Multi-omic profiling reveals associations between the gut mucosal microbiome, the metabolome, and host DNA methylation associated gene expression in patients with colorectal cancer

Qing Wang1,2,3, Jianzhong Ye1,2, Daqiong Fang1,2, Longxian Lv1,2, Wenrui Wu1,2, Ding Shi1,2, Yating Li1,2, Liya Yang1,2, Xiaoyuan Bian1,2, Jingjing Wu1,2, Xianwan Jiang1,2, Kaicen Wang1,2, Qiangqiang Wang1,2, Mark P. Hodson4,5, Loïc M. Thibaut3,6, Joshua W. K. Ho7,8, Eleni Giannoulatou3,9* and Lanjuan Li1,2*

From Joint 30th International Conference on Genome Informatics (GIW) & Australian Bioinformatics and Computational Biology Society (ABACBS) Annual Conference
Sydney, Australia. 9-11 December 2019

Abstract

Background: The human gut microbiome plays a critical role in the carcinogenesis of colorectal cancer (CRC). However, a comprehensive analysis of the interaction between the host and microbiome is still lacking.

(Continued on next page)
Background

Colorectal cancer (CRC) is the third most commonly diagnosed cancer and second leading cause of cancer death in both men and women globally, leading to an estimated 551,269 annual deaths worldwide [1]. Diet plays an important role in the initiation, promotion and progression of colon carcinogenesis [2]. Epidemiological surveys indicate that approximately 80% of CRC cases in Western countries are due to dietary factors [3].

The human intestine harbours a complex microbial community comprising more than $10^{14}$ microorganisms; it carries greater than 150 times the number of genes in the human genome [4]. Diet has been shown to have a dominant impact on the structure and composition of the gut microbiome, microbial generated metabolites and host metabolism [5, 6]. Many studies have demonstrated the effect of the gut microbiome on the pathogenesis of CRC, revealing potential pathogenic bacteria such as Fusobacterium as well as beneficial bacteria such as Lactobacillales [6–8].

Nevertheless, dietary factors have also been shown to induce epigenetic changes. The most extensively studied epigenetic mechanism is DNA methylation, which involves the transfer of methyl group to the cytosine of a CpG dinucleotide and is responsible for regulating the expression of genes related with critical pathological processes [9]. Carcinogenic gene silencing may occur due to aberrant gene-specific hypermethylation at the CpG rich region (CpG island) which is located in the promoters of the genes and interfere with transcription factor binding [10, 11].

One possible mechanism underlying the effects of diet and microbiome on CRC development is a potential interaction between the microbiome and the host, whereby the colonic metabolome is impacted, leading to a subsequent alteration in host epigenetic activity and gene expression [12, 13]. Research on selected aspects of the microbiome, metabolome, host epigenome and host transcriptome have been carried out in human, animal and cell models [14]. For instance, the interaction between the microbiome and metabolome has been extensively studied, revealing various epigenome-modulation-related metabolites such as butyrate and folate [15, 16]. In addition, the contribution of commensal bacteria to epigenetic control in the host large intestine has been demonstrated by comparing conventional and germ-free mice [17]. Associations between the microbiome and differentially methylated genes have also been investigated in patients with ulcerative colitis [18]. The interplay between the microbiome, host transcripts related to adhesion molecules and fatty acid biosynthesis was strongly supported in one study of inflammatory bowel disease [19].

Despite the mounting evidence of a potential host-microbiome interaction, a comprehensive human study integrating all the aforementioned omics is still lacking. Thus, in this pilot study, we generated and analysed four types of omic data: the microbiome (16S rRNA sequencing; 36 pairs), the metabolome (untargeted GC/MS; 17 pairs), the host transcriptome (RNA-seq; 4 pairs) and the host epigenome (Infinium HumanMethylation850 BeadChip array; 4 pairs), as measured from paired tumour and adjacent normal colonic mucosa tissues.
samples obtained from CRC patients (details of the study design in Additional file 1: Figure S1).

**Results**

**Comparison of microbial composition between tumour and adjacent normal tissues**

Non-metric multidimensional scaling (NMDS) analysis based on the unweighted UniFrac distance on operational taxonomic units (OTUs) revealed that the microbial community composition of the cancerous tissues could be clearly distinguished from the non-cancerous tissues, which was confirmed by analysis of similarities (Anosim) ($p$-value = 0.002; Fig. 1a).

