

RESEARCH ARTICLE

Open Access

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



Qian Zhang^{1,2}, Huan Zhao¹, Dedong Wu², Dayong Cao³ and Wang Ma^{1*}

Abstract

Background: The dysregulation of gut microbiota is pivotal in colorectal carcinogenesis. Meanwhile, altered gut microbiome may affect the development of intestinal diseases through interaction with the host genes. However, the synergy between the altered gut microbiota composition and differential expression of specific genes in colorectal cancer (CRC) remains elusive. Thus, we integrated the data from 16S rRNA gene sequences and RNA sequences to investigate the potential relationship between genes and gut microbes in patients with CRC.

Results: Compared with normal samples, the presence of *Proteobacteria* and *Fusobacteria* increased considerably in CRC samples; conversely, the abundance of *Firmicutes* and *Spirochaetes* decreased markedly. In particular, the genera *Fusobacterium*, *Catenibacterium*, and *Shewanella* were only detected in tumor samples. Meanwhile, a closely interaction between *Butyricimonas* and *Clostridium* was observed in the microbiome network. Furthermore, a total of 246 (differentially expressed genes) DEGs were identified between tumor and normal tissues. Both DEGs and microbiota were involved in bile secretion and steroid hormone biosynthesis pathways. Finally, genes like cytochrome P450 family 3 subfamily A member 4 (*CYP3A4*) and ATP binding cassette subfamily G member 2 (*ABCG2*) enriched in these two pathways were connected with the prognosis of CRC, and CRC patients with low expression level of *CYP3A4* and *ABCG2* had longer survival time.

Conclusion: Identifying the complicated interaction between gut microbiota and the DEGs contributed to further understand the pathogenesis of CRC, and these findings might 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 that diet is regarded as a notable risk factor of CRC [3, 4]. Chan et al. revealed

that excessive intake of red meat and animal fat might increase the risk of CRC [5]. It is reported that diet can modulate the composition of gut microbiota which serves a crucial role in maintaining intestinal homeostasis and is involved in the regulation of host inflammation and immune responses [6]. Different members of the intestinal microbiota can jointly regulate the host immune and metabolic systems, subsequently producing carcinogenic or anticancer substances [7, 8]. Lately, accumulating studies have reported the role of intestinal microbiota in health and disease [9, 10]. Flemer et al. proposed that the disharmony of intestinal microbiota

* Correspondence: doctormawang@126.com

¹Department of Oncology, The First Affiliated Hospital of Zhengzhou University, No. 1 Eastern Jianshe Road, Erqi District, Zhengzhou 450000, Henan, China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

might influence the pathogenesis of CRC [11]. Hence, applying strategies to manage the composition of gut microbiota to promote recovery of a favorable microbiota community may be feasible in the treatment of patients with CRC.

Additionally, it has been revealed that an altered gut microbiome may affect the development of intestinal diseases through interaction with the innate immune system and other host genes [12]. Huang and colleagues demonstrated that the possible pathogenic flora of colitis-related cancer was connected with the C-X-C motif receptor 2 (CXCR2) signaling axis during cancer progression [13]. Imhann et al. showed that the interaction between host genetics and intestinal microbiota was the basis of the occurrence and clinical manifestations of inflammatory bowel disease [14]. Moreover, enteric microbiota dysbiosis and genetic abnormalities led to disruption of the intestinal barrier, thus triggering early kidney injury in mice [15]. Furthermore, a large number of studies have shown that dysbiosis of microbiota contributes to cancer susceptibility by affecting multiple pathways. A previous study indicated that gut microbes induced epithelial-to-mesenchymal transitions through various signaling pathways, such as Wnt- and TGF β -signaling pathway, resulting in invasion and metastasis of CRC cells [16]. These findings emphasize that these specific pathways can influence the development of cancer through altering gene expression and microbiota composition [17]. However, pathways that involved in the altered microbiota composition and differential gene expression in CRC have not been well identified. Meanwhile, specific genes that may disrupt the gut microbial composition and ultimately cause CRC remain not well recognized.

With the development of biological information technology, high-throughput sequencing has been widely employed to investigate the pathogenesis of cancer. Meanwhile, multi-omics (metagenomics, transcriptomics, and proteomics) are rapidly expanding our knowledge of the gut microbiota in health and disease. Thompson et al. explored the correlation between the gut bacterial groups and host genes expression in patients with breast cancer by using RNA sequencing data and 16S ribosomal sequencing data [18]. However, a global investigation of the association of tumor gene expression with tumor metagenomics in CRC has not been well described. The obtainment, analysis, and comparison of multi-omics data are challenging tasks. Some actual difficulties must be considered. For example, samples collection, as well as RNA and DNA extraction are challenging; in addition, sequencing costs and run times of metagenomics and transcriptomics analyses are substantially higher and longer than single sequencing analysis. Thus, we used the sequencing data from public

databases for analysis. Normally, multi-omics studies should analyze the sequencing results from tumors and matched normal tissues that were from same patients [19, 20]. Unfortunately, analyzing the sequencing data of the same patients were difficult due to our data came from two databases. Therefore, we made an initial exploration of the relationship between microbial composition and gene expression. In the present study, the data of 16S rRNA sequencing and mRNA sequencing were downloaded, followed by identification of the significantly altered gut microbiota and differentially expressed genes (DEGs) between CRC sample and normal sample. In addition, Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment pathway analysis of differential OUT and DEGs was respectively performed, followed by integration to identify the co-enrichment pathways. Finally, survival analysis of DEGs involved in co-enrichment pathways was further conducted. This study described the relationship between intestinal flora and gene expression in patients with CRC, and provided valuable information for diagnosis and treatment of CRC.

