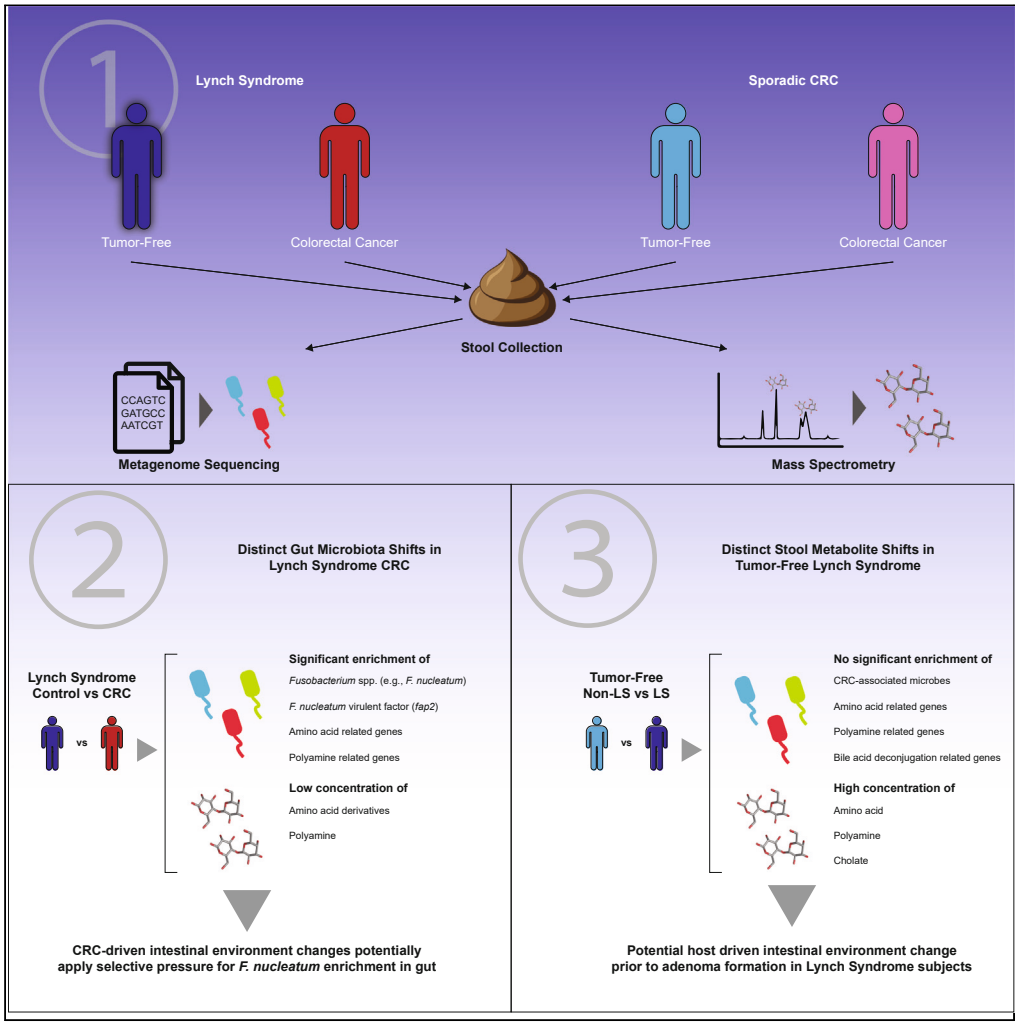
iScience

CellPress
OPEN ACCESS

Article

Fusobacterium species are distinctly associated with patients with Lynch syndrome colorectal cancer



Felix Salim, Sayaka Mizutani, Satoshi Shiba, ..., Shinji Fukuda, Shinichi Yachida, Takuji Yamada

syachida@cgi.med.osaka-u.ac.jp (S.Y.)
takuji@bio.titech.ac.jp (T.Y.)

Highlights

Lynch Syndrome (LS) subjects gut microbiota has lower diversity compared to non-LS

Distinct enrichment of *F. nucleatum* and *fap2* in LS CRC subjects

LS CRC gut microbiota exhibits increased lysine and arginine metabolism

Distinct fecal metabolites with minimal dysbiosis in LS subjects with no adenoma

Salim et al., iScience 27, 110181
July 19, 2024
https://doi.org/10.1016/j.isci.2024.110181

iScience

CellPress
OPEN ACCESS

Article

Fusobacterium species are distinctly associated with patients with Lynch syndrome colorectal cancer

Felix Salim,[1] Sayaka Mizutani,[1,2] Satoshi Shiba,[3] Hiroyuki Takamaru,[4] Masayoshi Yamada,[4] Takeshi Nakajima,[4] Tatsuo Yachida,[6] Tomoyoshi Soga,[7] Yutaka Saito,[4] Shinji Fukuda,[7,8,9,10,11,12] Shinichi Yachida,[5,13,15,*] and Takuji Yamada[1,11,12,14,15,16,*]

SUMMARY

Accumulating evidence demonstrates clear correlation between the gut microbiota and sporadic colorectal cancer (CRC). Despite this, there is limited understanding of the association between the gut microbiota and CRC in Lynch Syndrome (LS), a hereditary type of CRC. Here, we analyzed fecal shotgun metagenomic and targeted metabolomic of 71 Japanese LS subjects. A previously published Japanese sporadic CRC cohort, which includes non-LS controls, was utilized as a non-LS cohort (n = 437). LS subjects exhibited reduced microbial diversity and low-*Faecalibacterium* enterotypes compared to non-LS. Patients with LS-CRC had higher levels of *Fusobacterium nucleatum* and *fap2*. Differential fecal metabolites and functional genes suggest heightened degradation of lysine and arginine in LS-CRC. A comparison between LS and non-LS subjects prior to adenoma formation revealed distinct fecal metabolites of LS subjects. These findings suggest that the gut microbiota plays a more responsive role in CRC tumorigenesis in patients with LS than those without LS.

INTRODUCTION

Lynch Syndrome (LS) is an autosomal-dominant familial condition caused by a pathogenic germline mutation in DNA mismatch repair (MMR) genes (*MLH1*, *MSH2*, *PMS2*, and *MSH6*) or *EPCAM2*. Loss of function in MMR genes leads to the accumulation of gene mutations, thus increasing cancer risk in patients with LS, especially colorectal cancer (CRC) and endometrial cancer (EC).[1,2] The cancer risk in patients with LS varies depending on the MMR mutation variants, with a higher risk rate in LS carriers with *MLH1* and *MSH2* variants than in those with *MSH6* and *PMS2* variants.[3,4] A previous study reported that the epithelial cells have mutation rate similar to that of wild-type individuals prior to MMR deficiency, suggesting that the inactivation of the wild-type MMR allele is the key driver event in LS-associated CRC.[5]

The hallmarks of LS-associated CRC include MMR deficiency, rapid progression from colorectal adenoma to carcinoma, and improved prognosis.[1,2] MMR deficiency in LS-associated CRC is further associated with microsatellite instability (MSI), which has been reported as a potential marker of patient response to chemotherapy and immune checkpoint inhibitor treatment.[6] As LS-associated CRC progresses faster through the adenoma–carcinoma sequence,[1] LS carriers may serve as model cases to elucidate the role of the gut microbiota in MMR-deficient (dMMR) CRC tumorigenesis. However, this rapid progression may also diminish the effect of the gut microbiota on LS-associated CRC tumorigenesis.

[1]School of Life Science and Technology, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8550, Japan
[2]Research Fellow of Japan Society for the Promotion of Science, Tokyo, Japan
[3]Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan
[4]Endoscopy Division, National Cancer Center Hospital, Chuo-ku 104-0045, Tokyo, Japan
[5]Department of Cancer Genome Informatics, Graduate School of Medicine, Osaka University, Suita, Osaka 565-0871, Japan
[6]Department of Gastroenterology & Neurology, Faculty of Medicine, Kagawa University, Miki-cho, Kagawa 761-0793, Japan
[7]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0052, Japan
[8]Gut Environmental Design Group, Kanagawa Institute of Industrial Science and Technology, Kawasaki, Kanagawa 210-0821, Japan
[9]Transborder Medical Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8575, Japan
[10]Laboratory for Regenerative Microbiology, Juntendo University Graduate School of Medicine, Bunkyo-ku, Tokyo 113-8421, Japan
[11]Metagen, Inc., Tsuruoka, Yamagata 997-0052, Japan
[12]Metagen Theurapeutics, Inc., Tsuruoka, Yamagata 997-0052, Japan
[13]Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives (OTRI), Osaka University, Suita, Osaka 565-0871, Japan
[14]digzyme, Inc., Minato-ku, Tokyo 105-0004, Japan
[15]These authors contributed equally
[16]Lead contact
*Correspondence: syachida@cgi.med.osaka-u.ac.jp (S.Y.), takuji@bio.titech.ac.jp (T.Y.)
https://doi.org/10.1016/j.isci.2024.110181

Several reports have associated the human gut microbiota with CRC. Early studies have shown that specific microbes, such as *Fusobacterium nucleatum*, are enriched in the tumor tissues of patients with CRC compared to the neighboring non-tumorous tissues, with later studies reporting other microbial CRC markers, such as *Peptostreptococcus stomatis* and *Parvimonas micra*.[7–10] Mouse model experiments have elucidated the potential role of human gut microbes in CRC tumorigenesis. For example, *Streptococcus thermophilus* shows a tumor-suppressive effect via β-galactosidase secretion.[11] *Peptostreptococcus anaerobius* promotes tumorigenesis by interacting with colon cells and triggering increased cell proliferation and a proinflammatory immune microenvironment.[12] Furthermore, some bacterial strains may produce genotoxins, such as pks+ *Escherichia coli* colibactin, which are linked to an increased CRC risk.[13] Colibactin has also been reported to worsen MMR deficiency-associated mutations.[14]

Recent studies have shed light on the association between gut microbiota and dMMR CRC and LS-CRC. Mouse model studies have shown that microbial metabolites such as butyrate and the gut microbiota contribute to CRC tumorigenesis in MSH2-deficient mice.[15,16] Studies of patients with dMMR CRC and MSI-H CRC have reported distinct *Fusobacterium* enrichment in dMMR CRC gut microbiota and MSI-H CRC tissue samples.[17–19] A study on patients with LS and colorectal adenoma reported similar gut microbiota dysbiosis in LS adenoma and sporadic CRC but no *Fusobacterium* enrichment.[20] Several other studies have also reported distinct gut microbiota compositions between patients with and without LS, as well as between patients with LS and a CRC history and those without any CRC history.[21–24] However, as most previous studies have been based on 16S rRNA amplicon sequencing, there is limited information on how the composition and function of the gut microbiota differ between patients with LS-CRC compared to those with LS and no prior adenoma or carcinoma.

