13 different phyla were detected from these two groups. The result is shown in Figure 2A and B. Typically, four phyla were highly abundant among all the samples, including *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, and *Actinobacteria*. In the normal group, an increase in the abundance of *Firmicutes* (50.7%), *Bacteroidetes* (21.5%), and *Proteobacteria* (15.9%) was observed. In the tumor group, the quantity of *Proteobacteria* (21.7%) and *Fusobacteria* (3.8%) increased, whereas the abundance of *Firmicutes* (46.5%) and *Spirochaetes* (0.1%) decreased. Moreover, genus analysis demonstrated that the level of *Bacteroides* was advantaged across two groups (Figure 2C and D). Furthermore, *Fusobacterium* (6.7%), *Catenibacterium* (2.5%), and *Shewanella* (2.0%) were the genera specifically detectable in the tumor samples.

Differentially enriched OTU and pathway enrichment analysis

A total of 66 significantly differential enriched OTUs were identified in CRC samples, including 22 up-regulated and 44 down-regulated OTUs. The differential OTUs were visualized by using volcano plot and hierarchical clustering (Figure 3 A and B). There was significant difference between tumor and normal samples. *Actinobacteria* and *Fusobacteria* was in a significantly higher abundance in the tumor samples in comparison to normal samples. *Firmicute* and *Proteobacteria* were at significantly lower abundance in tumor samples in comparison to the normal samples. In particular, based on the functional prediction, we observed considerable differences in the KEGG pathways. A total of 67 pathways were significantly enriched in CRC (Figure 3C). These differences implicated that the differential OTUs were involved in CRC, MAPK signaling pathway, and p53 signaling pathway, as well as these pathways were closely connected with CRC.

Network analysis of microbiome

A network composed of 35 nodes and 55 edges was constructed to describe the complex relationships of microbiome (Figure 4). All nodes were assigned to eight bacterial phyla, including *Firmicutes* (57.55%), *Crenarchaeota* (10.40%), *Bacteroidetes* (10.39%), *Actinobacteria* (8.87%), *Proteobacteria* (6.52%), *Cyanobacteria* (5.21%), *Euryarchaeota* (1.03%), and *Spirochaetes* (0.03%). Additionally, total 16 genera

were identified. OTUs with highest number of connections in the correlation networks belonged to *Bacteroides*. Specially, for genera related to CRC, *Butyrivimonas* showed closely interacting with *Clostridium*.

DEG identification and KEGG pathway enrichment analysis

A total of 246 DEGs comprising 222 up-regulated and 24 down-regulated genes were screened between tumor and normal samples. By comparing the normal group with CRC group, we found that apolipoprotein B (*APOB*) and carbonic anhydrase 1 (*CA1*) were significantly up-regulated, whereas the angiopoietin like 5 (*ANGPTL5*) and shisa family member 7 (*SHISA7*) were down-regulated in the tumor samples. Furthermore, 22 of the up-regulated KEGG pathways were significantly enriched for specific DEGs (Figure 5A), such as genes involved in chemical carcinogenesis, drug metabolism-cytochrome P450, and bile secretion. The down-regulated DEGs were closely correlated to mineral absorption (Figure 5B).

Integrated analysis

The DEGs and OTUs specifically enriched in CRC samples were compared, and two pathways were filtrated, including bile secretion and steroid hormone biosynthesis. These two pathways could affect CRC not only at transcriptome levels, but also through the alteration of specific intestinal microbiota abundance (Figure 5C). Additionally, a set of 11 DEGs was significantly up-regulated in both of these pathways.

Survival analysis

According to the above candidate genes, the K-M survival curve was drawn. Three genes were significantly related to the survival analysis, including the solute carrier family 4 member 4 (*SLC4A4*), cytochrome P450 family 3 subfamily A member 4 (*CYP3A4*), and ATP binding cassette subfamily G member 2 (*ABCG2*). The K-M survival curve is visualized in Figure 6. The results of median survival indicated that the low-level expression of *CYP3A4* and *ABCG2* significantly prolonged the median survival of the patients with CRC.