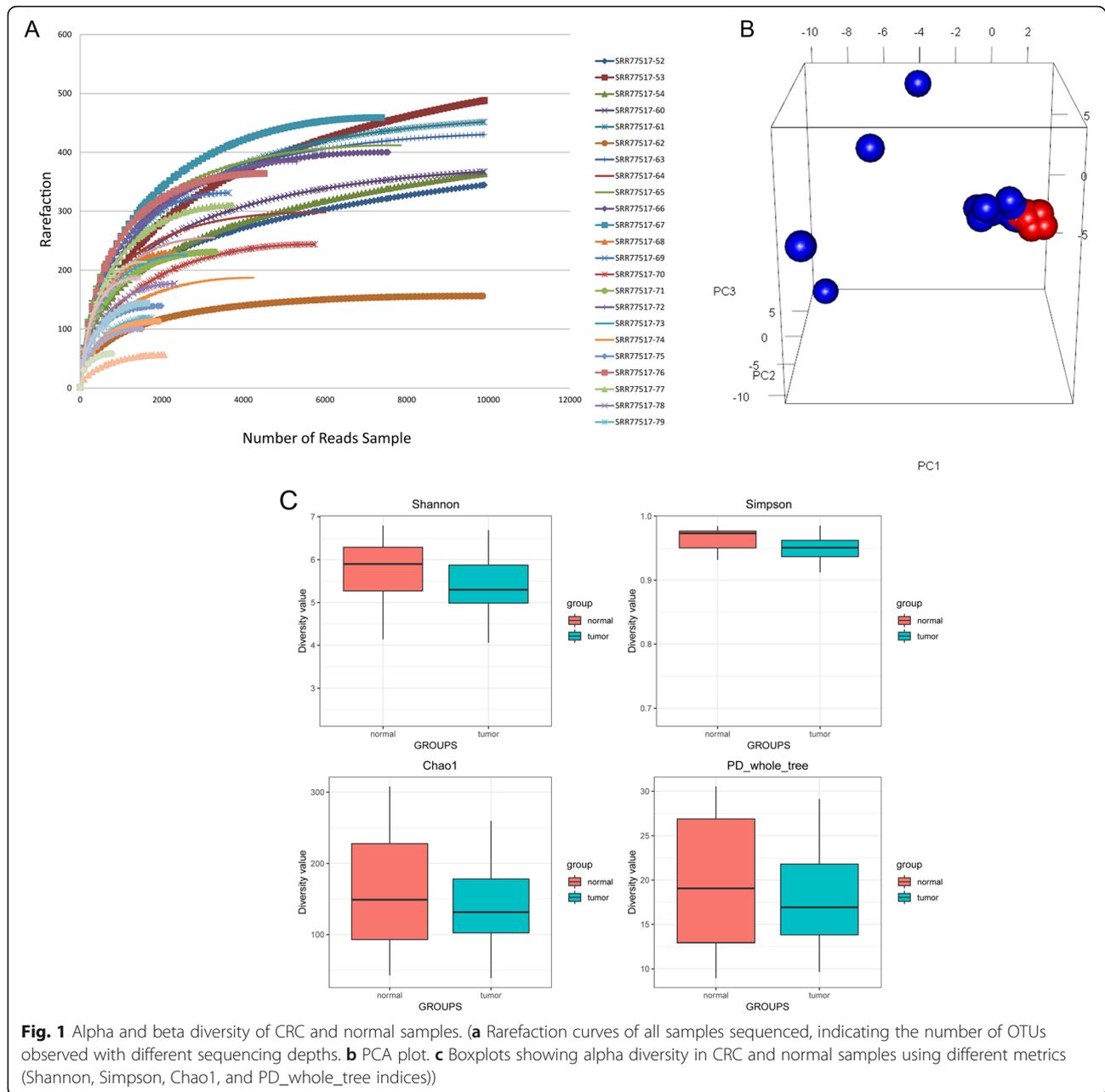
Results

Rarefaction curve and diversity analysis

Rarefaction curves of most samples tend to be flat, suggesting that the amount and depth of sequencing data were reasonable (Fig. 1a). The principal component analysis (PCA) plot showed differences in the gut microbiota composition between CRC and normal samples (Fig. 1b). In addition, we calculated the alpha diversity indices to estimate the diversity of gut microbiota. In the normal samples, the alpha diversity indices were 5.4, 0.9, 149.4, and 18.6 for Shannon, Simpson, chao1, and PD_{whole_tree}, while these indices for tumor samples were 5.5, 0.9, 147.7, and 18.5. No differences were observed among these four indicators between the two groups. Thus, the alpha diversity results revealed that there was no significant difference between the normal samples and tumor samples (Fig. 1c).

Taxonomic composition

A total of 13 different phyla were detected from these two groups. The results are shown in Fig. 2a and b. At the phylum level, the microbiota of the tumor and control samples shared 12 phyla, and the members of *Thaumarchaeota* were only identified in the control group. In addition, four dominant phyla were detected among all the samples, including *Firmicutes* (50.7% in normal group and 46.5% in tumor group), *Proteobacteria* (15.9% in normal group and 21.7% in tumor group), *Bacteroidetes* (21.5% in normal group and 22.4% in tumor group), and *Actinobacteria* (4.3% in normal group and 4.1% in tumor group). Compared with the normal group,