In this study, we performed shotgun metagenomics and targeted metabolomics on 71 stool samples collected from 71 patients with LS (patients with pathogenic variants of germline mutations in one or more MMR genes). Patients with LS were divided into four groups: Patients with no prior history of colorectal adenoma or carcinoma formation (LS control), those with colorectal adenoma formation (LS-adenoma), those with colorectal carcinoma formation (LS-CRC), and those with no current colorectal adenoma or carcinoma but who underwent a colectomy procedure (LS-surgery). Controls' microbiota and metabolome profiles were compared with those of LS-adenoma and LS-CRC groups to identify potential associations between the gut microbiota and tumorigenesis within the LS cohort. Furthermore, to identify potential LS-specific microbiota–carcinogenesis associations, associations identified within the LS cohort were compared with associations identified within a sporadic CRC cohort.

RESULTS

Overview of study subjects

Patients undergoing colonoscopy at the National Cancer Center Hospital, Tokyo, Japan, were enrolled in the study, totaling 71 patients with Lynch Syndrome. The inclusion criterion was a positive diagnosis of a pathogenic mutation in germline MMR genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*, or *EPCAM* + *MSH2*). Patients with a history of gastrectomy were excluded. Patients' stool samples were collected after bowel preparation on the day of colonoscopy and stored at −80°C before further processing.

Patients with LS were grouped into adenoma (LS-adenoma), CRC (LS-CRC), and post-colectomy (LS-surgery) groups based on the colonoscopy results and clinical records. Those without any adenoma, carcinoma, or history of colectomy were defined as controls (LS control). The lifestyle and dietary data of some patients were acquired using questionnaires based on a previous study.[25] Potential confounding factors such as age, sex, mutated MMR genes, smoking habits, meat consumption, and dietary fiber consumption were determined from clinical and questionnaire data and compared across the groups (Table 1). The age distribution between the groups was significantly different (ANOVA, $p = 0.045$; Table 1). LS-CRC had a higher male-to-female ratio, smoker-to-non-smoker ratio, and Brinkman index than LS control, but the difference was not significant. No significant differences were observed between patients with LS in mutated MMR genes and dietary fiber or meat intake.

Data on patients with sporadic CRC were obtained from previously published studies with some modifications.[26] Based on colonoscopy results, patients with sporadic CRC were grouped into five groups: (1) controls (no remarkable colonoscopic findings or records of adenoma); (2) multiple polypoid adenomas with low-grade dysplasia (more than three adenomas, mostly more than five adenomas); (3) intramucosal carcinoma (polypoid adenoma(s) with high-grade dysplasia), stage 0/pTis CRC (S0); (4) early CRC (Stages I or II); (5) advanced CRC (Stages III or IV). As previously mentioned, potential confounding factors such as age, sex, smoking habit, meat consumption, and dietary fiber consumption were determined from clinical and questionnaire data and compared across subject groups (Table S1). Similar to the LS cohort, the controls in the sporadic cohort were significantly younger than those with progressive CRC. Furthermore, the Brinkman Index was significantly higher in patients with CRC progression (Table S1).

Lynch syndrome gut microbiome has low alpha diversity and *Faecalibacterium* depleted enterotypes

Species alpha diversity, estimated using the Shannon–Wiener index, was significantly lower in patients with LS than those with sporadic CRC (Mann–Whitney U $p = 0.00014$, Figure 1A). Next, pairwise comparisons were made within the LS cohort to explore the potential association between alpha diversity and patient group, MMR mutations, or age. Based on the comparison results, no significant association was observed between alpha diversity and subject group, age, or genotype (Figure S1A).

Principal coordinate analysis (PCoA) was performed on Bray–Curtis dissimilarities of the patient's genus-level profile to visualize the gut microbiota composition (Figure 1B). In the PCoA plot, no clear separation was observed between the patients with LS and those with sporadic CRC. Focusing on the gut microbiota profile of the groups, the PCoA plot showed no clear separation between the LS control, LS-adenoma, LS-CRC, and LS-surgery groups (Figure 1C).

Table 1. Lynch Syndrome subject's clinical information

	LS-CTR[a]	LS-ADE[b]	LS-CRC[c]	LS-SUR[d]	P[g]
Subject Number	18	23	11	19	
Age (mean (SD))	39.56 (11.13)	50.74 (13.70)	46.82 (19.11)	52.95 (13.33)	0.025
Sex = Male (%)	5 (27.8)	11 (47.8)	7 (63.6)	7 (36.8)	0.248
BMI[e] (mean (SD))	21.16 (2.47)	24.12 (4.89)	21.88 (3.94)	24.06 (4.57)	0.089
Brinkman Index (mean (SD))	97.65 (163.08)	161.33 (227.30)	344.00 (372.71)	235.79 (389.48)	0.178
Total dietary fiber intake (mean (SD))	11.68 (6.11)	10.39 (7.38)	14.66 (6.78)	11.28 (4.68)	0.406
Meat intake (mean (SD))	90.44 (48.59)	90.23 (87.41)	102.45 (74.44)	72.28 (44.34)	0.68
Mutated MMR[f] variants (%)					0.319
MLH1	6 (33.3)	10 (43.5)	2 (18.2)	7 (36.8)	
MSH2	11 (61.1)	10 (43.5)	5 (45.5)	7 (36.8)	
MSH2+EPCAM	0 (0.0)	1 (4.3)	1 (9.1)	1 (5.3)	
MSH6	0 (0.0)	2 (8.7)	0 (0.0)	2 (10.5)	
PMS2	1 (5.6)	0 (0.0)	3 (27.3)	2 (10.5)	
Colectomy history = yes (%)	0 (0.0)	8 (34.8)	3 (27.3)	19 (100.0)	<0.001
Colectomy part (%)					<0.001
none	18 (100.0)	15 (65.2)	8 (72.7)	0 (0.0)	
right	0 (0.0)	3 (13.0)	2 (18.2)	11 (57.9)	
left	0 (0.0)	2 (8.7)	1 (9.1)	4 (21.1)	
both	0 (0.0)	3 (13.0)	0 (0.0)	4 (21.1)	

[a]LS-CTR = Lynch Syndrome (LS) subjects without adenoma, carcinoma, or history of colectomy.
[b]LS-ADE = LS subjects with colorectal adenoma.
[c]LS-CRC = LS subjects with CRC.
[d]LS-SUR = LS subjects with colectomy history, but no current colorectal adenoma or CRC.
[e]BMI = body mass index.
[f]MMR = mismatch repair.
[g]One-way analysis of variance (one-way ANOVA) test was used for numerical variables and chi-square test was used for categorical variables.

Permutational analysis of variance (PERMANOVA) was performed to verify whether LS affected the gut microbiota composition. Patient grouping (four LS groups and five sporadic CRC groups) was found to be the best explanatory variable (R^2 = 1.885%, p = 0.161), followed by LS diagnosis (R^2 = 0.328%, p = 0.110) (Figure S1B, left panel). PERMANOVA within the LS cohort showed that the best explanatory variables were patient groups (R^2 = 5.586%, p = 0.153) and MMR gene mutations (R^2 = 5.403%, p = 0.514) (Figure S1B, right panel). Similar to the previous study, while colectomy history was less explanatory compared to other variables (R^2 = 1.103%, p = 0.613), focusing on samples with colectomy history showed that colon parts removed by colectomy were the best explanatory variables (R^2 = 6.673%, p = 0.470) (Figure S1B, right panel).[20]

To check whether LS-associated enterotypes existed, the gut microbiota profile of each patient was fitted into a Dirichlet Multinomial Mixture (DMM) model. From the best-fit model, four enterotypes were determined, consisting of three *Bacteroides*-dominant enterotypes (Enterotypes 1, 2, and 3) and one *Prevotella*-dominant enterotype (Enterotype 4) (Figures 1D and S1C). Notably, Enterotypes 1 and 4 were characterized by a higher proportion of *Faecalibacterium* than the other two enterotypes (Figure 1E). Comparison of enterotype ratios between the LS and sporadic cohorts showed that the LS cohort had a significantly higher ratio of Enterotype 2 (Fisher's exact test $p = 8.50 \times 10^{-4}$) and a significantly lower ratio of Enterotypes 1 (p = 0.0102) and 4 (p = 0.0139) compared to the sporadic cohort (Figure 1F). However, *Faecalibacterium* abundance in the LS and sporadic cohorts was not significantly different (Figure S1D).

Overall, although PCoA visualization and PERMANOVA results showed no clear separation between the gut microbiota of patients with sporadic CRC and those with LS, we observed lower alpha diversity and a higher prevalence of *Faecalibacterium* poor enterotype in patients with LS. Both low alpha diversity and *Faecalibacterium* poor enterotypes have been associated with inflammatory bowel disease (IBD).[27,28]

Fusobacterium species enriched in the gut microbiota of patients with Lynch syndrome

To identify the gut microbial species that were differentially abundant between dysbiosis occurring in patients with and without LS after adenoma and CRC formation in more detail, we performed a differential abundance analysis within non-LS and LS cohorts.