The microbial population structure differed between tumour and healthy tissues at multiple levels (Fig. 1b-d). At the phylum level, lower abundance of Bacteroidetes, Firmicutes and Actinobacteria, and higher abundance of Proteobacteria and Fusobacteria, was observed in tumour tissues ($p$-value > 0.05). At the family level, there was an under-representation of Bacteroidaceae, Lachnospiraceae and Ruminococcaceae and predominance of Fusobacteriaceae in tumour tissues ($p$-value > 0.05). At the genus level, the most distinctive genera were Fusobacterium, Bacteroides and Faecalibacterium, with Fusobacterium higher in tumour tissues and the other two lower in tumour tissues ($p$-value > 0.05).

To further characterise the bacterial genera driving the difference between the paired samples, the paired zero-inflated Gaussian (ZIG) mixture model was used [20]. As shown in Fig. 2, 15 taxa were identified to be the most relevant to the differences between tumour and normal tissues (Additional file 2: Table. S1). We found that taxa in family Halomonadaceae, genus Halomonas, genus *Shewanella*, genus *Fusobacterium*, genus *Fretibacterium*, genus *Peptostreptococcus* and genus *Klebsiella* were highly enriched in cancer tissues ($p$-value < 0.05).

![Fig. 1](image-url) Microbial composition comparison between tumour tissues and paired normal tissues in CRC patients. a. NMDS plot of microbiota based on unweighted UniFrac metric on OTUs. Each point represents one sample with colour indicating tumour tissues (red) and adjacent normal tissues (blue). b-d. Distribution of the top 10 most abundant taxa in the samples at phylum (b), family (c) and genus (d) level.
Fig. 2a; details in Additional file 2: Table. S1), while genus *Paeniclostridium*, species *Bacteroides eggerthii*, genus *Lactobacillus*, genus *Prevotella 2*, *Enterococcus*, species *Clostridium perfringens* and genus *Veillonella* were less abundant in cancer tissues (p-value < 0.05, Fig. 2b; Additional file 2: Table. S1).

Taken together, these results reveal the different microbial environments between the cancerous and normal colon mucosa.

**Microbiota associated with CRC-related colonic metabolites**

Seventeen pairs of samples were retained for colonic metabolome analysis, fifteen of which had matched microbiome data. In total, 309 metabolites were detected and quantified using the untargeted gas chromatography-mass spectrometry (GC-MS) approach. The score plot from orthogonal partial least squares discriminant analysis (OPLS-DA) revealed differences in metabolic features between the tissues (R²Y = 0.937; Q² = 0.658; Fig. 3a). The reliability of the model was confirmed by seven-fold cross-validation and 200 response permutation tests on Fig. 3b, which has showed that all permuted values were lower than the original value. It indicated that the original classification model predicted the class better than the other newly calculated OPLS-DA models with randomly assigned class labels (Q²_inter = ‐0.377; Fig. 3b).

Differential abundant metabolites were further identified based on Variable Influence on Projection (VIP) from the OPLS-DA model (threshold > 1) and significance level of 0.05 in the two-tailed paired t test. As shown in Fig. 3c, eight metabolites were selected, including lactic acid, 4-hydroxybutyric acid (4-HB), glutamic acid, taurine, lyxose, tetracosane, oxalic acid and D-talose. The heatmap revealed that the levels of 4-HB, lyxose, oxalic acid and D-talose were lower in tumour tissues compared to the paired normal tissues. On the other hand, the concentrations of lactic acid, glutamic acid, taurine and tetracosane were elevated in tumour tissues.

Pearson correlation analysis between changes in the level of the differentially abundant metabolites and the microbial genera in cancer and healthy tissues revealed several significant associations. The elevated abundance of genus *Fusobacterium* was significantly correlated with the decreased 4-HB level (r = 0.722, p-value = 0.002, degree of freedom(df) = 13; Table 1; details in Additional file 3: Figure S2). On the other hand, decrease in abundance of genus *Prevotella 2* was correlated with the declined level of 4-HB in tumour tissues in comparison with matched normal tissues (r = – 0.593, p-value = 0.02, df = 13; Table 1). In addition, the increase of family
Halomonadaceae abundance exhibited a significant correlation with the increase in glutamic acid level between the tissues \((r = 0.541, p\text{-value} = 0.037, \text{df} = 13; \text{Table 1})\). The increase of lactic acid concentration in cancer tissues was also found to be significantly associated with the decrease in both genus Paeniclostridium level \((r = -0.537, p\text{-value} = 0.039, \text{df} = 13; \text{Table 1})\) and genus Enterococcus level \((r = -0.527, p\text{-value} = 0.044, \text{df} = 13; \text{Table 1})\).