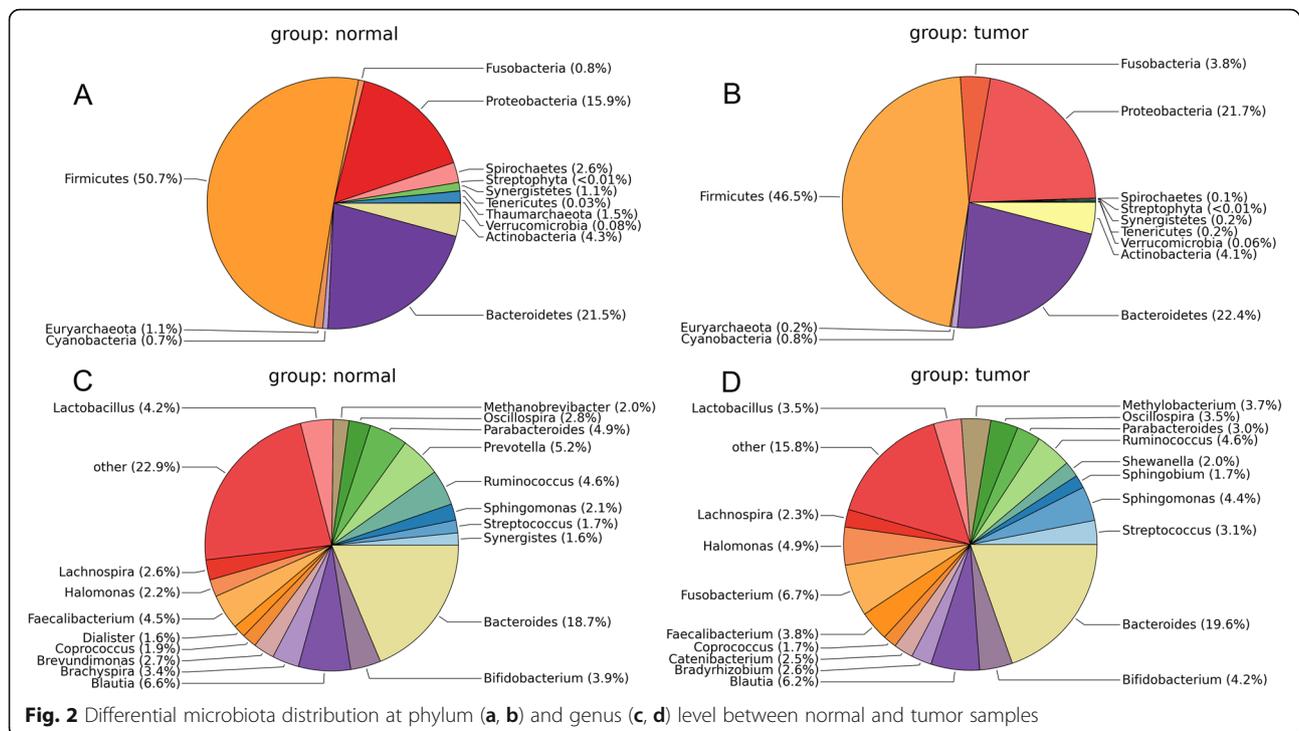


the relative abundance of *Firmicutes*, *Spirochaetes*, and *Euryarchaeota* in tumor group was decreased, while the abundance of *Fusobacteria*, *Proteobacteria*, and *Bacteroidetes* was increased. Meanwhile, the gut microbiota of the two groups shared 24 genera. The results demonstrated that the level of *Bacteroides* was advantaged across two groups (Fig. 2c and d). Specifically, four dominant genera, including *Bacteroides* (18.7%), *Blautia* (6.6%), *Prevotella* (5.2%), and *Parabacteroides* (4.9%), were observed in normal group. Meanwhile, the abundance of *Bacteroides* (19.6%), *Fusobacterium* (6.7%), and *Blautia* (6.2%) was higher in the tumor group compared

with normal group. Notably, *Fusobacterium* (6.7%), *Catenibacterium* (2.5%), and *Shewanella* (2.0%) were specifically detected in the tumor samples.

Differentially enriched operational taxonomic units (OTU) and pathway enrichment analysis

A total of 66 differentially enriched OTUs were identified between CRC samples and normal samples, including 22 up-regulated and 44 down-regulated OTUs. The differential OTUs were visualized by using volcano plot and hierarchical clustering (Fig. 3a and b). There was significant difference between tumor and normal



samples. The results indicated that the abundance of *Actinobacteria* and *Fusobacteria* in tumor samples was significantly higher than that in normal samples. Compared with normal samples, the relative abundance of *Firmicutes* and *Proteobacteria* was observably lower in tumor samples. Meanwhile, the differential OTUs were significantly involved in 67 pathways (Fig. 3c). The results implicated that these OTUs mainly participated in CRC, MAPK signaling pathway, and p53 signaling pathway.

Network analysis of microbiome

Network analysis of differential OTUs to reveal the relationship among microbes. The network composed of 38 nodes and 55 edges was constructed to describe the complex relationships of microbiome (Fig. 4). The 35 genera were from eight bacterial phyla, including 24 genera from *Firmicutes* (57.55%), one genus from *Crenarchaeota* (10.40%), two genera from *Bacteroidetes* (10.39%), two genera from *Actinobacteria* (8.87%), six genera from *Proteobacteria* (6.52%), one genus from *Cyanobacteria* (5.21%), one genus from *Euryarchaeota* (1.03%), and one genus from *Spirochaetes* (0.03%). Bacteria from *Bacteroides*, *Phascolarctobacterium*, and *Delftia* showed interaction with seven, six, and four genera, respectively. Specially, for the genera related to CRC, *Butyricimonas* showed closely connection with *Clostridium*.

DEG identification and pathway enrichment analysis

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

Integrated analysis

We integrated the pathways enriched by DEGs and OTUs, two overlapping pathways were obtained, namely, bile secretion and steroid hormone biosynthesis. These two pathways could affect CRC not only at transcriptome levels, but also at the intestinal microbiota level (Fig. 5c). Additionally, a total of 11 up-regulated DEGs, including aquaporin 8 (*AQP8*), carbonic anhydrase 2 (*CA2*), solute carrier family 4 member 4 (*SLC4A4*), ATP binding cassette subfamily G member 2 (*ABCG2*), cytochrome P450 family 3 subfamily A member 4 (*CYP3A4*), ATPase Na⁺/K⁺ Transporting subunit alpha 2 (*ATPIA2*), ATP binding cassette subfamily B Member 11 (*ABCB11*), hydroxy-delta-5-steroid dehydrogenase, 3 beta- And steroid delta-isomerase 2 (*HSD3B2*), UDP

