

BRIEF REPORT



Establishing a novel colorectal cancer predictive model based on unique gut microbial single nucleotide variant markers

Chenchen Ma^{a,b}, Kaining Chen^c, Yuanyuan Wang^{a,b}, Chaoping Cen^c, Qixiao Zhai^{d,e}, and Jiachao Zhang^{a,b}

^aCollege of Food Science and Engineering, Hainan University, Haikou, Hainan, P. R. China; ^bKey Laboratory of Food Nutrition and Functional Food of Hainan Province, Hainan University, Haikou, Hainan, China; ^cHainan General Hospital, Hainan Affiliated Hospital of Hainan Medical University, Haikou, P. R. China; ^dState Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, P. R. China; ^eSchool of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu, China

ABSTRACT

Current metagenomic species-based colorectal cancer (CRC) microbial biomarkers may confuse diagnosis because the genetic content of different microbial strains, even those belonging to the same species, may differ from 5% to 30%. Here, a total of 7549 non-redundant single nucleotide variants (SNVs) were annotated in 25 species from 3 CRC cohorts (n = 249). Then, 22 microbial SNV markers that contributed to distinguishing subjects with CRC from healthy subjects were identified by the random forest algorithm to construct a novel CRC predictive model. Excitingly, the predictive model showed high accuracy both in the training (AUC = 75.35%) and validation cohorts (AUC = 73.08%-88.02%). We further explored the specificity of these SNV markers in a broader background by performing a meta-analysis across 4 metabolic disease cohorts. Among these SNV markers, 3 SNVs that were enriched in CRC patients and located in the genomes of *Eubacterium rectale* and *Faecalibacterium prausnitzii* were CRC specific (AUC = 72.51%-94.07%).

ARTICLE HISTORY

Received 23 October 2020
Revised 10 December 2020
Accepted 16 December 2020






KEYWORDS

Metagenome; colorectal cancer; single nucleotide variants; gut microbiota; diagnostic markers

Introduction

Colorectal cancer (CRC) is the most common cancer, both in men (1-lung, 2-prostate, 3-colorectal, 4-pancreatic cancer) and women (1-lung, 2-breast, 3-colorectal, 4-pancreatic cancer)¹ and the second most common cause of cancer death after lung cancer.² In the last decade, some studies have highlighted the importance of the gut microbiome in CRC. In particular, many bacteria,³ including *Fusobacterium nucleatum*, *Bacteroides fragilis* and *Escherichia coli* and fungi,¹ including *Malasseziomycetes* and *Candida* are involved in the development of CRC due to pathogenicity and carcinogenicity via multiple mechanisms. Accordingly, numerous noninvasive microbial biomarkers based on metagenomic species or functional genes were developed for early-stage CRC diagnosis.⁴⁻⁷ However, the genetic content of different microbial strains, even those belonging to the same species, may differ from 5% to 30% or more,⁸ which may confuse the diagnosis and create a barrier to applying species-level biomarkers.

Therefore, analysis of gut microbial single nucleotide variants (SNVs) can provide an in-depth view of CRC pathogenesis. To date, only one study has reported that intestinal *Bacteroides coprocola* has a characteristic distribution of single nucleotide variants in the T2D patient group compared to healthy controls.⁹ Unfortunately, no study has addressed the association between CRC and gut microbiota at the SNV level, and the profile of gut microbial genomic variation in patients suffering from CRC is largely unknown. Here, for the first time, this challenge was addressed by identifying gut microbial SNVs in CRC patients from three discovery cohorts (n = 249). Additionally, we recruited a validation cohort and resampled discovery cohorts to evaluate the accuracy of the CRC predictive model. Finally, four disease cohorts were used to determine the specificity of the CRC SNV markers. Importantly, a new method was established to predict CRC with high accuracy based on SNV signatures of gut microbiota.

CONTACT Jiachao Zhang  zhjch321123@163.com  College of Food Science and Engineering, Hainan University, Haikou, Hainan 570228, P. R. China; Qixiao Zhai  zhqixiao@sina.com  State Key Laboratory of Food Science and Technology, Jiangnan University, Wuxi, Jiangsu 214122, P. R. China.
 Supplemental data for this article can be accessed on the publisher's website.