Discussion

In this study, a total of 22 OTUs were identified as up-regulated and 44 OTUs as down-regulated in CRC in comparison to normal samples. At the phyla level, the abundance of *Proteobacteria* and *Fusobacteria* increased, whereas that of *Firmicutes* decreased in the tumor samples. *Fusobacterium*, *Catenibacterium*, and *Shewanella* were specifically detectable in the tumor samples. Additionally, a total of 246 DEGs (222 up-regulated and 24 down-regulated in CRC in comparison to normal samples) were identified. Moreover,

bile secretion and steroid hormone biosynthesis were the two co-enrichment pathways, which were associated with both DEGs and differential enriched microbiota OTU in CRC. Furthermore, we found that the low expression of *CYP3A4* and *ABCG2* could significantly prolong the median survival of CRC patients.

The tumor microenvironment of CRC is a complex community of cancer cells, noncancerous cells, and diverse microbiota [22]. The imbalance of gut microbiota may contribute to carcinogenesis. Consistent with the findings of the present study, Aleksandar *et al.* have reported that OTUs of *Fusobacterium* are enriched in carcinomas, whereas those of *Firmicutes* are significantly decreased in tumors [23]. Specifically, *Fusobacterium* was markedly enriched in the tumor tissues. Previous studies have suggested that *Fusobacterium* expresses the virulence factor FadA, which activates the WNT signaling pathways, thus promoting tumor growth in CRC [24]. *Fusobacterium* has also been identified as being able to inhibit immune responses in CRC tumors [15]. *Fusobacterium* species are also known to induce host proinflammatory responses and possess virulence [25]. Our findings support the above reports, and highlight that the clinical relevance of *Fusobacterium* enrichment in CRC should be addressed in further studies. *Shewanella* was another genus particularly enriched in the tumor tissues. *Shewanella* has been reported to cause pulmonary and blood infections [26]. *Shewanella* algae could raise purulent pericarditis with greenish pericardial effusion [27]. However, its role has not been in CRC tumor progression is not well-defined. Our studies suggest that *Shewanella* might serve as a potential biomarker of CRC. However, further detailed studies are required to verify this possibility. Interestingly, a closely interaction between *Butyricimonas* and *Clostridium* was observed in the microbiome network. Wu *et al.* indicted that *Butyricimonas* was only detected in colorectal cancer group from mouse model [28]. Meanwhile, *Clostridium* was a risk factor for the development of CRC [21]. Therefore, we predicted that this relationship might play a pivotal role in the pathogenesis of CRC, and these two microbiota could be utilized as predictors of CRC diagnosis.

Two pathways were significantly enriched in both CRC associated DEGs and CRC enriched microbiota OTUs: the bile secretion and steroid hormone biosynthesis pathways. Eleven DEGs were up-regulated in these two pathways. Subsequently, when they were compared to the clinical characteristics data of CRC patients, the *CYP3A4* and *ABCG2* were considered to be specifically connected to CRC prognosis. The low-level expression of *CYP3A4* was related to a significantly prolonged median survival of CRC patients. *CYP3A4* encodes a member of the cytochrome P450 superfamily of enzyme [29]. Recent studies have provided evidence that the genotoxicity of the carcinogen is influenced by cytochrome P450 enzyme system, and *CYP3A* is highly expressed in colonic tissue [30]. Additionally, *CYP3A* is involved in the metabolism of carcinogen, and is associated with inactivation of anticancer drugs [31]. Evidences indicate that the *CYP3A* mRNA transcripts are present in the human colorectal epithelium and CRC cell lines [32]. In this study, we found that the expression of *CYP3A4* was up-regulated in tumor tissues, and this finding concords with the above evidences. Furthermore, survival analysis indicated that patients with a high level of *CYP3A4* showed a significantly shortened survival time. *CYP3A4* in CRC tumor samples may influence tumor sensitivity, and might induce drug resistance against certain colon cancer drugs [33]. Therefore, *CYP3A4* can lead to poor prognosis. Our study implicated that *CYP3A4* is involved

in steroid hormone biosynthesis pathway. A previous study on gastric cancer (GC) suggested that steroid hormone biosynthesis pathway and their receptors expressions could be altered by genetic variations, thereby contributing to susceptibility to GC [34]. Therefore, changes in the expression level of *CYP3A4* observed in this study may have an impact on the steroid hormone biosynthesis pathway, thereby affecting the prognosis of CRC. We propose that *CYP3A4* might serve as prognostic biomarker and a therapeutic target for CRC.