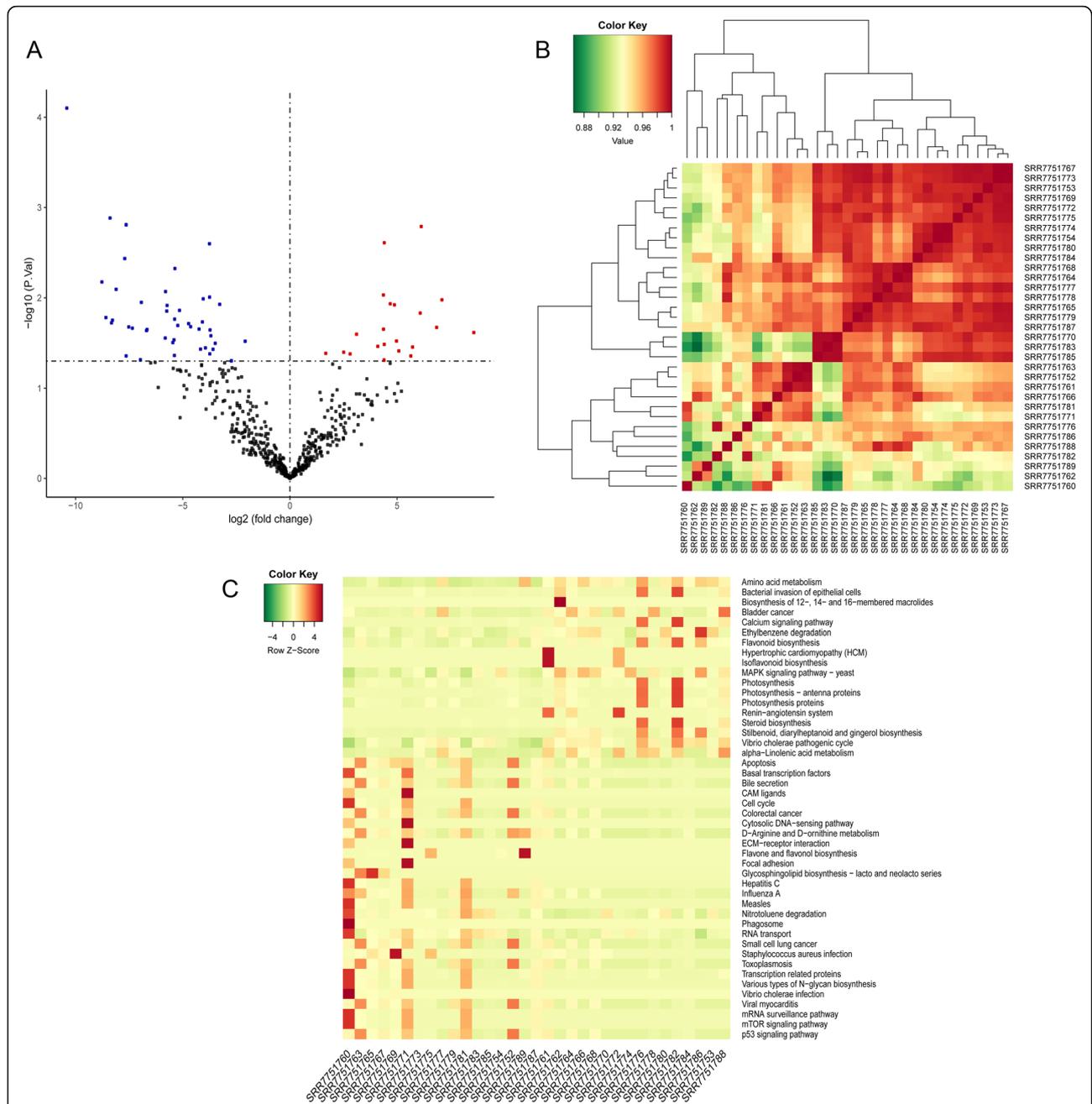
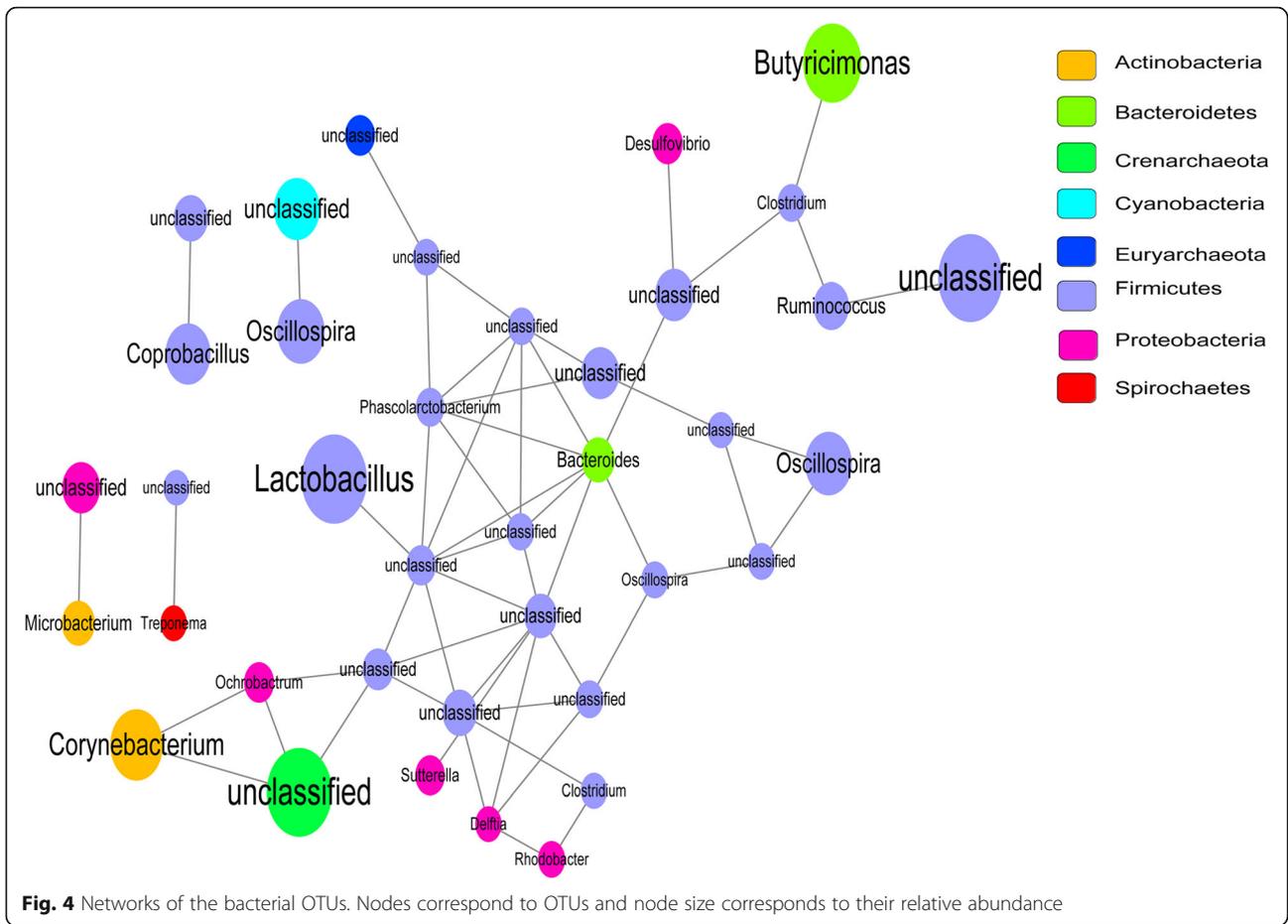


Fig. 3 Identification of the different OTUs and KEGG pathway enrichment. **(a)** The volcano plot of different OTUs. **(b)** The heat map of different OTUs. Green represents low expression, and red indicates high expression. **(c)** KEGG pathways enrichment analysis. Red refers to high expression, while green refers to low expression)

glucuronosyltransferase family 1 member A8 (*UGT1A8*), UDP glucuronosyltransferase family 2 member B10 (*UGT2B10*), and UDP glucuronosyltransferase family 1 member A3 (*UGT1A3*), were significantly involved in these two pathways.