© 2021 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Results

SNV annotation, construction and verification of CRC classification model

The depth of the metagenomic sequencing and the coverage of each strain directly affected the discrimination of intestinal microbial SNV identification. Therefore, by using MetaPhlan2 for species annotation, the species with the average relative abundance greater than 0.5% were selected for SNV annotation. The selected reference or representative strains from NCBI and their GenBank accessions are listed in supplemental material 3. In general, 7549 non-redundant SNVs were annotated in 25 species from 249 individuals of 3 cohorts, including Japanese, Australian and Italian individuals. Then, the random forest algorithm was used to find the SNV markers that contributed to distinguishing CRC from healthy controls from three discovery cohorts (mean decrease in accuracy > 0). The Wilcoxon rank-sum test was performed to determine the significantly different SNV markers shared among the three cohorts. Finally, a total of 22 SNV markers (including 4 SNVs enriched and 18 SNVs absent in CRC patients) were used to build the prediction model (supplemental materials 5A). The 22 SNVs belong to 4 species, among which 11 belong to *Eubacterium rectale*, nine belong to *Faecalibacterium prausnitzii*, and both *Bifidobacterium pseudocatenulatum* and *Bacteroides vulgatus* contain 1 SNV (supplemental materials 5A). Based on the presence profile of the 22 SNVs (supplemental material 2A), a novel CRC predictive model was constructed. The results showed that the predictive model had a high accuracy of 75.35% in the discovery cohorts (Figure 1a). To further evaluate the robustness of the model, 8 CRC patients and 12 healthy controls were recruited from Hainan Province, China. Our model achieved an area under the receiver-operating curve (AUC) of 88.02% for the validation samples (supplemental material 2B) (Figure 1b). In addition, we resampled to verify the robustness of the model again. We obtained AUCs of 79.53%, 76.25%, and 73.08%, respectively, for the three resampling processes (Figure 1c).

Functional annotations related to four enriched SNVs and cross-disease model verification

The 4 SNVs enriched in CRC patients were located in the genomes of *Eubacterium rectale* and *Faecalibacterium prausnitzii*. Interestingly, we did not observe any significant difference in the relative abundance of *Eubacterium rectale* and *Faecalibacterium prausnitzii* in all cohorts (supplemental materials 5B, C), which further underlined the sensitivity of gut microbial SNV biomarkers. To gain insights into the 4 enriched SNV functions in CRC, their respective functions were demonstrated according to the feature table from the public database. Further, we observed that all four SNVs were non-synonymous mutations and we predicted the proteins structure of the related genes using Phyre2¹⁰ (Figure 1d,e). The 2 SNVs located in the *Eubacterium rectale* genome were related to the function of fusaric acid resistance protein-like (FUSC family protein, WP_012744219.1). The other two CRC-enriched SNVs in *Faecalibacterium prausnitzii* were assigned to the function of methyltransferase and ZF-HC2 domain-containing protein. To further explore the specificity of the 4 CRC-enriched SNVs in a broader background, a meta-analysis was performed across four metabolic disease cohorts, including ulcerative colitis (UC), Crohn's disease (CD), liver cirrhosis (LC), and type 2 diabetes (T2D). The results showed that 3 of 4 SNVs were disease specific. In contrast, the SNV1 of *Faecalibacterium prausnitzii* showed no difference between CRC and intestinal IBD diseases, including CD and UC (Figure 1f,g). Subsequently, the three SNVs were used to distinguish CRC from other conditions, and the accuracy ranged from 72.51% to 94.01%, which implied the outstanding specificity of the 3 SNV markers in CRC disease (supplemental material 2 C) (Figure 1h).

Discussion

According to the above findings at the intestinal microbial SNV level, there are many potential applications worth discussing that are not limited to the development of potential drug targets reported by a previous study.⁹ Notably, fusaric acid (FA) is a ubiquitous but neglected fungal