Colonic metabolite-correlated microbial taxa associated with DNA methylation-related differential gene expression

In total, four tumour and normal tissue pairs were selected for transcriptome analysis using RNA-seq. After pre-processing and filtering, 15,832 gene transcripts were included in downstream analysis. Multi-dimensional scaling (MDS) was performed to visualise the differences in gene expression levels between cancer tissues and matched healthy tissues (Fig. 4a). After generalised linear model (GLM)-based paired differential comparison, 261 transcripts were found to be downregulated in cancer tissues while 333 were found to be

| Table 1 Significant correlation of differentially abundant microbial taxa and metabolites \((n = 15)\) |
|--------------------|-----------------|--------------|-------------|
| Microbial taxa     | Metabolites      | \(r\) value  | \(p\)-value |
| Genus Fusobacterium (increased in tumour) | 4-HB (decreased in tumour) | 0.722 | 0.002 |
| Genus Prevotella 2 (decreased in tumour) | 4-HB (decreased in tumour) | -0.593 | 0.020 |
| Family Halomonadaceae (increased in tumour) | glutamic acid (increased in tumour) | 0.541 | 0.037 |
| Genus Paeniclostridium (decreased in tumour) | lactic acid (increased in tumour) | -0.537 | 0.039 |
| Genus Enterococcus (decreased in tumour) | lactic acid (increased in tumour) | -0.527 | 0.044 |

\(a\)For the calculation of the correlations, the signed relative change of microbe abundance and metabolite level between paired tumour and normal tissues were used
significantly up-regulated in tumour tissues (fold change (FC) > 2, *p*-value < 0.05; Fig. 4b).

We then evaluated the methylation profile of four tumour and normal tissues pairs at the 594 differentially expressed genes, including all the probes that cover the promoters and gene body of these genes. After quality control, 14,108 probes were selected for subsequent analysis. As shown in Fig. 4c, the methylation patterns of these differentially expressed genes are clearly distinctive between the tumour and normal tissues. Of the 111 differentially methylated probes, 68 were highly methylated in cancer tissues (*p*-value < 0.05; Additional file 4: Table. S2). Due to the reported relationship of DNA hypermethylation located in gene promoter and gene transcription repression, genes with the opposite trend of DNA methylation at the promoter region and the
The reduction in the expression level of Halomonadaceae (decreased in tumour) was significantly associated with the decreased expression of MT1M (Fig. 4d).

We further validated our hypothesis of a possible microbiome and host transcriptome interaction by analysing the correlation between the two profiles. The microbial taxa that were associated with the colonic metabolome were included. As a result, the increase in genus Fusobacterium abundance was found to be significantly associated with the down-regulated expression of PI16 ($r = -0.999$, $p$-value = 0.008, df = 2; Table 2) and also significantly associated with the decreased expression of FCRLA ($r = 0.956$, $p$-value = 0.044, df = 2; Table 2) and LSP1 ($r = 0.954$, $p$-value = 0.046, df = 2; Table 2) in cancer tissues in comparison with the healthy tissues. The decrease in the abundance of genus Prevotella 2 was significantly associated with the decreased expression of MT1M in cancer tissues ($r = -0.999$, $p$-value = 0.001, df = 2; Table 2); the positive abundance change of family Halomonadaceae was significantly associated with the reduction in the expression level of RELN ($r = 0.957$, $p$-value = 0.043, df = 2; Table 2). Nevertheless, the reduced abundance of genus Paeniclostridium and genus Enterococcus in cancer tissues also exhibited significant associations with the expression difference of PLCB1 ($r = 0.995$, $p$-value = 0.01, df = 2; Table 2) and IGSF9, respectively ($r = -0.997$, $p$-value = 0.003, df = 2; Table 2).

**Discussion**

In our present study, we comprehensively investigated the host-microbiome interaction in CRC by assessing multi-omics, including the mucosal microbiome, mucosal metabolome, host transcriptome and host DNA methylation profile.

We replicated several differentially abundant microorganisms as reported in previous CRC-related studies. At the genus level, potential pathogenic microorganisms such as genus *Fusobacterium* were found to be significantly enriched in tumour tissues compared with normal tissues in our samples. The genus *Fusobacterium* is a well-known potential pathogenic gut microorganism and enrichment of this genus has been reported to be associated with CRC in several studies [21–23]. The over-representation of genus *Halomonas* and genus *Shewanella* has also been revealed in tumour-associated microbiota in patients with rectal and distal colon cancers [24]. In addition, genus *Peptostreptococcus* has been shown to be associated with CRC in several studies [8, 25, 26]. Similarly, genus *Klebsiella* has been implicated in the progression of CRC [6, 27]. On the other hand, genus *Lactobacillus* and genus *Enterococcus*, were over-represented in healthy tissues in the current study, which have been long regarded as anti-inflammation and anti-tumorigenic probiotics [28, 29]. A recent study also reported that genus *Lactobacillus* was able to counterbalance the dysbiosis induced by CRC in human and mice [26, 30]. In an animal study, genus *Enterococcus* was shown to inhibit chemically-induced CRC [31].