Survival analysis

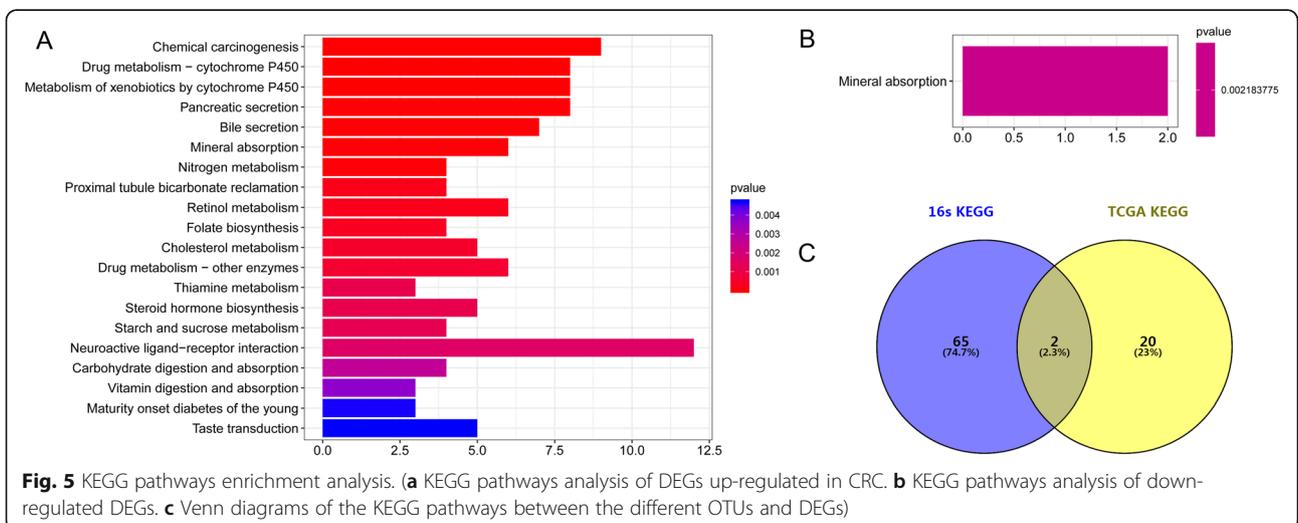
Survival analysis of above DEGs was performed using Kaplan-Meier (K-M) method. Among these candidate genes, three genes, including *SLC4A4*, *CYP3A4*, and *ABCG2*, were significantly related to the prognostic of CRC. As shown in Fig. 6, CRC patients with low expression level of *CYP3A4* and *ABCG2* had longer survival time.

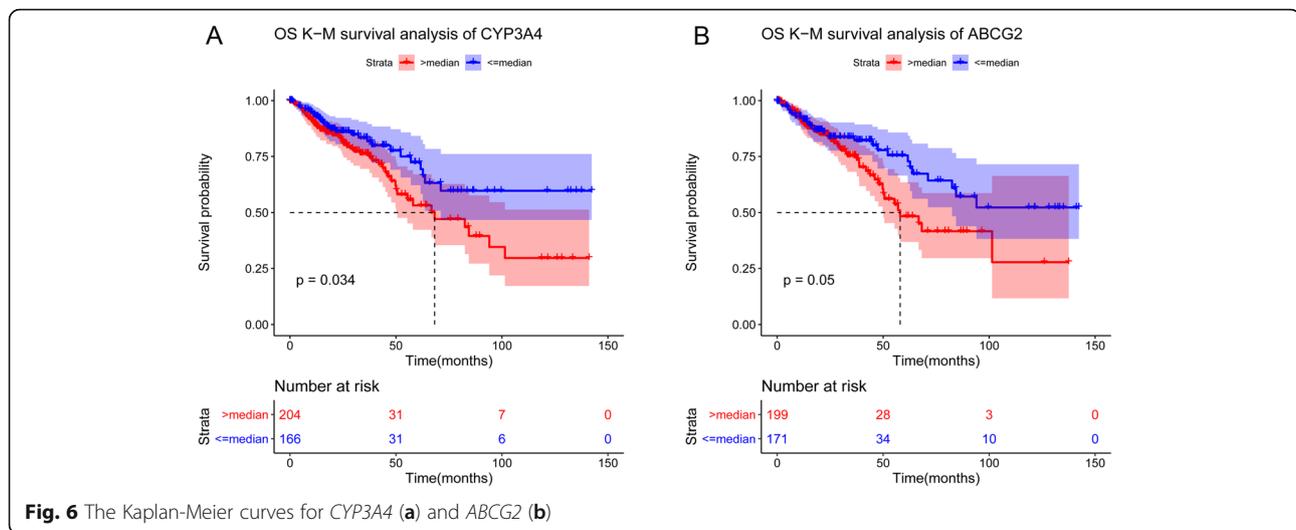


Discussion

Genes are known to regulate the pathogenesis of CRC and are associated with the survival outcomes of patients. In addition, host genes can also regulate the growth of microbiota, influencing the composition of the

intestinal microbial community [21]. Thus, we investigated the association between mRNA expression and microbiome composition in CRC tissues. In this study, we found that the abundance of *Proteobacteria* and *Fusobacteria* increased in the tumor samples, whereas





that of *Firmicutes* decreased. *Fusobacterium*, *Catenibacterium*, and *Shewanella* were specifically detected in the tumor tissues. Additionally, two co-enrichment pathways, including bile secretion and steroid hormone biosynthesis, were associated with both DEGs and differential OTU in CRC. Furthermore, CRC patients with low expression level of *CYP3A4* and *ABCG2* had longer survival time.