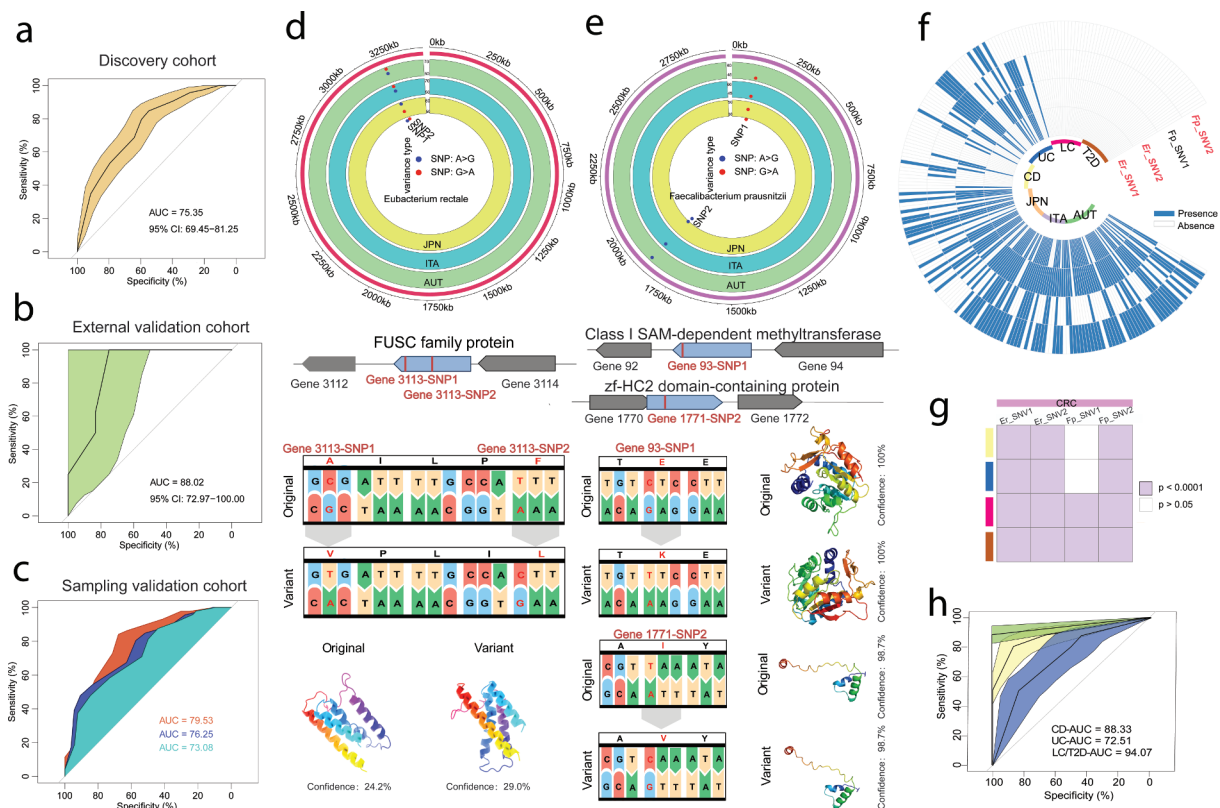


Figure 1. Construction and validation of CRC prediction model based on SNVs. (a) Prediction power in discovery cohorts with the accuracy of 75.35% using 22 SNV markers, including 4 SNVs enriched and 18 SNVs absent in CRC patients. (b) CRC classification accuracy for independent validation set recruited with 8 CRC patients and 12 healthy controls from Hainan Province, China. (c) The performance of the model in resampling validation cohorts. The sampling validation cohorts refer to the random sampling of 100 samples from discovery cohorts. (d) And (e) The related position, functions, amino acid and protein structure and of the four SNVs enriched in CRC. (f) And (g) Heatmaps were used to show the disease specificity of the four SNVs mentioned above, and three SNVs were enriched in CRC cohorts, including Er_SNV1, Er_SNV2 and Fp_SNV2. (h) The accuracy of enriched SNVs to classify various diseases, and the accuracy ranged from 72.51% to 94.01%.

toxin.¹¹ So, it should be emphasized that the human gut microbiota consists of not only bacteria but also viruses, fungi, and Archaea. Possible changes in FA resistance should be taken seriously in patients with CRC due to two significantly enriched SNVs related to FUSC family proteins. Interestingly, FA is a causative agent of esophageal cancer¹² and decreases p53 expression;¹³ on the other hand, FA has activity against head and neck squamous cell carcinoma^{14,15} and human esophageal epithelial carcinoma cells.¹⁶ Therefore, the SNV profiles indicate that the effect of FA on human colorectal cancer cells should be further investigated. In summary, our study suggests that SNV distributions should be further examined to determine, in-depth, potential changes in the function of the gut microbiota. Even so, verification with larger and more regional cohorts still needs to be carried out, and the isolation and confirmation of strains is

especially critical. The Study have summarized gut microbes associated with CRC development,³ including *Fusobacterium nucleatum*, *Bacteroides fragilis*, *Escherichia coli*, *Enterococcus faecalis*, *Helicobacter hepaticus*, *Peptostreptococcus anaerobius*, *Helicobacter pylori*, *Streptococcus bovis*, and *Porphyromonas gingivalis*. However, *Eubacterium rectale* and *Faecalibacterium prausnitzii* have not received enough attentions in patients suffering CRC. Only few studies implied *Faecalibacterium prausnitzii* as the potential probiotic because of its maintenance function of gut homeostasis.^{3,17} Interestingly, both *Eubacterium rectale* and *Faecalibacterium prausnitzii* were crucial intestinal microbes for butyrate producing,^{18,19} and butyrate represented SCFAs were considered as microbial metabolites with anti-tumorigenic properties and may contribute to the prevention of CRC.^{20,21} However, no significant difference was observed