The microbiota and its metabolic products have been reported to affect the health status of the host, especially in colon tissues [32]. In line with other metabolome studies in CRC, several metabolites were found to be significantly differentially abundant in cancer mucosal tissues in contrast to adjacent healthy tissues. Interestingly, the top three most distinguished differences between colon tissues in our study were lactic acid, 4-hydroxybutyric acid and glutamic acid. The elevation of

| Microbial taxa                     | Genes       | Coding protein          | $r$ value $^a$ | $p$-value |
|------------------------------------|-------------|-------------------------|----------------|-----------|
| Genus *Fusobacterium* (increased in tumour) | PI16 (decreased in tumour) | Peptidase inhibitor 16 | $-0.993$ | 0.008     |
|                                    | FCRLA (decreased in tumour) | Fc Receptor Like A | 0.956 | 0.044     |
|                                    | LSP1 (decreased in tumour) | Lymphocyte Specific Protein 1 | 0.954 | 0.046     |
| Genus *Prevotella 2* (increased in tumour) | MT1M (decreased in tumour) | Metallothionein | $-0.999$ | 0.001     |
| Family *Halomonadaceae* (increased in tumour) | RELN (decreased in tumour) | Reelin | 0.957 | 0.043     |
| Genus *Paeniclostridium* (decreased in tumour) | PLCB1 (increased in tumour) | Phospholipase C beta 1 | 0.995 | 0.010     |
| Genus *Enterococcus* (decreased in tumour) | IGZF9 (decreased in tumour) | Immunoglobulin Superfamily Member 9 | $-0.997$ | 0.003     |

$^a$For the calculation of the correlations, the signed relative change of microbe abundance and gene expression level between paired tumour and normal tissues were used.
lactic acid and glutamic acid in cancer tissues has been shown in multiple cohort studies of the CRC mucosa metabolome [33]. Additionally, taurine was found to be high in tumour tissues. Taurine is a sulphur amino acid, and studies have concluded that it is pro-inflammatory and genotoxic [34]. Oxalic acid has also been reported to be decreased in CRC samples [35]. Though tetracosane has been reported to be cytotoxic and potentially apoptotic, a high level of tetracosane in CRC tumour tissues has not yet been reported and warrants further study [36]. D-talose, which was found to be more abundant in normal tissues, is a rare sugar with possible anti-tumour application in the pharmaceutical industry [37, 38].

The key findings in our study were the associations between changes in microbial taxa, butyrate-related colonic metabolites, and DNA methylation-associated host gene expression between tumour and paired normal colon tissues. We identified a significant association between the increase of the pro-oncogenic genus *Fusobacterium* and decrease of 4-HB between the tissues. As a hydroxylated derivative of butyrate, 4-HB is involved in one of the four pathways of butyrate synthesis (the 4-aminobutyrate pathway) [39]. Although genus *Fusobacterium*, especially *F. nucleatum*, has been linked to colorectal tumorigenesis, the mechanism remains unclear [21–23]. Our finding suggests that, under tumour conditions, genus *Fusobacterium* might promote inflammatory response and carcinogenesis by interfering with butyrate synthesis in the tumour tissues. Butyrate is a short-chain fatty acid and has been revealed in several previous studies to be a potential anti-carcinogenic and anti-inflammatory metabolite [40, 41].

Butyrate is produced by the microbial fermentation of undigested polysaccharides that reach the colon [42]. In addition, butyrate also plays a crucial role in influencing gene expression in host colonic cells via epigenetic regulation [43, 44]. In human cells, there is evidence that butyrate has a direct effect on DNA methylation by regulating key enzymes, including methylcytosine dioxygenase (TET) and DNA methyltransferase 1 (DNMT1) [45, 46]. Butyrate has also been shown to modify DNA methylation and inhibit Wnt signalling in human gastric cancer cells, a pathway also known to be activated in CRC [47]. The potential association between *Fusobacterium* and DNA methylation-related carcinogenesis has been supported by a previous study [18]. It has also been demonstrated that *Fusobacterium* is enriched in CRC tissues, especially those with the methylation phenotype [22]. Butyrate may mediate the effect of *Fusobacterium* on host DNA methylation modifications.