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. Aleksandar et al. reported that *Fusobacterium* were enriched in carcinomas, whereas the abundance of *Firmicutes* significantly decreased in tumors [23], which was consistent with our findings. Specifically, *Fusobacterium* was markedly enriched in the tumor tissues. A previous study suggested that *Fusobacterium* expressed the virulence factor FadA and activated the WNT signaling pathways, thus promoting growth of CRC [24]. *Fusobacterium* has been shown to inhibit immune responses of CRC tumors [25]. Meanwhile, *Fusobacterium* species are also known to induce host proinflammatory responses and possess virulence [26]. Our findings were supported by the above reports, and highlighted that the clinical relevance of *Fusobacterium* in the development of CRC should be addressed in further studies. In this study, we found *Shewanella* was another genus particularly enriched in the tumor tissues. *Shewanella* was revealed to cause pulmonary and blood infections [27], and it could raise purulent pericarditis with greenish pericardial effusion [28]. Wang et al. suggested that increased *Shewanella* algae was a biomarker of colorectal adenoma [17]. However, its role in CRC tumor progression was not well-defined, further detailed studies were required to verify our findings. Interestingly, a closely interaction between

Butyricimonas and *Clostridium* was observed in the microbiome network. Wu et al. confirmed that *Butyricimonas* was only detected in CRC group from mouse model [29]. Meanwhile, *Clostridium* was a risk factor for the development of CRC [30]. A previous study suggested that transplanting *Clostridium symbiosum* to germ-free nutrition-deficient mice might promote protein synthesis in local gut epithelium, which might be considered as potential supporter to the development of carcinogenesis [31]. Additionally, *Clostridium symbiosum* was a promising biomarker for early and noninvasive detection of CRC [32]. Taken together, we hypothesized that this relationship might play a pivotal role in the pathogenesis of CRC, and these two microbiotas could be utilized as predictors of CRC diagnosis.

Two pathways were significantly enriched in both DEGs and OTUs, namely, bile secretion and steroid hormone biosynthesis pathways. Moreover, genes like *CYP3A4* and *ABCG2* were involved in these two pathways and were also associated with the prognosis of CRC. *CYP3A4* encodes a member of the cytochrome P450 superfamily of enzyme [33]. A recent study demonstrated that the genotoxicity of the carcinogen was influenced by cytochrome P450 enzyme system, and *CYP3A* was highly expressed in colonic tissue [34]. Zhang et al. pointed out that *CYP3A4* might be recognized as an anti-cancer target, indicating it could be used as a potential molecular marker for predicting and treating CRC [21]. Additionally, *CYP3A* was involved in the metabolism of carcinogen and was associated with inactivation of anticancer drugs [35]. A prior study implicated that the *CYP3A* mRNA transcripts were present in the human colorectal epithelium and CRC cell lines [36]. Similarly, we also showed that *CYP3A4* served an important role in the development of CRC. Furthermore, a study about the relationship between cytochrome P450

and CRC indicated that several P450s were the independent markers of CRC prognosis [34], which were consistent with our results. Pathway enrichment analysis revealed that *CYP3A4* was involved in steroid hormone biosynthesis pathway. The association between steroid hormone biosynthesis pathway and cancer development has been confirmed. A 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 [37]. Steroid hormones also played a central role in the progression of prostate cancer, and conversion of adrenal androgen precursors and other steroid-producing pathways might contribute to tumor progression and resistance to therapy [32]. Therefore, we speculated that *CYP3A4* might serve important role in the pathogenesis of CRC via affecting steroid hormone biosynthesis pathway, as well as it might be regarded as a prognostic biomarker and a therapeutic target for CRC.

ABCG2 is a member of the superfamily of ATP-binding cassette (ABC) transporter protein, which can induce drug resistance and treatment failure in tumor tissues [38]. Liu et al. suggested that *ABCG2* was highly expressed in CRC and it might be involved in progression and metastasis of advanced malignancy cancer [39], which was in line with our finding. Meanwhile, higher *ABCG2* mRNA expression also represented an unfavorable prognostic factor of esophageal squamous cell carcinoma [40]. Thus, we speculated that *ABCG2* might be regarded as a prognostic marker of CRC. In this analysis, we observed that *ABCG2* was involved in bile secretion pathway. *ABCG2* is a hepatobiliary efflux transporter and is involved in the biliary excretion of sulfate conjugates [41] and troglitazone sulfate of therapeutics [42]. Genes such as urothelial cancer associated 1 (*UCA1*) which participated in bile secretion pathway were over-expressed in hepatocellular carcinoma (HCC) tissues [43]. Additionally, bile acids role as tumor promoters have been confirmed by extensive experiments [44, 45]. From the above, *ABCG2* might play role in the pathogenesis of CRC via bile secretion pathway. Although the mechanism of *ABCG2* in CRC progression remained unclear, the importance of *ABCG2* in CRC should not be underestimated.

Taken together, our study highlighted that changes in gene expression and microbiota composition were linked to the specific pathways. We showed that differential expression of genes might cause the alteration of the bile secretion and steroid hormone biosynthesis in CRC tissues, thereby changing the abundance and composition of intestinal microbiota and eventually might trigger the occurrence of cancer. Our results have to be interpreted in light of some limitation. In this analysis, CRC related data were obtained from two different databases and it

was not from matched samples. An integrated study based on multi-omics data from the same patient will be our focus in the future. In addition, our study was based on bioinformatics analyses of the datasets from public databases, and further experimental studies and clinic trial must be conducted to validate and strengthen our results.

Conclusion

By integrating the results of microbiome and transcriptome, we revealed a potential relationship between the genes and gut microbes in patients with CRC, and gained better insight into the pathogenesis and prognosis of CRC. Our study might provide a new perspective for the diagnosis and treatment of CRC, and these genes and microbiota might serve as potential diagnostic markers and therapeutic targets for CRC.