in the abundance of *Eubacterium rectale* and *Faecalibacterium prausnitzii* in all cohorts, which implied even the microbial relative abundance has not changed, a number of significant functional mutations have occurred in microbial genome which participate in the development of CRC.²²

The present predictive model constructed by 22 CRC gut microbial SNVs exhibited high accuracy in training and validation cohorts, and the 3 CRC-enriched microbial SNVs were disease specific. Here, we not only explored the potential correlation between gut microbial genomic mutations and CRC disease but also developed a feasible noninvasive CRC diagnostic method. SNVs profiles of bacteria in the gut of CRC patients has been analyzed in this study, whereas it was observed that the CRC stage is strongly related to the fungal microbes.¹ So, fungal SNVs in the gut of patients with CRC should be analyzed in future studies.

Materials and methods

Sequence data collection

Fecal shotgun metagenomic data of human CRC patients and healthy cohorts were collected. For discovery cohorts, raw SRA files and sample information from three studies were downloaded from NCBI using the following accessions: ERP008729 for Austria,²³ DRA006684 for Japan,²⁴ and SRP136711 for Italy.²⁵ A total of 118 cases and 131 healthy controls were included in this meta-analysis (Table 1). For external validation cohorts, we recruited 8 CRC patients and 12 healthy controls from Hainan Province, China. The patient recruitment and sequencing pipeline can be found in supplemental materials 1. The sequence data have been deposited in the NCBI database under PRJNA663646. We also randomly sampled the discovery cohorts using Rstudio, performing the sampling for a total of three times; each time, 100 individuals were used to evaluate the prediction model, and the random sampling process is shown in supplemental materials 2. To investigate whether four enriched SNVs found from the CRC cohorts were disease specific, we also collected information for four other common diseases, and thirty samples were randomly selected for each disease (Table 1), including inflammatory bowel

Table 1. Fecal metagenomic studies included in this meta-analysis.

Cohorts	No. of cases	No. of controls	Accession
Discovery cohorts	46	63	ERP008729
Austria(AUT)			
Italy(ITA)	32	28	SRP136711
Japan(JPN)	40	40	DRA006684
External validation cohorts ^{#1}			
China(CHN)	8	12	PRJNA663646
Sampling validation cohorts ^{#2}			
Sampling cohort 1	44	56	-
Sampling cohort 2	46	54	-
Sampling cohort 3	48	52	-
Disease comparison cohorts			
IBD-UC	30	-	PRJNA400072
IBD-CD	30	-	PRJNA400072
Type 2 diabetes(T2D)	30	-	PRJNA422434
Liver cirrhosis(LC)	30	-	PRJEB6337

^{#1}: The data for external validation cohorts has not been published.

^{#2}: The sampling validation cohorts refer to the random sampling of 100 samples from discovery cohorts.

diseases (IBDs),²⁶ type 2 diabetes²⁷ and liver cirrhosis²⁸.

Identification of microbial taxonomy and SNV calling

Shotgun metagenomic sequencing and quality control information can be found in supplemental materials 1. For metagenomic species annotation, MetaPhlan2 software was applied for taxonomic classification.²⁹ We next employed MIDAS (Metagenomic Intra-Species Diversity Analysis System) to profile the species-level SNV frequency and gene contents in the gut microbiota.³⁰ Briefly, we constructed reference bacteria in a high-abundance genome database. Information on all 25 reference strains can be found in supplemental material 3. Then, the shotgun metagenomic sequencing reads were mapped to the database for SNV calling. More information on the code can be found in supplemental materials 4 and GitHub: <https://github.com/HNUMcc/CRC-SNP>.

Statistics statement

The statistical analyses were conducted using R software. Randomforest test was performed by the “randomForest” package. Further, we selected the differential SNVs shared among three discovered cohorts based on randomforest results using the Wilcoxon rank-sum test

($p < .05$), which were considered to be potential biomarkers. Boxplot was shown by the “ggplot2” package. Receiver operator characteristic (ROC) analysis was used to assess the performance of the microbial biomarkers using the “pROC” package in R. The heatmap was constructed using TBtools software.³¹

Acknowledgments

We sincerely thank all the volunteers for their participation.