A pairwise comparison of species abundances between the LS controls and other groups detected 105 species with differential abundances (Mann–Whitney U, $p < 0.05$) in at least one comparison (Table S2). Focusing on the differentially abundant species between the LS control and LS-CRC groups, 19 were enriched, and 29 were depleted in the LS-CRC group (Figure 2A). Of the 19 species enriched in LS-CRC, 12 belonged to the *Fusobacterium* genus, such as *F. nucleatum* subsp. *nucleatum* ($p = 6.84 \times 10^{-4}$) and *F. nucleatum* subsp. *animalis*

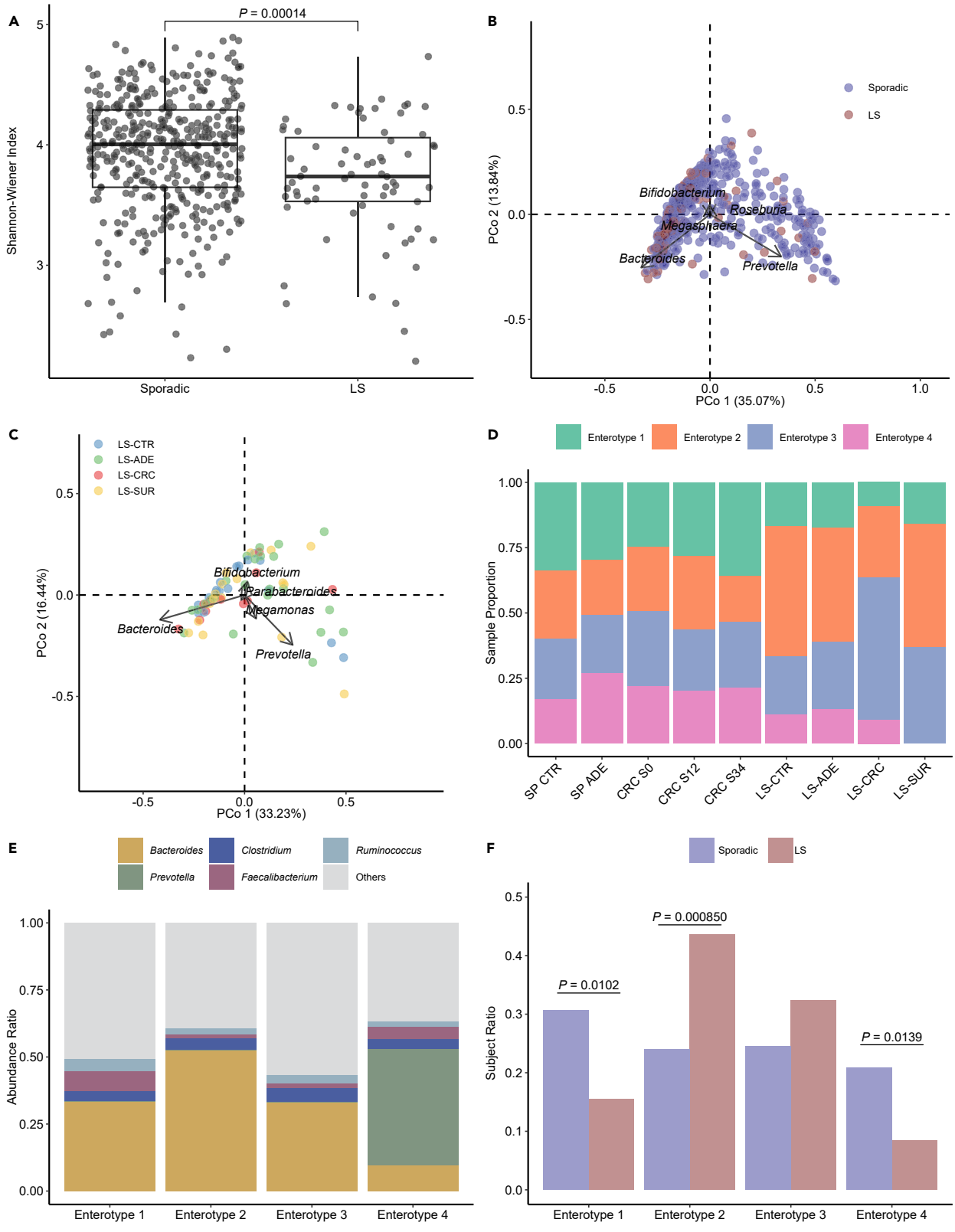

A
P = 0.00014
Shannon-Wiener Index
Sporadic
LS
B
Sporadic
LS
PCo 2 (13.84%)
PCo 1 (35.07%)
Bifidobacterium
Roseburia
Megasphaera
Bacteroides
Prevotella
C
LS-CTR
LS-ADE
LS-CRC
LS-SUR
PCo 2 (16.44%)
PCo 1 (33.23%)
Bifidobacterium
Parabacteroides
Megamonas
Bacteroides
Prevotella
D
Enterotype 1
Enterotype 2
Enterotype 3
Enterotype 4
Sample Proportion
SP CTR
SP ADE
CRC S0
CRC S12
CRC S34
LS-CTR
LS-ADE
LS-CRC
LS-SUR
E
Bacteroides
Clostridium
Ruminococcus
Prevotella
Faecalibacterium
Others
Abundance Ratio
Enterotype 1
Enterotype 2
Enterotype 3
Enterotype 4
F
Sporadic
LS
Subject Ratio
P = 0.0102
P = 0.000850
P = 0.0139
Enterotype 1
Enterotype 2
Enterotype 3
Enterotype 4

Figure 1. Gut microbiota composition overview

(A) Alpha-diversity (Shannon–Weiner Index) comparison between the LS and non-LS cohorts. $p < 0.05$ (Mann-Whitney U test) is considered statistically significant.

(B) Principal coordinate analysis (PCoA) based on the genus profiles of all patients.

(C) Principal coordinate analysis (PCoA) based on the genus profiles of patients with LS.

(D) Enterotype distribution within the LS and non-LS cohorts. The y axis represents the proportion of patient enterotypes in each group.

(E) Distribution of the top five most abundant genera within each enterotype.

(F) Comparison of non-LS and LS patient proportions for each enterotype. $p < 0.05$ (Fisher's exact test) is considered statistically significant.

($p = 8.84 \times 10^{-4}$). Other CRC-related species, such as *Peptostreptococcus stomatis* ($p = 0.738$), *P. anaerobius* ($p = 0.723$), *Parvimonas micra* ($p = 0.383$), and *Gemella morbillorum* ($p = 0.126$), were not significantly enriched in LS-CRC (Figure S2A).

Although not significant, we identified similar dysbiosis trends, e.g., *Fusobacterium nucleatum* enrichment, in LS subjects with cancer history compared to LS subjects with no cancer history in an independent Israelian LS cohort[24] (Table S2).

Machine learning-based feature selection validated *Fusobacterium* species enrichment as an Lynch syndrome-colorectal cancer marker

Our results suggest *Fusobacterium* spp. enrichment is a shared CRC-related dysbiosis across LS and sporadic CRC. However, our small sample size had limited detection power. Furthermore, no previous studies using whole-genome shotgun sequences of LS-CRC have validated our results. In addition, the results are not necessarily reproducible in other taxonomic annotation pipelines. However, accumulated knowledge has been obtained from a considerable amount of microbial data on sporadic cancer.[29–31] To examine whether the microbiome features of sporadic CRC characterize LS-CRC, we used a sporadic CRC classifier to screen for potential dysbiosis markers in LS-CRC. Based on the screening results, we hoped to find the resemblance between the gut microbiota profiles of sporadic and LS-CRC or lack of such resemblance, especially whether *Fusobacterium* species were enriched in LS -CRC.

We trained a gradient-boosted decision tree (GBDT) classifier to differentiate between non-LS controls and patients with sporadic CRC based on their species profiles. The trained classifier performed comparable to previous CRC studies,[26,29,30] as evaluated by its mean area under the receiver operating curve (AU-ROC) across 5-fold cross-validation (mean AU-ROC = 0.805; Figure S2B). Feature importance analysis showed that the top contributors were mostly CRC-related species, such as *F. nucleatum* subsp. *nucleatum*, *Gemella morbillorum*, *P. stomatis*, and *P. micra* (Figure S2D).

The constructed sporadic CRC classifier was tested for its ability to discriminate LS CRC from LS controls using the species profiles of patients with LS as the input. The model distinguished LS-adenoma, LS-CRC, and LS-surgery from LS control with AU-ROC of 0.575, 0.869, and 0.684, respectively (Figure 2B). The normalized CRC probability predicted by the classifier was higher for LS-CRC and lower for LS controls, whereas LS-adenoma and LS-surgery showed a more uniform distribution (Figure S2C). These results indicate the potential of sporadic CRC microbiota-based characterization of LS CRC microbiota.

Local feature importance analysis identified *Fusobacterium nucleatum* subsp. *nucleatum* as the top microbial marker for Lynch syndrome-colorectal cancer

The classifier performance and feature importance analysis results indicated that CRC-related species abundance was the main predictor for differentiating between LS-CRC and LS controls. However, it was not sufficient to screen for LS CRC markers because feature importance was estimated for the entire classifier (global importance), which was trained on sporadic species profiles.

A previous study proposed a framework utilizing the Shapley Additive Explanations (SHAP) value to estimate feature importance for specific subjects (local importance) and showed distinct clusters within patients with sporadic CRC based on the estimated local importance.[31] Here, we used the SHAP value-based local feature importance analysis framework to validate whether *Fusobacterium* spp. were the main predictors of LS-CRC.

The PCA of the estimated SHAP values showed a distinct separation between control subjects and patients with CRC (Figure 2C). Next, the SHAP value principal components of correctly predicted patients with CRC were clustered via k-means clustering (optimal k = 4, determined by an elbow plot [Figure S2E]). Of the four clusters, only Clusters 2 and 4 contained subjects with LS-CRC (Figure 2D). In contrast, patients with sporadic CRC spread over the four clusters.

From each cluster's LS-CRC patient SHAP value, we found that the top-ranking species in Cluster 2 were CRC-related microbiota, such as *G. morbillorum*, *F. nucleatum* subsp. *nucleatum*, *P. micra*, and *P. stomatis* (Figure 2E, left panel). The top-ranking species in Cluster 4 were similar to those in Cluster 2, with a highly abundant *F. nucleatum* subsp. *nucleatum* as the top contributor (Figure 2E, right panel). Overall, the local feature importance analysis suggests that the sporadic CRC classifier distinguished LS-CRC from LS control, mainly based on *Fusobacterium* spp. presence, especially *F. nucleatum* subsp. *nucleatum*.

Enriched *fap2* in patients with Lynch syndrome-colorectal cancer

F. nucleatum genes have been reported as virulence factors (Fn VF) associated with CRC progression, such as *fadA*, *fap2*, *cbpF*, *radD*, and *FnDps*.[32–37] Here, we explored the LS metagenome data to identify the potential enrichment of these virulence factors.