Methods

Data resource

The intestinal microbiota data with the number SRP158779 (<http://www.ncbi.nlm.nih.gov/sra/SRP158779>) were retrieved from the NCBI Sequence Read Archive (SRA) database. This dataset contained 38 samples from 19 patients, including 19 CRC tumors samples and 19 paired non-neoplastic tissues. DNA was extracted and purified using the QIAGEN DNA extraction kit. The library was generated based on the V3-V4 region of the 16S rRNA, and then was sequenced on an Illumina HiSeq 2000 platform by using paired-end sequencing. In addition, the mRNA sequencing data (level 3, raw counts) and clinical characteristics of CRC were downloaded from The Cancer Genome Atlas (TCGA) database (<http://firebrowse.org/>) for a total of 422 samples (371 CRC tumors and 51 normal samples).

OTU cluster and taxonomy classification

The raw data were converted to fastq format using fastq-dump software (parameter: split-3). Raw data containing low-quality reads that could affect the results of following analysis. Thus, quality control was carried out to obtain high-quality clean reads. Quantitative Insights Into Microbial Ecology (QIIME) (version 1.4.0) [25] software was employed to perform further analysis. Primarily, the paired-end reads were assigned to samples based on their unique barcodes, and then the amplified primers were excised and chimera sequences were removed. Additionally, the clean reads were used for diversity analysis and taxonomic composition based on the Greengenes database (release 13.5, <http://greengenes.secondgenome.com/>) [46]. Filtered sequences were clustered into OTUs at 97% similarity using UCLUST (version 1.2.22q, <http://www.drive5.com/usearch/>) [47]. Thereafter, the sequence with the highest abundance in

each OTU was selected as the representative sequence of this OTU. Ultimately, based on the number of sequences included in each OTU, the OTU table abundance in each sample was constructed. In addition, taxonomic assignments of OTUs that reached 97% similarity level were performed using RDP classifier (<https://sourceforge.net/projects/rdp-classifier/>) [48] by comparing with the Greengene database (Release 13.5, <http://greengenes.secondgenome.com/>) [49].

Alpha and beta diversity analysis

Abundance and diversity of microbial communities could be reflected by alpha diversity. The Shannon, Simpson, Chao1, and PD_whole_tree indices were calculated to estimate alpha diversity. Concretely, the Shannon and Simpson indices were used to represent the community diversity, and chao 1 indicated the community richness, as well as PD_whole_tree symbolized the phylogenetic diversity. All the statistical analyses were performed using the R phyloseq package [50]. Additionally, the rarefaction curve was plotted to reveal whether the amount of sequencing data was reasonable. Beta diversity analysis could examine the similarity of community structure among different samples. In the present study, beta diversity was calculated by the QIIME software and cluster analysis was conducted by PCA; thereafter, RGL package in R software was applied to visualize the results.

Screening of differentially OTU

The OTU data were preprocessed by using trimmed mean of M values (TMM) method from edgeR package in R software [51]. Subsequently, the differential analysis was carried out using the quasi-likelihood (QL) F-test from edgeR package. P value < 0.05 was considered as statistically significant.

Network analysis of microbiome

In order to further explore the relationship among differential OTUs, a reciprocal interaction network among microbiome was constructed. Based on the differential OTUs, the correlation coefficient matrix between OTUs was calculated by using R with igraph (version 1.2.2) and psych (version 1.8.4) packages. The pairs with both p value < 0.05 and $|r| > 0.6$ were selected to construct network, and was visualized by using Cytoscape (version 2.8).

Prediction of the function of differential OTUs

Phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) is a computational approach to predict the function of bacteria according to the obtained 16S rRNA gene sequences. In this study, PICRUSt program was used to predict the

functional profile of the microbial communities. The main procedures were displayed as following: 1) based on the full-length sequence of 16S rRNA gene of the measured microbial genome, the genetic functional profiles of their common ancestors of differential OTUs were predicted; 2) the functional profiles of other untested species in the Greengenes 16S rRNA gene full-length sequence database were deduced, and then the genetic function prediction spectrum of the entire lineage of archaea and bacteria domain was constructed; and 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.

Identification of DEGs and pathway enrichment analysis

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 DEGs were selected with $|\log FC| > 1.5$ and false discovery rate (FDR) < 0.05 .

The pathway enrichment analysis of DEGs was carried out by using clusterProfiler [30] of R package. Thereafter, gene count ≥ 2 and p value < 0.05 were set as the cut-off criterion.

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 shared between the two sets of data were selected, and the obtained pathways were regarded as CRC-related functions affected by intestinal microbiota.

Survival analysis

The prognosis outcomes of CRC patients, including overall survival (OS) and survival status, were obtained from TCGA database. The survival analysis of genes involved in obtained pathways was performed. All samples were divided into high and low expression groups according to the median expression level of genes. The Kaplan–Meier survival curves were plotted and statistical significance was assessed using the log-rank tests. $P < 0.05$ was set as the cut-off criteria for statistical significance.

Abbreviations

CRC: Colorectal cancer; NCBI: National Center for Biotechnology Information; SRA: Sequence Read Archive; OTUs: Operational taxonomic units; TCGA: The Cancer Genome Atlas; DEGs: Differentially expressed genes; CXCR2: C-X-C motif receptor 2; KEGG: Kyoto Encyclopedia of Genes and Genomes; QIIME: Quantitative Insights Into Microbial Ecology; PCA: Principal component analysis; OS: Overall survival; APOB: Apolipoprotein B; CAT: Carbonic anhydrase 1; ANGPTL5: Angiopoietin like 5; SHISA7: Shisa family member 7; SLC4A4: Solute carrier family 4 member 4; CYP3A4: Cytochrome

P450 family 3 subfamily A member 4; *ABCG2*: ATP binding cassette subfamily G member 2; *ABC*: ATP-binding cassette

Acknowledgements

None.

Authors' contributions

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

Funding

None.

Availability of data and materials

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

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Oncology, The First Affiliated Hospital of Zhengzhou University, No. 1 Eastern Jianshe Road, Erqi District, Zhengzhou 450000, Henan, China. ²Department of Oncology, The First People's Hospital of Zhengzhou, Zhengzhou 450004, Henan, China. ³Department of Burns, The First People's Hospital of Zhengzhou, Zhengzhou 450004, Henan, China.