In our study, the change in abundance of genus *Fusobacterium* was also correlated with DNA methylation-related expression differences in genes PI16, FCRLA and LSP1 in colon tissues. PI16 is a peptidase inhibitor; consistent with our findings, other studies have detected decreased expression of PI16 in colon cancer [48]. FCRLA is an unusual member of the lymphocyte receptor family and is associated with multiple Ig isotypes, including IgM, IgG and IgA [49, 50]. As a B cell-specific protein, FCRLA may play a crucial role in the mucosal immune response [51]. Down-regulation of LSP1 has been shown in several cancers and its over-expression is associated with cell apoptosis [52, 53].

Additionally, butyrate might also play a role in the effect of genus *Prevotella* 2 on CRC. The decrease of genus *Prevotella* 2 abundance and 4-HB concentration in tumour tissues showed a significant correlation. In terms of the correlation between genus *Prevotella* 2 and 4-HB, more detailed genomic analysis of its strains is warranted to identify the metabolic pathway. The change of genus *Prevotella* 2 abundance was also associated with the reduced expression of MT1M in cancer tissues compared with healthy tissues. A member of the metallothionein (MT) family, MT1M is a cysteine-rich metal-binding protein that is induced by inflammation and protects the cell against carcinogens [54].

Another observation in our study is the positive relationship between the increase of glutamic acid and microbial taxa in the potential pathogenic family *Halomonadaceae* in cancer tissues compared with normal tissues. In the large intestine, glutamic acid is partially originated from the digestion of alimentary and endogenous proteins by the colonic microbes [55]. Glutamic acid can also serve as butyrate precursor and at least five various pathways have been reported to be responsible for the fermentation of glutamic to butyrate by anaerobic bacteria [56]. Nonetheless, glutamic acid can also be metabolized to α-ketoglutarate (α-KG), which is co-factor of methylcytosine dioxygenase (TET) [57]. The link between *Halomonadaceae* and glutamic acid identified in our finding suggests that taxa in this family might involve in glutamic acid metabolism. Furthermore, the elevation of the taxa in family *Halomonadaceae* also correlated with the decreased expression of Reelin (*RELN*) in tumour tissues. The downregulation of Reelin has been found to be related to cell migration ability, tumour invasiveness and patients’ survival rates in several cancers, including gastric, breast, pancreatic and liver cancers [58–60]. Importantly, large CpG islands are located at Reelin promoter sites and its transcriptional silence has been shown to be strongly controlled by promoter hypermethylation [61]. Our finding might offer additional evidence of the mechanism of epigenetic regulation of RELN in colon cancer.

We also reported a significant association between the increase in lactic acid level and the decrease in abundance of genus *Paeniclostridium* as well as genus
Enterococcus in cancer tissues, in contrast with normal tissues. This may in part be due to the excess lactic acid produced by cancer cells as a result of the high consumption of glucose in glycolysis, which is characteristic of tumour metabolism; this is termed the Warburg Effect [62]. Nonetheless, due to the synergic metabolism of lactate-producers such as genus Enterococcus and lactate-utilizing bacteria in the colon, less lactate might be disposed by lactate-utilizing bacteria in tumour tissues [63]. Lactate fermented by lactate-utilizing bacteria can be used to generate butyrate, which, as mentioned, might influence gene expression by modifying DNA methylation [64]. To explain the correlation between genus Paeniclostridium and lactic acid, further functional studies are required to explore their relationship. Interestingly, the reduced level of genus Enterococcus in tumour tissues was also found to be associated with the down-regulated expression of IGSF9 (Immunoglobulin Superfamily Member 9), which contributes to regulating immune cell activation [65]. Moreover, the reduced abundance of genus Paeniclostridium was related to the increased expression of Phospholipase C beta 1 (PLCBI). Phospholipase C (PLC) is critical for cell proliferation and apoptosis in several cell systems [66, 67]. The knockdown of PLCBI is reported to initiate apoptosis, leading to strong growth arrest [68].

One limitation of our study is the use of 16sRNA sequencing technology; as such, we are unsure which specific species might distinguish tumour and normal tissues. However, the current shotgun metagenomic method is infeasible for mucosal microbiome study because of human DNA contamination [69]. Future development of sequencing technology would be critical for the identification of CRC-related microbes. Due to a relatively small sample size, the heterogeneity of CRC was also not investigated. However, our main aim was to conduct a pilot study to investigate the potential interaction between the host and microbiome from multiple “omics”. Future studies with larger sample sizes would help disentangle the complicated host-microbiome relationship in different stages and different types of CRC. Another issue is the causality of the reported relationship. Further animal intervention studies are needed to validate the correlation identified from our findings.