Four out of seven Fn VFs were detected in our LS cohort, which were *fap2* (average prevalence 5.0%), *radD* (average prevalence 35.7%), *fadA* (average prevalence 7.1%), and *cbpF* (average prevalence 10.4%), while *fadA2*, *fadA3*, and *FnDps* were not detected in our LS cohort

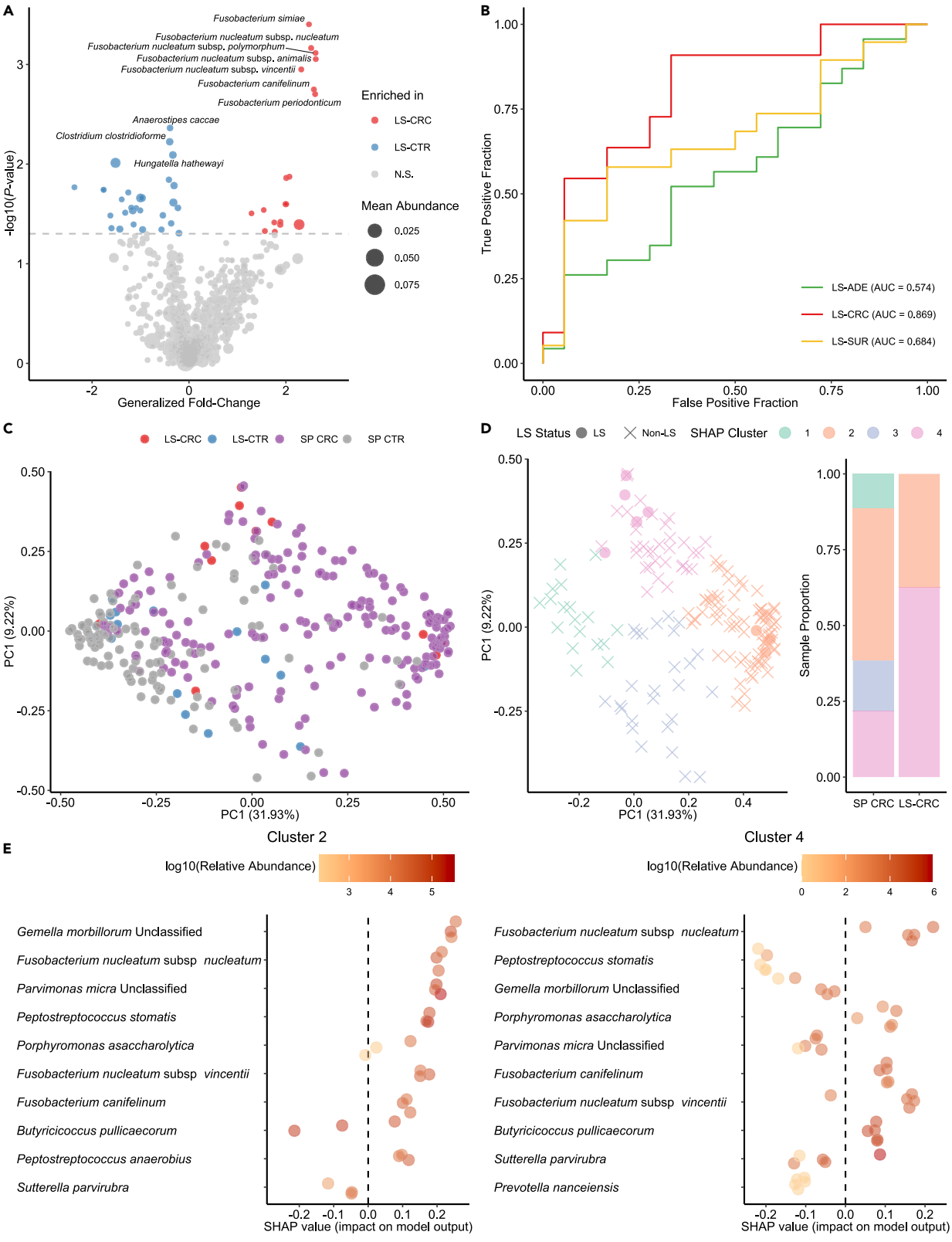

A
Fusobacterium simiae
Fusobacterium nucleatum subsp. nucleatum
Fusobacterium nucleatum subsp. polymorphum
Fusobacterium nucleatum subsp. animalis
Fusobacterium nucleatum subsp. vincentii
Fusobacterium canifelinum
Fusobacterium periodonticum
Anaerostipes caccae
Clostridium clostridioforme
Hungatella hathewayi
-log10(P-value)
Generalized Fold-Change
Enriched in
LS-CRC
LS-CTR
N.S.
Mean Abundance
0.025
0.050
0.075
B
True Positive Fraction
False Positive Fraction
LS-ADE (AUC = 0.574)
LS-CRC (AUC = 0.869)
LS-SUR (AUC = 0.684)
C
LS-CRC
LS-CTR
SP CRC
SP CTR
PC1 (9.22%)
PC1 (31.93%)
D
LS Status
LS
Non-LS
SHAP Cluster
1
2
3
4
PC1 (9.22%)
PC1 (31.93%)
Sample Proportion
SP CRC
LS-CRC
E
Cluster 2
log10(Relative Abundance)
Gemella morbillorum Unclassified
Fusobacterium nucleatum subsp nucleatum
Parvimonas micra Unclassified
Peptostreptococcus stomatis
Porphyromonas asaccharolytica
Fusobacterium nucleatum subsp vincentii
Fusobacterium canifelinum
Butyricicoccus pullicaecorum
Peptostreptococcus anaerobius
Sutterella parvirubra
SHAP value (impact on model output)
Cluster 4
log10(Relative Abundance)
Fusobacterium nucleatum subsp nucleatum
Peptostreptococcus stomatis
Gemella morbillorum Unclassified
Porphyromonas asaccharolytica
Parvimonas micra Unclassified
Fusobacterium canifelinum
Fusobacterium nucleatum subsp vincentii
Butyricicoccus pullicaecorum
Sutterella parvirubra
Prevotella nanceiensis
SHAP value (impact on model output)

Figure 2. *Fusobacterium spp.* enriched in patients with LS CRC and acts as CRC marker

(A) Comparison of relative species abundance between the LS controls (n = 18) and LS-CRC (n = 11). The x axis represents the generalized fold-change between LS-CRC and LS controls, and the y axis represents the negative log10 transformed p-values of the two-sided Mann–Whitney U test. The horizontal dotted line shows a -log10 transformed p-value of 0.05. The sizes of the dots indicate their relative abundances.

(B) ROC-AUC plot of LS classification task between LS controls and LS-adenoma, LS-CRC, or LS-surgery.

(C) Principal component analysis (PCA) of estimated SHAP values, annotated by the LS and CRC status of a patient.

(D) Principal component analysis (PCA) of estimated SHAP values, annotated by k-means cluster of a patient. Bar plot shows prevalence ratio of clusters within sporadic CRC and LS CRC subjects.

(E) Waterfall plot of species contribution to the CRC probability of patients with LS-CRC stratified for each LS CRC patient cluster (Clusters 2 and 4, Figure 2D). The x axis represents the SHAP values and the y axis is ordered by the sum of the absolute SHAP values. Dotted colors indicate the relative abundance of the species in each subject.

(Figures 3A and S3A). Of these four detected genes, *fap2* (Mann–Whitney U, $p = 0.008$) was significantly enriched in LS-CRC compared to LS controls, whereas the other three genes were not significantly different ($p > 0.05$) in both LS controls and LS-CRC subjects (Figure 3A). In comparison, all seven CRC-associated genes were detected in the sporadic CRC cohort, with five genes significantly enriched in CRC subjects, especially those with late-stage CRC subjects (Figure S3B).

Gene prevalence was similar between the LS and sporadic cohorts (Figure S3A), indicating a low number of patients with LS as a potential reason for the absence of *fadA2*, *fadA3*, and *FnDps* in the LS cohort. Despite the low sample numbers, *fap2* was highly prevalent in LS CRC and was enriched compared to LS controls. As *fap2* has been reported to play a role in host–microbe adhesion (GalNAc)[33] and immune system modulation (TIGIT),[38] a high abundance of *fap2* may be an indicator of intratumoral *Fusobacterium* species and an immunogenic tumor microenvironment.

Escherichia coli and colibactin biosynthesis gene cluster showed no significant change in patients with Lynch syndrome

Other than *Fusobacterium nucleatum* and its adhesins, other microbial genes have also been reported to have a causal relationship with CRC tumorigenesis. Recently, a study reported that colibactin, a genotoxin produced by pks+ *Escherichia coli*, exacerbates dMMR-related mutations.[14] As dMMR is a hallmark of LS,[2] we explored our metagenomic data to identify potential colibactin biosynthesis-related gene enrichment in patients with LS.

Pairwise comparisons of *E. coli* abundance and colibactin biosynthesis-related genes between LS controls and other groups showed no significant changes (Table S3). A similar analysis of sporadic CRC cohorts also showed no significant changes (Table S3).

Gut microbe under Lynch syndrome-colorectal cancer had higher amino acid degradation and polyamine-related metabolism

In addition to direct cell adhesion, gut microbiota-derived metabolites and other metabolic activities may affect the host. To explore the functions of gut microbial genes, we quantified the relative abundance of KEGG Orthologies (KOs). To investigate differences or similarities in other gut microbiota metabolic activities, pairwise comparisons were performed on fecal metabolite concentrations and KO-level functional gene profiles.

A pairwise comparison of the fecal metabolite concentrations between the LS controls and other groups showed 65 metabolites with differential concentrations in at least one comparison (Table S4). Focusing on LS controls and LS-CRC results, three metabolites (glycerophosphorylcholine ($p = 0.00531$), glycerol ($p = 0.00517$), and betaine aldehyde ($p = 0.01960$)) had significantly higher concentrations in LS-CRC. In contrast, 25 metabolites, such as gamma-aminobutyric acid (GABA) ($p = 0.00252$), malonate ($p = 0.00879$), and 2-hydroxyglutarate ($p = 0.00501$), had significantly lower concentrations (Figure 3B). A pairwise comparison of KOs between the LS controls and other groups detected 753 KOs with differential abundance in at least one comparison (Table S5). Focusing on LS control and LS-CRC comparisons, 275 KOs were significantly more abundant and 62 KOs had a significantly lower abundance in LS-CRC ($p < 0.05$) (Figure 3C).

Overrepresentation analysis (ORA) of the KEGG module was performed with fecal metabolites and KOs, showing statistical significance ($p < 0.05$) to determine whether any metabolic function was significantly altered. At KEGG MODULE level, nine modules were significantly over-represented ($q < 0.1$), including "Lysine degradation, L-lysine => succinate" (M00956) ($q = 0.0295$) and "GABA shunt" (M00027) ($q = 0.0573$) (Figure 3D). ORA was performed using the EnteroPathway module (EPM) to cover functions undefined in the KEGG MODULE. Three modules ("Lysine degradation, L-lysine => acetoacetyl-coA, butyril-coA" (EPM0634) ($q = 0.0251$), "Arginine degradation" (EPM0734) ($q = 0.0313$), and "Ornithine degradation, ornithin => putrescine" (EPM0739) ($q = 0.0768$)) were significantly overrepresented (Figure 3E).