In addition to *CYP3A4*, *ABCG2* has also been related to the prognosis of CRC, and patients with a low transcript level of *ABCG2* could prolong the median survival of patients with CRC. *ABCG2* is a member of the superfamily of ATP-binding cassette (ABC) transporter protein. *ABCG2* can induce drug resistance and treatment failure in tumor tissues [35]. Our findings were supported by a previous study, which suggested that *ABCG2* was highly expressed in CRC, and that *ABCG2* might be involved in progression and metastasis of advanced malignancy cancer [36]. In a previous study, higher *ABCG2* mRNA expression also represented an unfavorable prognostic factor of esophageal squamous cell carcinoma [37]. These findings support our view that *ABCG2* could be regarded as a prognostic marker of CRC. Moreover, we observed that *ABCG2* was associated with bile secretion pathway. *ABCG2* is a hepatobiliary efflux transporter and is involved in the biliary excretion of sulfate conjugates [38] and troglitazone sulfate of therapeutics [39]. Reportedly, bile acids role as tumor promoters have been tested using extensive experiments [40, 41]. As a result, we speculated that *ABCG2* might ultimately affect the prognosis of CRC by influencing bile secretion.

Taken together, our study highlights that changes in gene expression and microbiota composition are linked to the functional enrichment of specific pathways. Differential expression of genes caused the alteration of the bile secretion and steroid hormone biosynthesis in CRC tissues, thereby changing the abundance and composition of intestinal microbiota, and eventually triggering the occurrence of cancer. However, our study was based on bioinformatics analyses of the datasets from a public database, and further experimental studies must be conducted to further validate and strengthen our results.

Conclusion

By integrating the results of 16S rRNA gene sequencing and transcriptome sequence datasets, we revealed a relationship between the DEGs and CRC enriched microbiota, and gained better insights into pathogenesis and prognosis of CRC. Our study might provide a new perspective for the diagnosis and treatment of CRC by targeting novel genes and microbiota biomarkers.

List Of Abbreviations

colorectal cancer (CRC)

National Center for Biotechnology Information (NCBI)

Sequence Read Archive (SRA)

operational taxonomic units (OTUs)

The Cancer Genome Atlas (TCGA)

differentially expressed genes (DEGs)

C-X-C motif receptor 2 (CXCR2)

Kyoto Encyclopedia of Genes and Genomes (KEGG)

Quantitative Insights Into Microbial Ecology (QIIME)

principal component analysis (PCA)

overall survival (OS)

apolipoprotein B (*APOB*)

carbonic anhydrase 1 (*CA1*)

angiopoietin like 5 (*ANGPTL5*)

shisa family member 7 (*SHISA7*)

solute carrier family 4 member 4 (*SLC4A4*)

cytochrome P450 family 3 subfamily A member 4 (*CYP3A4*)

ATP binding cassette subfamily G member 2 (*ABCG2*)

ATP-binding cassette (ABC)

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

None.

Authors' contributions

Conception and design of the research: QZ, WM; acquisition of data: DW; analysis and interpretation of data: HZ; statistical analysis: QZ, DC;; drafting the manuscript: QZ; revision of manuscript for important intellectual content: WM. All authors read and approved the final manuscript.

Acknowledgements

None.

References

- 1.Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi REM, Corcione F: *Worldwide burden of colorectal cancer: a review. Updates in surgery* 2016, *68*(1):7–11.
- 2.Lynch HT, De la Chapelle A: *Hereditary colorectal cancer. New England Journal of Medicine* 2003, *348*(10):919–932.
- 3.Baena R, Salinas P: *Diet and colorectal cancer. Maturitas* 2015, *80*(3):258–264.
- 4.Rosato V, Guercio V, Bosetti C, Negri E, Serraino D, Giacosa A, Montella M, La Vecchia C, Tavani A: *Mediterranean diet and colorectal cancer risk: a pooled analysis of three Italian case–control studies. British journal of cancer* 2016, *115*(7):862.
- 5.Chan DS, Lau R, Aune D, Vieira R, Greenwood DC, Kampman E, Norat T: *Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. PloS one* 2011, *6*(6):e20456.
- 6.Shapiro H, Thaiss CA, Levy M, Elinav E: *The cross talk between microbiota and the immune system: metabolites take center stage. Current opinion in immunology* 2014, *30*:54–62.
- 7.Sekirov I, Russell SL, Antunes LCM, Finlay BB: *Gut microbiota in health and disease. Physiological reviews* 2010, *90*(3):859–904.