Apart from integrating multiple “omics”, another strength of our study is that we measured the microbiome from the mucosa sampled from CRC patients. The mucosal microbiome is different from the faecal microbiome, which is only an estimate of the former. The mucosal microbiome exhibits a more intimate and direct interaction with the host colonic epithelium; thus, it has a greater impact on the pathogenesis of disease [70].

Conclusions

In conclusion, our study identified associations between previously reported CRC-related microbial taxa, butyrate-related metabolites and DNA methylation-associated gene expression in tumour and paired normal colonic mucosa tissues from CRC patients, which uncovered a possible mechanism of the role of microbiome in the carcinogenesis of CRC. In addition, these findings offer insights into potential new biomarkers, therapeutic and/or prevention strategies for CRC.

Methods

Patient recruitment and DNA/RNA isolation

A total of 36 pairs of tumour tissues and adjacent healthy colon tissues were obtained from CRC patients in the Department of Colorectal and Anal Surgery, First Affiliated Hospital of Zhejiang University, China. Mucosa tissues were collected during surgery from both the tumour site and from an adjacent non-tumorous site that was more than 5 cm away from the tumour. The samples were flash-frozen in liquid nitrogen immediately after colonic resection and were stored at −80 °C. Patients with diabetes or infectious diseases, those undergoing radiotherapy or chemotherapy, and those who had received antibiotics or probiotics within 4 weeks were excluded from the study. The study was approved by the Ethics Committee of the First Affiliated Hospital of the Medical School of Zhejiang University. Written informed consent was obtained from all subjects.

Total DNA and RNA were extracted from the tissue samples using AllPrep DNA/RNA mini kit (Qiagen, Valencia, CA, USA) according to the manufacturer’s instructions.

Mucosal microbiome analysis

After assessing DNA concentration and quality, PCR amplification of the V3-V4 regions of the 16S rRNA gene was performed. Then, the qualified amplicon libraries were sequenced on the Ion S5TM XL platform (Thermo Scientific) following the manufacturer’s protocol. The sequencing reads were processed using QIIME pipeline (v1.9.1) [71]. Quality filtering were performed under specific conditions according to the Cutadapt (Martin, 2011) [72]. Chimera sequences were identified using the UCHIME algorithm and were removed. After quality control and trimming, the reads were merged and clustered into operational taxonomic units (OTUs) using UPARSE software (v7.0.1001) with a 97% similarity threshold [73]. Taxonomy was assigned to each OTU cluster based on the representative reads using the RDP classifier based on the SILVA Database version 123 (confidence threshold: 80%) [74]. OTUs present in less than 25% of the total samples and with less than 3 reads were excluded. Non-metric multidimensional scaling
The dried residue was added to 80 under a nitrogen stream (Aosheng, Hangzhou, China). From the previous step. Next, the samples were dried supernatant was then combined with the supernatant ous and non-cancerous tissues with a generalized linear model (GLM) [77]. Multi-dimensional scaling (MDS) plot was generated to visualize the difference between pairs of RNA samples. The resulting p-values were corrected using the Benjamini and Hochberg approach at False Discovery Rate (FDR) of 5% [78]. Genes with an adjusted p-value< 0.05 and fold change (FC) > 2 were considered differentially expressed between the groups.

Host transcriptome analysis
The RNA sequencing library was prepared as previously described [75] and sequenced on an Illumina HiSeq 2000 platform. After quality control, the paired-end clean reads were mapped to the reference genome hg19 using Hisat2 (v2.0.5). FeatureCounts (v1.5.0-p3) was used to count the number of reads mapped to genes [76]. Then, the gene expression value was quantified by the length of the gene and total read counts of the gene as count per million (CPM). The transcripts were subsequently filtered with the cut-off of CPM > 0.5 in at least 4 libraries. The bioconductor package edgeR was used to normalize and analyse differential gene expression levels between paired cancerous and non-cancerous tissues with a generalized linear model (GLM) [77]. Multi-dimensional scaling (MDS) plot was generated to visualize the difference between pairs of RNA samples.