Overrepresented metabolic modules and components with differential abundance showed that both metabolites and related KOs were altered in LS-CRC (Figures 3F, S3E, and S3F). For example, lysine decarboxylase *ldcC* (K01582), which is involved in the first step of the L-lysine degradation pathway, and two other KOs (*gabT* (K07250) and *gabD* (K00135)) that are involved in subsequent reactions in this pathway, were highly abundant in the LS-CRC group. In addition, the fecal concentrations of 5-aminovalerate and 2-hydroxyglutarate, two intermediate metabolites of the L-lysine degradation pathway, were lower in the LS-CRC group. In addition to l-lysine degradation, *gabT* (K07250) and *gabD* (K00135) are also involved in GABA shunt formation. Furthermore, fecal GABA concentrations were lower in patients with LS-CRC.

The l-lysine degradation module, which leads to butyrate production, was also overrepresented. The beta-lysine 5,6-aminomutase complex (*kamD* [K01844] and *kamE* [K18011]), L-*erythro*-3,5-diaminohexanoate dehydrogenase (*kdd* [K18012]), and acetate CoA/acetoacetate CoA-transferase alpha subunit (*atoD* [K01034]) were enriched in LS-CRC.

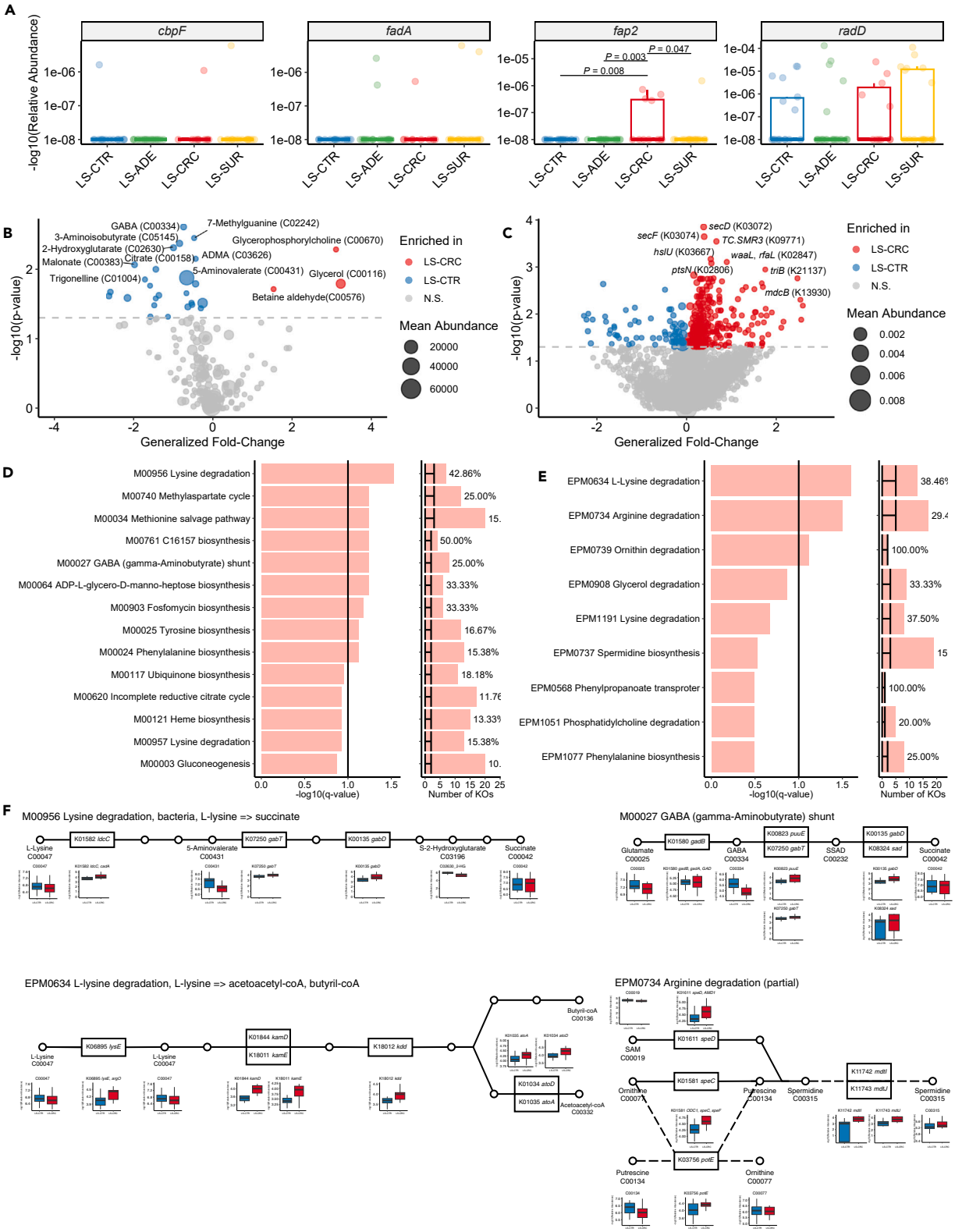

A
cbpF
fadA
fap2
radD
-log10(Relative Abundance)
P = 0.003
P = 0.008
P = 0.047
LS-CTR
LS-ADE
LS-CRC
LS-SUR
B
GABA (C00334)
7-Methylguanine (C02242)
3-Aminoisobutyrate (C05145)
Glycerophosphorylcholine (C00670)
2-Hydroxyglutarate (C02630)
ADMA (C03626)
Malonate (C00383)
Citrate (C00158)
5-Aminovalerate (C00431)
Glycerol (C00116)
Trigonelline (C01004)
Betaine aldehyde(C00576)
-log10(p-value)
Generalized Fold-Change
Enriched in
LS-CRC
LS-CTR
N.S.
Mean Abundance
20000
40000
60000
C
secD (K03072)
secF (K03074)
TC.SMR3 (K09771)
hslU (K03667)
waaL, rfaL (K02847)
ptsN (K02806)
triB (K21137)
mdcB (K13930)
-log10(p-value)
Generalized Fold-Change
Enriched in
LS-CRC
LS-CTR
N.S.
Mean Abundance
0.002
0.004
0.006
0.008
D
M00956 Lysine degradation
M00740 Methylaspartate cycle
M00034 Methionine salvage pathway
M00761 C16157 biosynthesis
M00027 GABA (gamma-Aminobutyrate) shunt
M00064 ADP-L-glycero-D-manno-heptose biosynthesis
M00903 Fosfomycin biosynthesis
M00025 Tyrosine biosynthesis
M00024 Phenylalanine biosynthesis
M00117 Ubiquinone biosynthesis
M00620 Incomplete reductive citrate cycle
M00121 Heme biosynthesis
M00957 Lysine degradation
M00003 Gluconeogenesis
42.86%
25.00%
50.00%
25.00%
33.33%
33.33%
16.67%
15.38%
18.18%
15.38%
-log10(q-value)
Number of KOs
E
EPM0634 L-Lysine degradation
EPM0734 Arginine degradation
EPM0739 Ornithin degradation
EPM0908 Glycerol degradation
EPM1191 Lysine degradation
EPM0737 Spermidine biosynthesis
EPM0568 Phenylpropanoate transproter
EPM1051 Phosphatidylcholine degradation
EPM1077 Phenylalanine biosynthesis
38.46%
100.00%
33.33%
37.50%
100.00%
20.00%
25.00%
-log10(q-value)
Number of KOs
F
M00956 Lysine degradation, bacteria, L-lysine => succinate
K01582 kdcC
K07250 gabT
K00135 gabD
L-Lysine C00047
5-Aminovalerate C00431
S-2-Hydroxyglutarate C03196
Succinate C00042
M00027 GABA (gamma-Aminobutyrate) shunt
K01580 gadB
K00823 puuE
K07250 gabT
K00135 gabD
K08324 sad
Glutamate C00025
GABA C00334
SSAD C00232
Succinate C00042
EPM0634 L-lysine degradation, L-lysine => acetoacetyl-coA, butyril-coA
K06895 lysE
K01844 kamD
K18011 kamE
K18012 kdd
K01034 atoD
K01035 atoA
L-Lysine C00047
Butyril-coA C00136
Acetoacetyl-coA C00332
EPM0734 Arginine degradation (partial)
K01611 speD
K01581 speC
K03756 potE
K11742 mdtI
K11743 mdtJ
SAM C00019
Ornithine C00077
Putrescine C00134
Spermidine C00315

Figure 3. Enriched *fap2* and overrepresentation of metabolites and genes involved in lysine and arginine degradation in LS CRC

(A) Boxplots of *Fusobacterium nucleatum* virulence factor abundance. Statistically significant results (Mann–Whitney U test $p < 0.05$) were annotated.

(B) Comparison of fecal metabolite concentrations between LS controls ($n = 18$) and LS-CRC ($n = 11$). The x axis represents the generalized fold-change between LS CRC and LS controls, and the y axis represents the negative log10 transformed *p*-values of the two-sided Mann–Whitney *U* test. The horizontal dotted line shows the -log10 transformed *p*-value of 0.05. The size of the dots indicates relative abundance.

(C) Comparison of KO relative abundances between LS controls ($n = 18$) and LS-CRC ($n = 11$). The x axis represents the generalized fold-change between LS-CRC and LS controls, and the y axis represents the negative log10 transformed *p*-values of the two-sided Mann–Whitney *U* test. The horizontal dotted line shows the -log10 transformed *p*-value of 0.05. The size of the dots indicates relative abundance.

(D) Overrepresented KEGG metabolic modules (KEGG MODULE) in LS CRC (*p*-value <0.05). Nine KEGG MODULEs showed significant overrepresentation in LS CRC (FDR *p*-value <0.1). Modules are arranged in ascending order based on *p*-value. The right figure shows the number of Kos in each module (red bar) and the percentage of enriched Kos within the module (black line).

(E) Overrepresented EnteroPathway metabolic modules in LS CRC (*p*-value <0.05). Three EnteroPathway MODULEs showed significant overrepresentation in LS CRC (FDR *p*-value <0.1). Modules are arranged in ascending order based on *p*-value. The right figure shows the number of KOs in each module (red bar) and the percentage of enriched KOs within the module (black line).

(F) Diagrams of metabolic modules overrepresented in LS CRC compared to LS controls. Circular nodes represent key metabolites or statistically significant metabolites. Rectangular nodes represent key KOs or statistically significant KOs. Statistical significance threshold was defined as Mann–Whitney U $p < 0.05$.

In addition to L-lysine degradation, ornithine decarboxylase *speC/F* (K01581) and putrescine: ornithine antiporter *potE* (K03756), which are involved in putrescine formation and transport, were also enriched in LS-CRC. In addition, S-adenosylmethionine decarboxylase *speD* (K01611) and spermidine export proteins *mdtI* (K11742) and *mdtJ* (K11743), which are involved in spermidine biosynthesis and export, were enriched in the LS-CRC.