Received: 26 April 2020 Accepted: 9 August 2020

Published online: 13 October 2020

References

- Favoriti P, Carbone G, Greco M, Pirozzi F, Pirozzi REM, Corcione F. Worldwide burden of colorectal cancer: a review. *Updat Surg*. 2016;68(1):7–11.
- Lynch HT, De la Chapelle A. Hereditary colorectal cancer. *N Engl J Med*. 2003;348(10):919–32.
- Baena R, Salinas P. Diet and colorectal cancer. *Maturitas*. 2015;80(3):258–64.
- 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. *Br J Cancer*. 2016;115(7):862.
- 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.
- Koopman M, El Aidy S. Depressed gut? The microbiota-diet-inflammation triad in depression. *Curr Opin Psychiatry*. 2017;30(5):369–77.
- Shapiro H, Thaiss CA, Levy M, Elinav E. The cross talk between microbiota and the immune system: metabolites take center stage. *Curr Opin Immunol*. 2014;30:54–62.
- Sekirov I, Russell SL, Antunes LCM, Finlay BB. Gut microbiota in health and disease. *Physiol Rev*. 2010;90(3):859–904.
- 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.
- Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol*. 2014;12(10):661.
- Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, O'riordan M, Shanahan F, O'toole PW. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut*. 2017;66(4):633–43.
- 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. *Sci Rep*. 2016;6:25271.
- 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 Sci*. 2018;109(3):666–77.
- Imhann F, Vich Vila A, Bonder MJ, Fu J, Gevers D, Visschedijk MC, Spekhorst LM, Alberts R, Franke L, van Dullemen HM, et al. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*. 2018;67(1):108–19.
- 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. *Exp Mol Med*. 2017;49(8):e370.
- Hofman P, Vouret-Craviari V. Microbes-induced EMT at the crossroad of inflammation and cancer. *Gut Microbes*. 2012;3(3):176–85.
- Elderman M, Hugenholtz F, Belzer C, Boekschoten M, de Haan B, de Vos P, Faas M. Changes in intestinal gene expression and microbiota composition during late pregnancy are mouse strain dependent. *Sci Rep*. 2018;8(1):10001.
- 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.
- Lam K, Pan K, Linnekamp JF, Medema JP, Kandimalla R. DNA methylation based biomarkers in colorectal cancer: a systematic review. *Biochim Biophys Acta*. 2016;1866(1):106–20.
- Zhu Y, Lu H, Zhang D, Li M, Sun X, Wan L, Yu D, Tian Y, Jin H, Lin A, et al. Integrated analyses of multi-omics reveal global patterns of methylation and hydroxymethylation and screen the tumor suppressive roles of HADHB in colorectal cancer. *Clin Epigenetics*. 2018;10:30.
- Wahlström A, Sayin SI, Marschall HU, Bäckhed F. Intestinal crosstalk between bile acids and microbiota and its impact on host metabolism. *Cell Metab*. 2016;24(1):41–50.
- Sfanos KS, Yegnasubramanian S, Nelson WG, De Marzo AM. The inflammatory microenvironment and microbiome in prostate cancer development. *Nat Rev Urol*. 2018;15(1):11–24.
- 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 Res*. 2012;22(2):292–8.
- 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.
- Abed J, Emgård JE, Zamir G, Feroja M, Almogly 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):215–25.
- Castellarin 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 Res*. 2012;22(2):299–306.
- Poovorawan K, Chatsuwat T, Lakananurak N, Chansaenroj J, Komolmit P, Poovorawan Y. *Shewanella haliotis* associated with severe soft tissue infection, Thailand, 2012. *Emerg Infect Dis*. 2013;19(6):1019.
- Tan C-K, Lai C-C, Kuar W-K, Hsueh P-R. Purulent pericarditis with greenish pericardial effusion caused by *Shewanella* algae. *J Clin Microbiol*. 2008;46(8):2817–9.
- 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.
- Yeom CH, Cho MM, Baek SK, Bae OS. Risk factors for the development of *Clostridium difficile*-associated colitis after colorectal cancer surgery. *J Korean Soc Coloproctol*. 2010;26(5):329–33.
- Blanton LV, Charbonneau MR, Salih T, Barratt MJ, Venkatesh S, Ilkaveya O, Subramanian S, Manary MJ, Trehan I, Jorgensen JM, et al. Gut bacteria that prevent growth impairments transmitted by microbiota from malnourished children. *Science*. 2016;351(6275).
- Xie YH, Gao QY, Cai GX, Sun XM, Sun XM, Zou TH, Chen HM, Yu SY, Qiu YW, Gu WQ, et al. Fecal *clostridium symbiosum* for noninvasive detection of early and advanced colorectal cancer: test and validation studies. *EBioMedicine*. 2017;25:32–40.
- Stage TB, Graff M, Wong S, Rasmussen LL. Dicloxacillin induces CYP2C19, CYP2C9 and CYP3A4 in vivo and in vitro. *Br J Clin Pharmacol*. 2018;84(3):510–9.
- 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.
35. 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. *Br J Cancer*. 2002;87(6):681–86.
 36. 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. *Pathol Res Pract*. 2014;210(4):242–9.
 37. 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.
 38. Doyle LA, Ross DD. Multidrug resistance mediated by the breast cancer resistance protein BCRP (ABCG2). *Oncogene*. 2003;22(47):7340–58.
 39. 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):845–8.
 40. 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):251–8.
 41. Hirano M, Maeda K, Matsushima S, Nozaki Y, Kusuha H, Sugiyama Y. Involvement of BCRP (ABCG2) in the biliary excretion of pitavastatin. *Mol Pharmacol*. 2005;68(3):800–7.
 42. Enokizono J, Kusuha 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 Metab Dispos*. 2007;35(2):209–14.
 43. Qin LT, Tang RX, Lin P, Li Q, Yang H, Luo DZ, Chen G, He Y, Li P. Biological function of UCA1 in hepatocellular carcinoma and its clinical significance: investigation with in vitro and meta-analysis. *Pathol Res Pract*. 2018;214(9):1260–72.
 44. 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. *Int J Cancer*. 2016;139(8):1764–75.
 45. Ocvirk S, O'Keefe SJ. Influence of bile acids on colorectal cancer risk: potential mechanisms mediated by diet-gut microbiota interactions. *Curr Nutr Rep*. 2017;6(4):315–22.
 46. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJ Complex Syst*. 2006;1695(5):1–9.
 47. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–1.
 48. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol*. 2007;73(16):5261–7.
 49. 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):5112–20.
 50. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One*. 2013;8(4):e61217.
 51. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26(1):139–40.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