Host methylation analysis
An EZ DNA Methylation Kit was used for bisulfite modification of the genomic DNA. Then, DNA methylation analysis was performed using the Infinium HumanMethylation850 BeadChip array (Illumina San Diego, CA, USA), which covers more than 850,000 CpG sites and 99% of the RefSeq genes. DNA methylation data analysis was performed following a cross-package Bioconductor workflow [79]. Methylation levels were quantified as betal value, which describes the proportion of methylation at each CpG locus. After background subraction and data normalization, probes that were located at the region of the differentially expressed genes were selected for subsequent analysis. Differential DNA methylation between the groups was assessed by paired linear model using the limma R package and the p-values were adjusted for multiple testing using Benjamini-Hochberg method at FDR of 5% [78]. The probes with an adjusted p-value less than 0.05 were considered statistically significantly differentially methylated.

Interactions between omics
The correlations between change in the level of differentially abundant microbes and metabolites were estimated with Pearson correlation tests using the function cor.test in R (version 3.5.2). Due to the interdependence between

Mucosal metabolomics analysis
50 mg of colon mucosa tissues was mixed with 800 μl ice-cold methanol and homogenized. After centrifugation, the supernatant was transferred to a new Eppendorf tube containing 20 μl of internal standard (1 mg/ml heptadecanoic). The remainder was extracted with 800 μl ice-cold methanol/chloroform (3:1, v/v) and the supernatant was then combined with the supernatant from the previous step. Next, the samples were dried under a nitrogen stream (Aosheng, Hangzhou, China). The dried residue was added to 80 μl of 20 mg/ml methoxylamine hydrochloride in anhydrous pyridine. The mixture was incubated at 37 °C for 24 h after vortexing. Then, the sample was mixed with 100 μl of N,N-bis(trimethylsilyl)fluoroacetamide (BSTFA) [with 1% trimethylsilyl chloride (TMCS)] (Sigma-Aldrich, St. Louis, MO, USA) and derivatised (70 °C for 2 h). Untargeted metabolomics analysis was carried out by gas chromatography-mass spectrometry (GC-MS) with an Agilent 7890A GC and 5975C inert mass selective detector (MSD) system (Agilent Technologies, Santa Clara, CA, USA). The raw data from GC-MS was transformed by ChemStation software (version E02.02.1431, Agilent, CA, USA) and pre-treated by ChromaTOF software (version 4.34, LECO, St. Joseph, MI, USA). The Fiehn database was referred to for identification of the metabolites. The GC-MS dataset was quantile normalized, auto-scaled and log transformed using MetaboAnalyst 4.0 [www.metaboanalyst.ca/]. Subsequently, principal component analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA) was performed using SIMCA (Umetrics, Umeå Sweden) to measure the differences in metabolites between the tissue samples. Model quality was summarized by R2Y and Q2 parameters, which measures the goodness of fit and the predictive ability of the model respectively (26). To measure the reliability of the OPLS-DA model from overfitting, we performed sevenfold cross-validation and 200 response permutation tests which randomly permuted the group attributes (Y) whereas the variables (X) remained same. Metabolites were considered statistically significant if the Variable Influence on Projection (VIP) value, a measure of the relative influence on the model, was greater than 1 and the p-value returned from the two-tailed Student’s t test on normalized data was less than 0.05.
paired tumour and normal tissues, the signed relative change of microbe abundance and metabolite concentration between paired tissues were used for the calculation of the correlations. Correlations were considered statistically significant if the \( p \)-value was less than 0.05. The same correlation tests were also conducted between the levels of differentially abundant microbes and differentially expressed genes that also exhibited opposite trend of differential methylation at the promoter region. Metabolites and methylation-related differentially expressed genes that were associated with the same microbe were reported.

**Supplementary information**

Additional file 1 Figure S1. Flow chart of study design.

Additional file 2 Table S1. Significantly differentially abundant microbial genera between tumour and normal colon tissues from CRC patients (\( n = 36 \)).

Additional file 3 Figure S2. Illustration of correlation between microbial abundance and metabolite concentration.

Additional file 4 Table S2. Significantly differentially methylated probes that located within the differentially expressed genes between tumour and normal colon tissues from CRC patients (\( n = 4 \)).

**Abbreviations**

CRC: Colorectal cancer; NMDS: Non-metric multidimensional scaling; OTU: Operational taxonomic unit; Anosim: Analysis of similarities; FDR: False discovery rate; ZIG: Zero-inflated Gaussian; GC-MS: Gas chromatography-mass spectrometry; PCA: Principal component analysis; OPLS-DA: Orthogonal partial least squares discriminant analysis; VIP: Variable Influence on Projection; MDS: Multi-dimensional scaling; GLM: Generalized linear model; FC: Fold change; 4-HB: 4-hydroxybutyric acid; DNMT1: DNA methyltransferase 1; TET: Methylcytosine dioxygenase

**Acknowledgements**

We acknowledge the support of the Victor Chang Cardiac Research Institute Innovation Centre, funded by the NSW Government.