We also explored the metagenome profile of an independent Israelian LS cohort[24] to validate our findings, and observed significant over-representation of "GABA shunt" module. Over-representation of lysine degradation, arginine degradation, or polyamine biosynthesis was not observed (Table S5). This might be due to the lack of subjects with ongoing CRC in the Israel cohort, indicating CRC-specific intestinal environment as the main driver of gut microbiota function differences.

Overall, differential fecal metabolites and functional genes between LS-CRC and LS controls indicated potential shifts in gut microbiota metabolism, specifically higher lysine degradation and polyamine biosynthesis.

Patients with Lynch syndrome have distinct fecal metabolite patterns prior to adenoma formation, but minimal gut microbe changes

Next, we compared LS controls (n = 18) and non-LS controls (n = 16) to examine the gut microbiota and fecal metabolite changes that occurred prior to adenoma formation in patients with LS. To account for sample number imbalance and confounding factors (Figure S4A), non-LS control samples were matched to LS control samples based on age and sex distributions (Figure 4A).

A comparison between the LS and non-LS controls showed 24 species with differential abundances (Figure 4B; Table S6). Of the 22 species, three (*Bacteroides helcogenes* [$p = 0.0041$), *Bacteroides plebeius* [$p = 0.022$], and *Cronobacter turicensis* [$p = 0.0479$]) were enriched in LS controls. In contrast to the comparison within LS subjects, significant enrichment of *Fusobacterium* spp. and other CRC-related gut microbes in LS controls was not observed (Figure S4C).

A comparison of fecal metabolite concentrations showed that 57 metabolites had significantly different concentrations between sporadic and LS controls, 52 of which were higher in the LS controls (Figure 4C; Table S6). Metabolites with higher concentrations in LS controls included several known oncometabolites, such as asymmetric dimethylarginine (ADMA) ($p = 0.00200$) and spermidine ($p = 0.00183$).[39–42] Differential abundance analysis of the KO profile showed 100 KOs higher in LS controls and 213 KOs lower in LS controls than in non-LS controls (Figure 4D; Table S6).

ORA on fecal metabolites and KOs showed 25 KEGG MODULES were significantly over-represented, including "Arginine biosynthesis, glutamate => arginine" (M00845) ($q = 1.26 \times 10^{-4}$), "Arginine biosynthesis, ornithine => arginine" (M00844) ($q = 1.82 \times 10^{-4}$), "Ornithine biosynthesis, glutamate => ornithine" (M00028) ($q = 0.0137$), and "Polyamine biosynthesis, arginine => spermidine" (M00133) ($q = 0.0563$) (Figure 4E). Three modules in EnteroPathway were significantly overrepresented ("Arginine biosynthesis, ornithine => arginine" (EPM1197) ($q = 0.0312$), "Conjugated CA deconjugation" (EPM0129) ($q = 0.0312$), and "GNB biosynthesis" (EPM0819) ($q = 0.0534$)) (Figure 4F). However, the ORA results were mostly driven by shifts in fecal metabolite concentrations with minimal changes in functional gene abundance (Figures 4G, S4C, and S4D).

Overall, a comparison between LS and non-LS controls showed minimal changes in gut microbiota composition and function. However, distinct fecal metabolite profiles and gut microbiota activities were observed, indicating that a distinct LS-specific intestinal environment may be present prior to adenoma formation.

Integrated multi-omics analysis showed potential links between gut microbiota and amino acid metabolism in Lynch syndrome-colorectal cancer subjects

To identify potential association between gut microbiota dysbiosis, fecal metabolite and LS-CRC, we performed correlation analysis between taxonomic profile, functional gene profile and fecal metabolite profiles.

We observed several positive and negative correlations between microbe-metabolite (Table S7). For example, *Fusobacterium ulcerans* and lysine showed negative correlation, while *Peptostreptococcus anaerobius* showed a positive correlation. Positive correlation was also

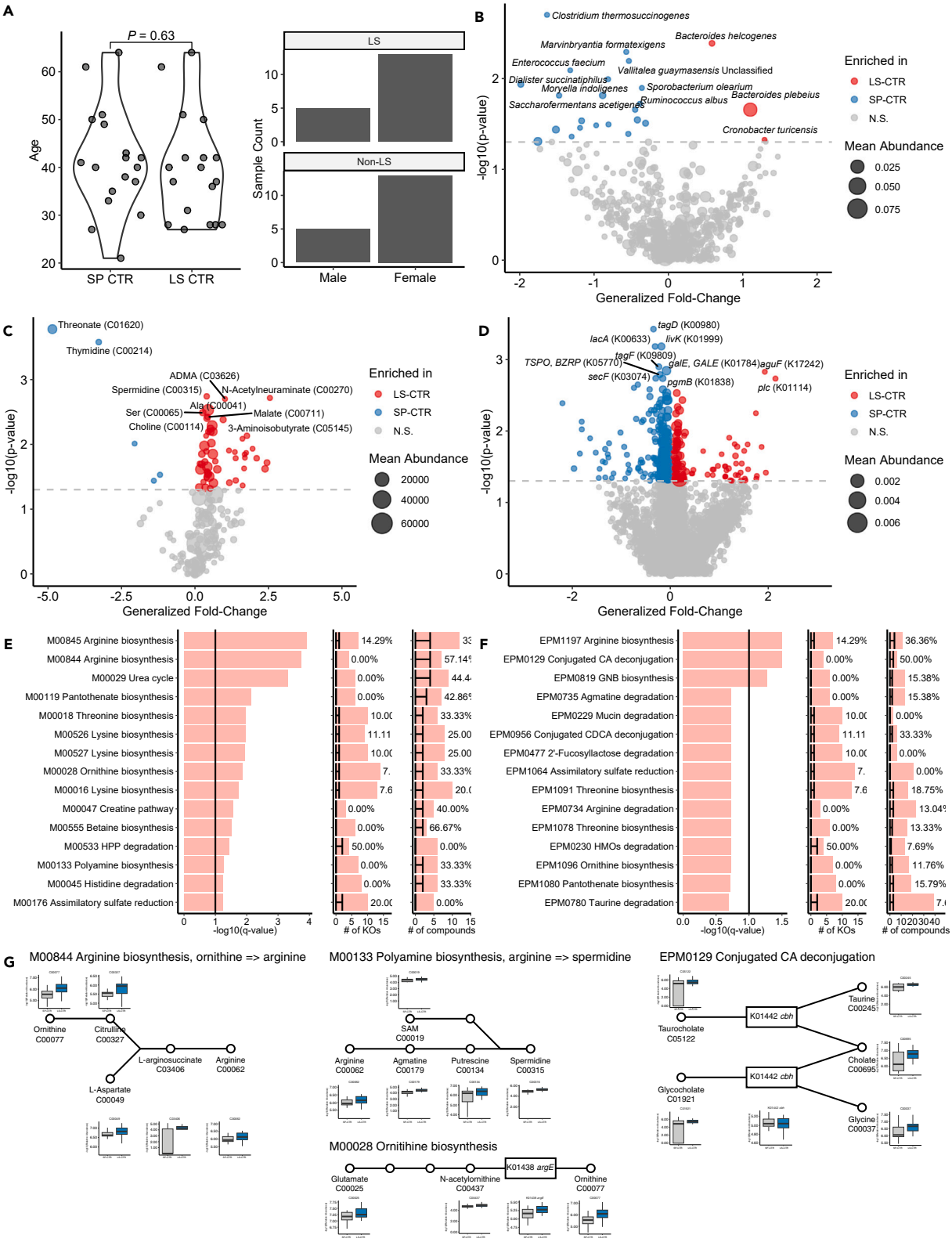

A
P = 0.63
Age
SP CTR
LS CTR
LS
Non-LS
Sample Count
Male
Female
B
Clostridium thermosuccinogenes
Bacteroides helcogenes
Marvinbryantia formatexigens
Enterococcus faecium
Vallitalea guaymasensis Unclassified
Dialister succinatiphilus
Sporobacterium olearium
Moryella indolicgenes
Ruminococcus albus
Bacteroides plebeius
Saccharofermentans acetigenes
Cronobacter turicensis
-log10(p-value)
Generalized Fold-Change
Enriched in
LS-CTR
SP-CTR
N.S.
Mean Abundance
0.025
0.050
0.075
C
Threonate (C01620)
Thymidine (C00214)
ADMA (C03626)
Spermidine (C00315)
N-Acetylneuraminate (C00270)
Ala (C00041)
Ser (C00065)
Malate (C00711)
Choline (C00114)
3-Aminoisobutyrate (C05145)
-log10(p-value)
Generalized Fold-Change
Enriched in
LS-CTR
SP-CTR
N.S.
Mean Abundance
20000
40000
60000
D
tagD (K00980)
lacA (K00633)
livK (K01999)
tagF (K09809)
TSPO, BZRP (K05770)
galE, GALE (K01784)
aguF (K17242)
secF (K03074)
pgmB (K01838)
plc (K01114)
-log10(p-value)
Generalized Fold-Change
Enriched in
LS-CTR
SP-CTR
N.S.
Mean Abundance
0.002
0.004
0.006
E
M00845 Arginine biosynthesis
M00844 Arginine biosynthesis
M00029 Urea cycle
M00119 Pantothenate biosynthesis
M00018 Threonine biosynthesis
M00526 Lysine biosynthesis
M00527 Lysine biosynthesis
M00028 Ornithine biosynthesis
M00016 Lysine biosynthesis
M00047 Creatine pathway
M00555 Betaine biosynthesis
M00533 HPP degradation
M00133 Polyamine biosynthesis
M00045 Histidine degradation
M00176 Assimilatory sulfate reduction
-log10(q-value)
# of KOs
# of compounds
F
EPM1197 Arginine biosynthesis
EPM0129 Conjugated CA deconjugation
EPM0819 GNB biosynthesis
EPM0735 Agmatine degradation
EPM0229 Mucin degradation
EPM0956 Conjugated CDCA deconjugation
EPM0477 2'-Fucosyllactose degradation
EPM1064 Assimilatory sulfate reduction
EPM1091 Threonine biosynthesis
EPM0734 Arginine degradation
EPM1078 Threonine biosynthesis
EPM0230 HMOs degradation
EPM1096 Ornithine biosynthesis
EPM1080 Pantothenate biosynthesis
EPM0780 Taurine degradation
-log10(q-value)
# of KOs
# of compounds
G
M00844 Arginine biosynthesis, ornithine => arginine
Ornithine C00077
Citrulline C00327
L-arginosuccinate C03406
Arginine C00062
L-Aspartate C00049
M00133 Polyamine biosynthesis, arginine => spermidine
SAM C00019
Arginine C00062
Agmatine C00179
Putrescine C00134
Spermidine C00315
M00028 Ornithine biosynthesis
Glutamate C00025
N-acetylornithine C00437
K01438 argE
Ornithine C00077
EPM0129 Conjugated CA deconjugation
Taurocholate C05122
K01442 cbh
Taurine C00245
Glycocholate C01921
K01442 cbh
Cholate C00695
Glycine C00037

Figure 4. Distinct fecal metabolites in LS controls, but minimal changes in related genes and species abundance

(A) Age and sex distribution of LS controls and matched non-LS controls. $p < 0.05$ (Mann-Whitney U test) is considered statistically significant.