Disclosure of potential conflicts of interest

No potential conflicts of interest were disclosed.

Funding

This work was supported by Key R & D programs in Hainan (No. ZDYF2019150 and ZDYF2018111).

ORCID

Jiachao Zhang  <http://orcid.org/0000-0001-8099-6749>

Ethics approval and consent to participate

The participants from Hainan, China in the study were informed about the study and were provided with written consent. The protocol for the study was approved by the Ethics Committee Hainan University.

Availability of data and materials

The external validation cohort sequence data reported in this paper have been deposited in the NCBI database (metagenomic sequencing data: PRJNA663646). The analysis code for metaphlan2 and SNV calling can be found in supplemental materials 4 and had been deposited in Github: <https://github.com/HNUmcc/CRC-SNP>.

Authors' contributions

The study was designed by J.Z. and Q.Z. The experiment was performed by C.M., K.C., Y.W. and C.C. Data collection was performed by C.M., K.C., Y.W. and C.C. Data analysis was performed by C.M., J.Z. and Q.Z. The manuscript was written by C.M., J.Z. and Q.Z. All authors read and approved the final manuscript.

References

1. Kazmierczak-Siedlecka K, Dvorak A, Folwarski M, Daca A, Przewlocka K, Makarewicz W. Fungal gut microbiota dysbiosis and its role in colorectal, oral, and pancreatic carcinogenesis. *Cancers (Basel)*. 2020;12:12. doi:10.3390/cancers12051326.
2. WHO:- <https://www.who.int/news-room/fact-sheets/detail/cancer> .[2020.10.02]
3. Kazmierczak-Siedlecka K, Daca A, Fic M, van de Wetering T, Folwarski M, Makarewicz W. Therapeutic methods of gut microbiota modification in colorectal cancer management - fecal microbiota transplantation, prebiotics, probiotics, and synbiotics. *Gut Microbes*. 2020;11:1518–1530. doi:10.1080/19490976.2020.1764309.
4. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, Chan FKL, Kristiansen K, Sung JY, Wong SH, et al. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome*. 2018;6(1):70. doi:10.1186/s40168-018-0451-2.
5. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, Fleck J, Voigt AY, Palleja A, Ponnudurai R, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med*. 2019;25:679–689. doi:10.1038/s41591-019-0406-6.
6. Xie YH, Gao QY, Cai GX, Sun XM, Sun XM, Zou TH, Yu S-Y, Qiu Y-W, Gu W-Q, Chen X-Y, et al. Fecal clostridium symbiosum for noninvasive detection of early and advanced colorectal cancer: test and validation studies. *EBioMedicine*. 2017;25:32–40. doi:10.1016/j.ebiom.2017.10.005.
7. Wong SH, Kwong TNY, Chow TC, Luk AKC, Dai RZW, Nakatsu G, Lam TYT, Zhang L, Wu JCY, Chan FKL, et al. Quantitation of faecal *Fusobacterium* improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut*. 2017;66:1441–1448. doi:10.1136/gutjnl-2016-312766.
8. Zhu A, Sunagawa S, Mende DR, Bork P. Inter-individual differences in the gene content of human gut bacterial species. *Genome Biol*. 2015;16:82. doi:10.1186/s13059-015-0646-9.
9. Chen Y, Li Z, Hu S, Zhang J, Wu J, Shao N, Bo X, Ni M, Ying X. Gut metagenomes of type 2 diabetic patients have characteristic single-nucleotide polymorphism distribution in *Bacteroides coprocola*. *Microbiome*. 2017;5:15. doi:10.1186/s40168-017-0232-3.
10. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015;10:845–858. doi:10.1038/nprot.2015.053.
11. Abdul NS, Nagiah S, Anand K, Chuturgoon AA. Molecular docking and mechanisms of fusaric acid induced mitochondrial sirtuin aberrations in glycolytically and oxidatively poised human hepatocellular