**About this supplement**

This article has been published as part of BMC Microbiology Volume 20 Supplement 1, 2020: Proceedings of the Joint International GIW & ABACBS-2019 Conference: microbiology. The full contents of the supplement are available online at https://bmcmicrobiol.biomedcentral.com/articles/supplements/volume-20-supplement-1

**Authors’ contributions**

QW, JZ, Y, DQF and LXL designed and coordinated the study. JWKH, EG and LL supervised the project. QW and WRW collected the samples. QW, DS, YT, LYY, XYB and JJW carried out the experiments. XWJ, KCW and QHW helped collect clinical data. MPH contributed to metabolome data analysis. QW, LT, JWKH and EG participated in the data analysis and result interpretation. QW wrote the manuscript. LT, MPH, JWKH, EG and LLL reviewed the manuscript. All authors have contributed and approved the final manuscript.

**Funding**

Funding for study design, data collection and data generation was provided by the National Natural Science Foundation of China (grant nos. 81709063, 81570512), the Science Fund for Creative Research Groups of the National Natural Science Foundation of China (grant no. 81721091) and the National Key Research and Development Program of China (2018YFC2000500). JWKH participated in the data analysis and result interpretation and was supported by Career Development Fellowship by the National Health and Medical Research Council (1105271) and Future Leader Fellowship by the National Heart Foundation of Australia (100848). EG participated in the data analysis and result interpretation and is supported by the NSW Health Early-Mid Career Fellowship and the National Heart Foundation of Australia Future Leader Fellowship (101204). Publication costs are funded by the National Natural Science Foundation of China (grant no. 81709063).

**Availability of data and materials**

The datasets used and/or analysed during the current study are available from Sequence Read Archive (SRA) database (accession number: PRJNA599023). All data generated or analysed during this study are also available from the corresponding author on reasonable request.

**Ethics approval and consent to participate**

All participants in this study provided written informed consent before sample collection. The research is in accordance with the ethical guidelines of the 1975 Declaration of Helsinki and was approved by the Ethics Committee of the First Affiliated Hospital of Zhejiang University (permit number: 2017–059–1).

**Consent for publication**

Not Applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

1 State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital of Zhejiang University, Hangzhou, China. 2 Collaborative Innovation Centre for Diagnosis and Treatment of Infectious Diseases, Hangzhou, China. 3 Computational Genomics Laboratory, Victor Chang Cardiac Research Institute, Sydney, Australia. 4 Freedman Foundation Metabolomics Facility, Victor Chang Innovation Centre, Victor Chang Cardiac Research Institute, Sydney, Australia. 5 School of Pharmacy, University of Queensland, Woolloongabba QLD 4102, Australia. 6 School of Mathematics and Statistics, UNSW Sydney, Sydney, Australia. 7 Bioinformatics and Systems Medicine Laboratory, Victor Chang Cardiac Research Institute, Sydney, Australia. 8 School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. 9 St Vincent’s Clinical School, UNSW Sydney, Sydney, Australia.

**Received:** 24 February 2020 Accepted: 23 March 2020

**Published:** 23 April 2020

**References**

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424. https://doi.org/10.3322/caac.21492.
2. Raskin H, Pommergaad H-C, Burchart J, Rosenberg J. Colorectal carcinogenesis-update and perspectives. World J Gastroenterol. 2014;20:188–191. https://doi.org/10.3748/wjg.v20.i4.18151.
3. Bingham SA. Diet and colorectal cancer prevention. Biochem Soc Trans. 2000;28:12–6. https://doi.org/10.1042/bst0280012.
4. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207–14. https://doi.org/10.1038/nature11234.
5. De Filippo C, Cavaleri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. Proc Natl Acad Sci U S A. 2010;107:14691–6. https://doi.org/10.1073/pnas.1005963107.
6. Wang T, Cai G, Qiu Y, Fei N, Zhang M, Pang X, et al. Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers. ISME J. 2012;6:320–9. https://doi.org/10.1038/ismej.2011.109.
7. Sobhani I, Tapi J, Roudot-Thoraval F, Roppec JP, Letulle S, Langella P, et al. Microbial dysbiosis in colorectal cancer (CRC) patients. PLoS One. 2011;6:e16393. https://doi.org/10.1371/journal.pone.0016393.
8. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal Cancer. PLoS One. 2012;7:e39743. https://doi.org/10.1371/journal.pone.0039743.