(B) Comparison of species relative abundances between LS controls (n = 18) and matched non-LS controls (n = 18). The x axis represents the generalized fold-change between LS and non-LS controls, and the y axis represents the negative log10 transformed *p*-values of the two-sided Mann–Whitney *U* test. The horizontal dotted line shows the -log10 transformed *p*-value of 0.05. The size of the dots indicates relative abundance.

(C) Comparison of fecal metabolite concentrations between LS controls (n = 18) and matched non-LS controls (n = 18). The x axis represents the generalized fold-change between LS and non-LS controls, and the y axis represents the negative log10 transformed *p*-values of the two-sided Mann–Whitney *U* test. The horizontal dot line shows the -log10 transformed *p*-value of 0.05. The size of the dots indicates relative abundance.

(D) Comparison of KO relative abundances between LS controls (n = 18) and matched non-LS controls (n = 18). The x axis represents the generalized fold-change between LS and non-LS controls, and the y axis represents the negative log10 transformed *p*-values of the two-sided Mann–Whitney *U* test. The horizontal dotted line shows the -log10 transformed *p*-value of 0.05. The size of the dots indicates relative abundance.

(E) Top 15 KEGG MODULEs overrepresented in LS controls (*p*-value <0.05). Modules are arranged in ascending order based on *p*-value. The right figures represent the number of KOs and metabolites in each module (red bar) and the percentage of enriched KOs and metabolites within the module (black line).

(F) Top 15 EnteroPathway MODULEs overrepresented in LS controls (*p*-value <0.05). Three EnteroPathway MODULEs showed significant overrepresentation in LS controls (FDR *p*-value <0.1). Modules are arranged in ascending order based on *p*-value. The right figures represent the number of KOs and metabolites in each module (red bar) and the percentage of enriched KOs and metabolites within the module (black line).

(G) Diagrams of overrepresented metabolic modules in LS controls compared to non-LS controls. Circular nodes represent key or statistically significant metabolites. Rectangular nodes represent key or statistically significant KOs. Statistical significance threshold was defined as Mann–Whitney U $p < 0.05$.

observed between other CRC-associated microbe and arginine, although no statistical significance was detected. Exploring gene-metabolite correlation, we observed a negative correlation between *gabD* and C00431 (5-aminovalerate), as well as C00042 (succinate), while *gabD* showed a positive correlation with C00025 (Glu) (Table S7). Several other metabolites were correlated with *gabD* abundance, indicating its promiscuous nature, and involved in several metabolism pathways. Although not significant, *arcA* showed a positive correlation with arginine. In microbe-gene correlation, we observed a positive correlation between CRC-associated microbes and *arcA* (arginine deiminase, K01478), indicating a potential link between amino acid metabolism and CRC-associated microbes (Table S7).

DISCUSSION

In this study, we profiled the fecal microbiome and metabolome of 71 patients with LS. Our LS cohort included patients without adenoma or carcinoma and no prior history of adenoma or carcinoma formation, those with colorectal adenoma, those with CRC, and those with a history of colectomy due to past carcinoma formation. Furthermore, we obtained data from patients with non-LS CRC using identical sampling protocol.[26] Our study is the first to analyze the changes in gut microbiota and fecal metabolite across all stages of adenoma-carcinoma progression in LS, along with a comparison with non-LS CRC.

In line with previous studies, LS subjects had lower alpha diversity than sporadic CRC and non-LS controls.[21–23] We did not detect significant associations between alpha diversity and colonoscopy findings, age, or MMR mutations. Despite a significant association between LS status and alpha diversity, beta-diversity analysis did not detect a significant association between LS status and gut microbiota composition. Subject's CRC progression was identified as the most explanatory variable for the LS cohort's gut microbiota composition. Although previous studies have reported LS-specific gut microbiota composition, recent studies showed no distinct gut microbiota between non-LS non-tumor control and non-tumor LS carriers, in consensus with our observation.

Both LS and non-LS subjects were clustered into enterotypes based on their genus profiles to identify potential LS-specific enterotype.[43,44] Based on enterotype clustering and dominant genus analysis of each enterotypes, LS subjects tended to cluster into enterotypes with high *Bacteroides* abundance and low *Faecalibacterium* abundance (*Faecalibaterium*-poor enterotype). Enterotype proportions had a minimal association with adenoma/carcinoma formation or colectomy. As low alpha diversity and *Faecalibacterium*-poor enterotypes have been associated with IBD, our findings indicate the intestinal inflammation of LS subjects, regardless of CRC progression.[27,28] Activated immune profiles due to MMR loss-of-function have been reported in LS subjects across all stages of CRC progressions,[45–48] which may induce metabolic reprogramming of the intestinal environment, affecting gut microbiota composition.[49]

One of the most important findings of this study was the discovery of a distinct association between *Fusobacterium* species and LS-CRC. Although several species, such as *F. nucleatum*, *P. stomatis*, *P. micra*, and *G. morbillorum*, have been associated with sporadic CRC,[26,29,30] only *Fusobacterium* species showed significant enrichment in patients with LS-CRC. ML-based marker screening validated that enriched *F. nucleatum* acts as CRC marker both in sporadic and LS CRC. In line with a previous study, *F. nucleatum* enrichment was not observed in patients with LS and adenoma was not found,[20] suggesting carcinoma formation preludes *F. nucleatum* enrichment in LS-CRC. Although not significant, we validated *Fusobacterium* enrichment in LS subjects with cancer history from an independent Israel LS cohort.[24] In consensus with our findings, high intra-tumoral *F. nucleatum* abundance have been reported in dMMR/MSI-H CRC, including those of hereditary nature.[18,50–52] To our knowledge, this study is the first to detect enriched *Fusobacterium* in LS-CRC fecal sample.

F. nucleatum has been associated with immune modulation, cancer metastasis, chemoresistance, and immune checkpoint inhibitor (ICI) response.[53–57] Chemoresistance and ICI response have been associated with dMMR/MSI-H CRC, indicating *F. nucleatum* may play some roles in CRC treatment. Furthermore, *F. nucleatum* has also been associated with aspirin roles in CRC prevention, in which aspirin was shown to have an antibiotic effect against specific *F. nucleatum* strains. Unfortunately, we did not have aspirin intake data in our cohort to study this phenomenon.

Consistent with the enrichment of *F. nucleatum* in LS CRC, we found *fap2*, a known *F. nucleatum* virulence factor (Fn VF), was significantly more abundant in LS CRC. Previous studies have reported the potential role of *fap2* in host–microbe interactions and immunomodulatory properties,[33,38] which may offer an advantage for *F. nucleatum* growth in the CRC tumor microenvironment.

Notably, our LS cohort had a low detection rate for *fadA* and its homologs, which have been shown to cause DNA damage and CRC cell growth.[32,58] Knockout of *fadA* minimizes *F. nucleatum* effect on CRC progression.[58] Low *fadA* levels may be associated with a better disease prognosis in patients with LS-CRC, in line with reports of a better prognosis in LS-CRC compared to sporadic CRC.[2] However, as our LS cohort has a limited sample size, the low *fadA* detection rate might be due to a lack of detection power. Thus, validation in a larger LS cohort is essential to study the exact roles of *F. nucleatum* in LS CRC and identify key virulence factors.

Differential abundance analysis of fecal metabolites and gene profiles revealed putative functional shifts in LS CRC gut microbes compared to LS controls, such as the overrepresentation of the lysine degradation pathway, GABA shunts, and polyamine biosynthesis. GABA shunt and lysine degradation were linked with glutamine metabolism, where lysine acts as a precursor to glutamine, which is further metabolized into succinate in the GABA shunt.[59,60] Serum concentration of both glutamine and lysine were reported to be elevated in CRC subjects.[61] Polyamines, such as putrescine and spermidine, play many roles in host and microbial biology, including immune system modulation and bacterial response to environmental changes, such as increased oxidative damage.[62,63] Higher amino acid catabolism and increased polyamine metabolism by gut microbes may be adaptive responses to CRC-driven changes in the intestinal environment.[64] Notably, *Fusobacterium nucleatum* has been reported to utilize amino acids as an energy source, as well as utilizing polyamine metabolism for its biofilm formation in the oral microenvironment.[65,66] Co-culture of *F. nucleatum* and human cell line have reported amino acid degradation activity by *F. nucleatum*, with succinate and SCFA as the end product, suggesting cross-feeding activity between *F. nucleatum* and CRC cell.[67]

Another important finding of this study was the discovery of distinct fecal metabolite compositions in patients with LS without cancerous lesions, with minimal gut microbiota composition and functional changes. Amino acids, polyamines, and bile acids have been reported to modulate the immune system, suggesting fecal metabolites disruption association with immune system activation in LS-control.[62,68,69] Other than immune modulation, bile acids have been associated with increased CRC risk, especially DCA-induced DNA oxidative damage.[70] As DNA oxidative damage is one of the repair targets of mismatch repair genes,[71] the accumulation of bile acids might significantly impact CRC tumorigenesis in patients with LS. Distinct immune profiles have been reported in patients with LS with dMMR colonic crypts, even prior to adenoma and carcinoma formation.[45] As activated immune systems have been reported to induce metabolic reprogramming,[49] the observed shifts in fecal metabolites might be a marker of the activated host immune system triggered by early dMMR mutations. In consensus with our findings, a previous study has shown serum metabolites of non-tumor LS subjects were similar to sporadic CRC, with elevated inflammatory markers.[72]

To search for potential links between omics, we explored microbe-gene, microbe-metabolite, and gene-metabolite correlation. From the correlation analysis results, we identify potential links between CRC-associated microbe and amino acids metabolism, such as lysine and arginine degradation. Arginine degradation is linked with polyamine biosynthesis, while lysine degradation is linked to GABA shunts and TCA cycle, both pathways that were affected by metabolic reprogramming, associated with immune system activation and carcinoma formation. Thus, from multi-omics integration, we identified amino acid metabolism as a potential link between gut microbiota and host intestinal environment changes.