- 8.Sonnenburg ED, Smits SA, Tikhonov M, Higginbottom SK, Wingreen NS, Sonnenburg JL: *Diet-induced extinctions in the gut microbiota compound over generations. Nature* 2016, 529(7585):212.
- 9.Louis P, Hold GL, Flint HJ: *The gut microbiota, bacterial metabolites and colorectal cancer. Nature reviews microbiology* 2014, 12(10):661.
- 10.Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, O'riordain M, Shanahan F, O'toole PW: *Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut* 2017, 66(4):633–643.
- 11.Park CH, Han DS, Oh Y-H, Lee A-r, Lee Y-r, Eun CS: *Role of Fusobacteria in the serrated pathway of colorectal carcinogenesis. Scientific reports* 2016, 6:25271.
- 12.Song H, Wang W, Shen B, Jia H, Hou Z, Chen P, Sun Y: *Pretreatment with probiotic Bifico ameliorates colitis-associated cancer in mice: Transcriptome and gut flora profiling. Cancer science* 2018, 109(3):666–677.
- 13.Hu J, Luo H, Wang J, Tang W, Lu J, Wu S, Xiong Z, Yang G, Chen Z, Lan T: *Enteric dysbiosis-linked gut barrier disruption triggers early renal injury induced by chronic high salt feeding in mice. Experimental & molecular medicine* 2017, 49(8):e370.
- 14.Thompson KJ, Ingle JN, Tang X, Chia N, Jeraldo PR, Walther-Antonio MR, Kandimalla KK, Johnson S, Yao JZ, Harrington SC: *A comprehensive analysis of breast cancer microbiota and host gene expression. PloS one* 2017, 12(11):e0188873.
- 15.Abed J, Emgård JE, Zamir G, Faroja M, Almogy G, Grenov A, Sol A, Naor R, Pikarsky E, Atlan KA: *Fap2 mediates Fusobacterium nucleatum colorectal adenocarcinoma enrichment by binding to tumor-expressed Gal-GalNAc. Cell host & microbe* 2016, 20(2):215–225.
- 16.Csardi G, Nepusz T: *The igraph software package for complex network research. InterJournal, Complex Systems* 2006, 1695(5):1–9.
- 17.Edgar RC: *Search and clustering orders of magnitude faster than BLAST. Bioinformatics* 2010, 26(19):2460–2461.
- 18.McMurdie PJ, Holmes S: *phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one* 2013, 8(4):e61217.
- 19.Chang WSW, Chou RH, Wu CW, Chang JY: *Human tissue kallikreins as prognostic biomarkers and as potential targets for anticancer therapy. Expert Opinion on Therapeutic Patents* 2007, 17(10):1227–1240.
- 20.Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD: *Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol* 2013, 79(17):5112–5120.

- 21.Yeom CH, Cho MM, Baek SK, Bae OS: *Risk factors for the development of Clostridium difficile-associated colitis after colorectal cancer surgery. Journal of the Korean Society of Coloproctology* 2010, 26(5):329.
- 22.Sfanos KS, Yegnasubramanian S, Nelson WG, De Marzo AM: *The inflammatory microenvironment and microbiome in prostate cancer development. Nature Reviews Urology* 2018, 15(1):11–24.
- 23.Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM, Ojesina AI, Jung J, Bass AJ, Tabernero J: *Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome research* 2012, 22(2):292–298.
- 24.Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW: *Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/ β -catenin signaling via its FadA adhesin. Cell host & microbe* 2013, 14(2):195–206.
- 25.Castellari M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J, Barnes R, Watson P, Allen-Vercoe E, Moore RA: *Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome research* 2012, 22(2):299–306.
- 26.Poovorawan K, Chatsuwan T, Lakananurak N, Chansaenroj J, Komolmit P, Poovorawan Y: *Shewanella haliotis associated with severe soft tissue infection, Thailand, 2012. Emerging infectious diseases* 2013, 19(6):1019.
- 27.Tan C-K, Lai C-C, Kuar W-K, Hsueh P-R: *Purulent pericarditis with greenish pericardial effusion caused by Shewanella algae. Journal of clinical microbiology* 2008, 46(8):2817–2819.
- 28.Wu M, Wu Y, Deng B, Li J, Cao H, Qu Y, Qian X, Zhong G: *Isoliquiritigenin decreases the incidence of colitis-associated colorectal cancer by modulating the intestinal microbiota. Oncotarget* 2016, 7(51):85318.
- 29.Plant N: *The human cytochrome P450 sub-family: transcriptional regulation, inter-individual variation and interaction networks. Biochimica et Biophysica Acta (BBA)-General Subjects* 2007, 1770(3):478–488.
- 30.Bethke L, Webb E, Sellick G, Rudd M, Penegar S, Withey L, Qureshi M, Houlston R: *Polymorphisms in the cytochrome p 450 genes CYP1A2, CYP1B1, CYP3A4, CYP3A5, CYP11A1, CYP17A1, CYP19A1 and colorectal cancer risk. BMC cancer* 2007, 7(1):123.
- 31.Martinez C, Garcia-Martin E, Pizarro R, Garcia-Gamito F, Agúndez J: *Expression of paclitaxel-inactivating CYP3A activity in human colorectal cancer: implications for drug therapy. British journal of cancer* 2002, 87(6):681.
- 32.Plewka D, Plewka A, Szczepanik T, Morek M, Bogunia E, Wittek P, Kijonka C: *Expression of selected cytochrome P450 isoforms and of cooperating enzymes in colorectal tissues in selected pathological conditions. Pathology-Research and Practice* 2014, 210(4):242–249.