- carcinoma (HepG2) cells. *Toxicon*. 2020;173:48–56. doi:10.1016/j.toxicon.2019.11.009.
12. Butchko RA, Brown DW, Busman M, Tudzynski B, Wiemann P. Lae1 regulates expression of multiple secondary metabolite gene clusters in *Fusarium verticillioides*. *Fungal Genet Biol* 2012; 49:602–12.
 13. Ghazi T, Nagiah S, Chaturgoon AA. Fusaric acid decreases p53 expression by altering promoter methylation and m6A RNA methylation in human hepatocellular carcinoma (HepG2) cells. *Epigenetics*. 2020;1–13. doi:10.1080/15592294.2020.1788324.
 14. Stack BC Jr., Ye J, Willis R, Hubbard M, Hendrickson HP. Determination of oral bioavailability of fusaric acid in male sprague-dawley rats. *Drugs R D*. 2014;14:139–145. doi:10.1007/s40268-014-0051-y.
 15. Stack BC Jr., Hansen JP, Ruda JM, Jaglowski J, Shvidler J, Hollenbeak CS. Fusaric acid: a novel agent and mechanism to treat HNSCC. *Otolaryngol Head Neck Surg*. 2004;131:54–60. doi:10.1016/j.otohns.2004.02.038.
 16. Devnarain N, Tiloke C, Nagiah S, Chaturgoon AA. Fusaric acid induces oxidative stress and apoptosis in human cancerous oesophageal SNO cells. *Toxicon*. 2017;126:4–11. doi:10.1016/j.toxicon.2016.12.006.
 17. Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. Growth requirements and fermentation products of *Fusobacterium prausnitzii*, and a proposal to reclassify it as *Faecalibacterium prausnitzii* gen. nov. *comb nov Int J Syst Evol Microbiol*. 2002;52:2141–2146.
 18. Nilsen M, Madelen Saunders C, Leena Angell I, Arntzen MO, Lodrup Carlsen KC, Carlsen KH, Haugen G, Hagen LH, Carlsen MH, Hedlin G, et al. Butyrate levels in the transition from an infant- to an adult-like gut microbiota correlate with bacterial networks associated with eubacterium rectale and ruminococcus gnavus. *Genes (Basel)*. 2020;11:11. doi:10.3390/genes11111245.
 19. Mukherjee A, Lordan C, Ross RP, Cotter PD. Gut microbes from the phylogenetically diverse genus *Eubacterium* and their various contributions to gut health. *Gut Microbes*. 2020;12:1802866. doi:10.1080/19490976.2020.1802866.
 20. Chen J, Inflammation-Modulating VL. Effect of butyrate in the prevention of colon cancer by dietary fiber. *Clin Colorectal Cancer*. 2018;17:e541–e4. doi:10.1016/j.clcc.2018.05.001.
 21. Peng L, Li ZR, Green RS, Holzman IR, Lin J. Butyrate enhances the intestinal barrier by facilitating tight junction assembly via activation of AMP-activated protein kinase in Caco-2 cell monolayers. *J Nutr*. 2009;139:1619–1625. doi:10.3945/jn.109.104638.
 22. Yang J, Li D, Yang Z, Dai W, Feng X, Liu Y, Jiang Y, Li P, Li Y, Tang B, et al. Establishing high-accuracy biomarkers for colorectal cancer by comparing fecal microbiomes in patients with healthy families. *Gut Microbes*. 2020;11:918–929. doi:10.1080/19490976.20.1712986.
 23. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun*. 2015;6:6528. doi:10.1038/ncomms7528.
 24. Erawijantari PP, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, Saito Y, Fukuda S, Yachida S, Yamada T, et al. Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. *Gut*. 2020;69:1404–1415. doi:10.1136/gutjnl-2019-319188.
 25. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, Beghini F, Manara S, Karcher N, Pozzi C, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med*. 2019;25(4):667–678. doi:10.1038/s41591-019-0405-7.
 26. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, Fornelos N, Haiser HJ, Reinker S, Vatanen T, Hall AB, Mallick H, McIver LJ, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol*. 2019;4(2):293–305. doi:10.1038/s41564-018-0306-4.
 27. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490(7418):55–60. doi:10.1038/nature11450.
 28. Mitchell AL, Scheremetjew M, Denise H, Potter S, Tarkowska A, Qureshi M, Salazar GA, Pesseat S, Boland MA, Hunter F, et al. EBI Metagenomics in 2017: enriching the analysis of microbial communities, from sequence reads to assemblies. *Nucleic Acids Res*. 2018;46(D1):D726–D35. doi:10.1093/nar/gkx967.
 29. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlan2 for enhanced metagenomic taxonomic profiling. *Nat Methods*. 2015;12:902–903.
 30. Nayfach S, Rodriguez-Mueller B, Garud N, Pollard KS. An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res*. 2016;26:1612–1625. doi:10.1101/gr.201863.115.
 31. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant*. 2020;13:1194–1202. doi:10.1016/j.molp.2020.06.009.