- 33.Gervasini G, García-Martín E, Ladero JM, Pizarro R, Sastre J, Martínez C, García M, Diaz-Rubio M, Agúndez JA: *Genetic variability in CYP3A4 and CYP3A5 in primary liver, gastric and colorectal cancer patients. BMC cancer* 2007, 7(1):118.
- 34.Cho LY, Yang JJ, Ko K-P, Ma SH, Shin A, Choi BY, Han DS, Song KS, Kim YS, Chang S-H: *Genetic susceptibility factors on genes involved in the steroid hormone biosynthesis pathway and progesterone receptor for gastric cancer risk. PloS one* 2012, 7(10):e47603.
- 35.Doyle LA, Ross DD: *Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). Oncogene* 2003, 22(47):7340.
- 36.Liu HG, Pan YF, You J, Wang OC, Huang KT, Zhang XH: *Expression of ABCG2 and its significance in colorectal cancer. Asian Pac J Cancer Prev* 2010, 11(4):845–848.
- 37.Tsunoda S, Okumura T, Ito T, Kondo K, Ortiz C, Tanaka E, Watanabe G, Itami A, Sakai Y, Shimada Y: *ABCG2 expression is an independent unfavorable prognostic factor in esophageal squamous cell carcinoma. Oncology* 2006, 71(3–4):251–258.
- 38.Hirano M, Maeda K, Matsushima S, Nozaki Y, Kusuhara H, Sugiyama Y: *Involvement of BCRP (ABCG2) in the biliary excretion of pitavastatin. Molecular pharmacology* 2005, 68(3):800–807.
- 39.Enokizono J, Kusuhara H, Sugiyama Y: *Involvement of breast cancer resistance protein (BCRP/ABCG2) in the biliary excretion and intestinal efflux of troglitazone sulfate, the major metabolite of troglitazone with a cholestatic effect. Drug metabolism and disposition* 2007, 35(2):209–214.
- 40.Xie G, Wang X, Huang F, Zhao A, Chen W, Yan J, Zhang Y, Lei S, Ge K, Zheng X: *Dysregulated hepatic bile acids collaboratively promote liver carcinogenesis. International journal of cancer* 2016, 139(8):1764–1775.
- 41.Ocvirk S, O'Keefe SJ: *Influence of bile acids on colorectal cancer risk: potential mechanisms mediated by diet-gut microbiota interactions. Current nutrition reports* 2017, 6(4):315–322.

Figures

Figure 1

Alpha and beta diversity of CRC and normal samples. (A. Rarefaction curves of all samples sequenced, indicating the number of OTUs observed with different sequencing depths. B. Principal component analysis (PCA). C. Boxplots showing alpha diversity in CRC and normal samples using different metrics (Shannon, Simpson, Chao1, and PD_whole_tree indices))

Figure 2

Differential microbiota distribution at phylum (A, B) and genus (C, D) level between normal and tumor samples.

Figure 3

Identification of the different OTUs and KEGG pathway enrichment. (A: The volcano plot of different OTUs. B: The heat map of different OTUs. Green represents low expression, and red indicates high expression. C: KEGG pathways enrichment analysis. Red refers to high expression, while green refers to low expression.)

KEGG pathways enrichment analysis. (A: KEGG pathways analysis of DEGs up-regulated in CRC. B: KEGG pathways analysis of down-regulated DEGs. C: Venn diagrams of the KEGG pathways between the different OTUs and DEGs.)

Figure 6

The Kaplan-Meier curves for CYP3A4 (A) and ABCG2 (B).