Overall, based on the findings in this study, we hypothesized that compared with sporadic CRC, host genetics, e.g., MMR mutation and loss-of-function, likely acts as the main driver of CRC tumorigenesis in LS.[1,2] Early MMR loss-of-function may elevate mutation accumulation rate, such as frameshift peptides, which acts as novel antigen, inducing immune system activation.[73] Activated immune system might induce shifts in gut microbiota composition and function, selecting for gut microbes with innate immune evasion and amino acid utilization capabilities, such as *Fusobacterium nucleatum*.[38,65–67] Furthermore, gut microbiota dysbiosis, such as *F. nucleatum* enrichment, may play a role in CRC progression, metastasis, disease prognosis, and treatment response.[19,32,36,54,55,57,58,67]

Limitations of the study

We acknowledge several limitations of our study. Due to the cross-sectional nature of our study, all results were observational, thus requiring validation experiments and additional cohorts. Recent study has shown the potential effect of non-antibiotic drugs against gut microbiota. Combined with the recent recommendation of daily aspirin intake for CRC prevention in the LS population, future studies should consider the potential interaction between aspirin and gut microbiota. Thus, future longitudinal or prospective studies with dietary and drug intake, intestinal inflammation marker, dMMR loss-of-function validation, paired serum-fecal metabolite, and immune profiles may provide further evidence to support our hypothesis. Additionally, a co-culture or mouse model experiment is essential to elucidate the causal relations of microbe-CRC interaction in LS subjects.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110181.

ACKNOWLEDGMENTS

We thank all patients and their families who participated in this study, N. Sezawa, R. Ohira, H. Sekiguchi, and R. Itsuki (National Cancer Center Hospital, Tokyo, Japan), for expert technical assistance. This work was supported by grants from the National Cancer Center Research and Development Fund (2020-A-4 to Y.S. and S.Y. and 2023-A-6 to Y.S.); Practical Research for Innovative Cancer Control from the Japan Agency for Medical Research and Development (AMED) (JP22ck0106546 to Y.S., T. Yamada, and S.Y.; JP23ck0106799h to Y.S., T. Yamada, and S.Y.); Project for Cancer Research and Therapeutic Evolution (P-CREATE) from AMED (JP21cm0106477 to S.S., T. Yamada, and S.Y.); Project for Promotion of Cancer Research and Therapeutic Evolution (P-PROMOTE) from AMED (JP23ama221404 to Y.S., T. Yamada, and S.Y.; JP24ama221430 to Y.S., T. Yamada, and S.Y.); United States-Japan Cooperative Medical Science Program from AMED (JP23jk0210009 to S.Y.); AMED-CREST (JP23gm1010009 to S.F.); AMED Moonshot (JP21zf012700 to S.F. and T. Yamada); JSPS KAKENHI (22H03541 to S.F.; 20H03662 to S.Y.; 23H02892 to S.Y.; 18K15800 to S.S.; 22K16336 to S.S.); JST ERATO (JPMJER1902 to S.F.); JST-AIP Acceleration Research (JPMKCR19U3 to T. Yamada and S.Y.); Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary Research Initiatives, Osaka University (to S.Y.); Joint Research Project of the Institute Medical Science, the University of Tokyo (to S.Y.); the Takeda Science Foundation (to S.Y.); the Yasuda Medical Foundation (to S.Y.); the Mitsubishi Foundation (to S.Y.); the Princess Takamatsu Cancer Research Fund (to S.Y.); Yakult Bio-Science Foundation (to S.Y.); the Food Science Institute Foundation (to S.F.).

AUTHOR CONTRIBUTIONS

S.Y., T.N., Y.S., S.F., and T. Yamada. contributed to the study concept and design. S.Y., T.N., S.S., H.T., M.Y., T. Yachida, and Y.S. collected the clinical samples and information. S.Y., T.S., and S.F. performed metagenome and metabolome experiments. F.S. and S.M. performed bioinformatics analyses. F.S., S.M., S.Y., and T. Yamada wrote the article.

DECLARATION OF INTERESTS

Takuji Yamada (T.Y.) and Shinji Fukuda (S.F.) are founders of Metagen Inc. and Metagen Therapeutics Inc. Metagen Inc. focuses on the design and control of the gut environment for human health. Metagen Therapeutics Inc. focuses on drug discovery and development, which utilizes microbiome science. T.Y. is the founder of digzyme Inc. digzyme Inc. focuses on the development of novel enzymes. None of the companies had control over the interpretation, writing, or publication of this work.

Received: December 7, 2023
Revised: March 11, 2024
Accepted: June 1, 2024
Published: June 4, 2024

REFERENCES

1. Fearon, E.R. (2011). Molecular Genetics of Colorectal Cancer. Annu. Rev. Pathol. *6*, 479–507. https://doi.org/10.1146/annurev-pathol-011110-130235.
2. Lynch, H.T., Snyder, C.L., Shaw, T.G., Heinen, C.D., and Hitchins, M.P. (2015). Milestones of Lynch syndrome: 1895–2015. Nat. Rev. Cancer *15*, 181–194. https://doi.org/10.1038/nrc3878.
3. Bonadona, V., Bonaïti, B., Olschwang, S., Grandjouan, S., Huiart, L., Longy, M., Guimbaud, R., Buecher, B., Bignon, Y.J., Caron, O., et al. (2011). Cancer Risks Associated With Germline Mutations in MLH1, MSH2, and MSH6 Genes in Lynch Syndrome. JAMA *305*, 2304–2310. https://doi.org/10.1001/jama.2011.743.
4. Ten Broeke, S.W., van der Klift, H.M., Tops, C.M.J., Aretz, S., Bernstein, I., Buchanan, D.D., de la Chapelle, A., Capella, G., Clendenning, M., Engel, C., et al. (2018). Cancer Risks for PMS2-Associated Lynch Syndrome. J. Clin. Oncol. *36*, 2961–2968. https://doi.org/10.1200/JCO.2018.78.4777.
5. Lee, B.C.H., Robinson, P.S., Coorens, T.H.H., Yan, H.H.N., Olafsson, S., Lee-Six, H., Sanders, M.A., Siu, H.C., Hewinson, J., Yue, S.S.K., et al. (2022). Mutational landscape of normal epithelial cells in Lynch Syndrome patients. Nat. Commun. *13*, 2710. https://doi.org/10.1038/s41467-022-29920-2.
6. Battaglin, F., Naseem, M., Lenz, H.-J., and Salem, M.E. (2018). Microsatellite Instability in Colorectal Cancer: Overview of Its Clinical Significance and Novel Perspectives. Clin. Adv. Hematol. Oncol. *16*, 735–745.
7. Zeller, G., Tap, J., Voigt, A.Y., Sunagawa, S., Kultima, J.R., Costea, P.I., Amiot, A., Böhm, J., Brunetti, F., Habermann, N., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol. Syst. Biol. *10*, 766. https://doi.org/10.15252/msb.20145645.
8. Kostic, A.D., Gevers, D., Pedamallu, C.S., Michaud, M., Duke, F., Earl, A.M., Ojesina, A.I., Jung, J., Bass, A.J., Tabernero, J., et al. (2012). Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res. *22*, 292–298. https://doi.org/10.1101/gr.126573.111.
9. Castellarin, M., Warren, R.L., Freeman, J.D., Dreolini, L., Krzywinski, M., Strauss, J., Barnes, R., Watson, P., Allen-Vercoe, E., Moore, R.A., and Holt, R.A. (2012). Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res. *22*, 299–306. https://doi.org/10.1101/gr.126516.111.
10. Yu, J., Feng, Q., Wong, S.H., Zhang, D., Liang, Q.Y., Qin, Y., Tang, L., Zhao, H., Stenvang, J., Li, Y., et al. (2017). Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut *66*, 70–78. https://doi.org/10.1136/gutjnl-2015-309800.
11. Li, Q., Hu, W., Liu, W.-X., Zhao, L.-Y., Huang, D., Liu, X.-D., Chan, H., Zhang, Y., Zeng, J.-D., Coker, O.O., et al. (2021). Streptococcus thermophilus Inhibits Colorectal Tumorigenesis Through Secreting β-Galactosidase. Gastroenterology *160*, 1179–1193.e14. https://doi.org/10.1053/j.gastro.2020.09.003.
12. Long, X., Wong, C.C., Tong, L., Chu, E.S.H., Ho Szeto, C., Go, M.Y.Y., Coker, O.O., Chan, A.W.H., Chan, F.K.L., Sung, J.J.Y., and Yu, J. (2019). Peptostreptococcus anaerobius promotes colorectal carcinogenesis and modulates tumour immunity. Nat. Microbiol. *4*, 2319–2330. https://doi.org/10.1038/s41564-019-0541-3.
13. Dalmasso, G., Cougnoux, A., Delmas, J., Darfeuille-Michaud, A., and Bonnet, R. (2014). The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. Gut Microb. *5*, 675–680. https://doi.org/10.4161/19490976.2014.969989.
14. Dougherty, M.W., Valdés-Mas, R., Wernke, K.M., Gharaibeh, R.Z., Yang, Y., Brant, J.O., Riva, A., Muehlbauer, M., Elinav, E., Puschhof, J., et al. (2023). The microbial genotoxin colibactin exacerbates mismatch repair mutations in colorectal tumors. Neoplasia *43*, 100918. https://doi.org/10.1016/j.neo.2023.100918.
15. Belcheva, A., Irrazabal, T., Robertson, S.J., Streutker, C., Maughan, H., Rubino, S., Moriyama, E.H., Copeland, J.K., Surendra, A., Kumar, S., et al. (2014). Gut Microbial Metabolism Drives Transformation of Msh2-Deficient Colon Epithelial Cells. Cell *158*, 288–299. https://doi.org/10.1016/j.cell.2014.04.051.
16. Pieters, W., Hugenholtz, F., Kos, K., Cammeraat, M., Moliej, T.C., Kaldenbach, D., Klarenbeek, S., Davids, M., Drost, L., de Konink, C., et al. (2022). Pro-mutagenic effects of the gut microbiota in a Lynch syndrome mouse model. Gut Microb. *14*, 2035660. https://doi.org/10.1080/19490976.2022.2035660.
17. Hale, V.L., Jeraldo, P., Chen, J., Mundy, M., Yao, J., Priya, S., Keeney, G., Lyke, K., Ridlon, J., White, B.A., et al. (2018). Distinct microbes, metabolites, and ecologies define the microbiome in deficient and proficient mismatch repair colorectal cancers. Genome Med. *10*, 78. https://doi.org/10.1186/s13073-018-0586-6.
18. Ono, T., Yamaguchi, T., Takao, M., Kojika, E., Iijima, T., and Horiguchi, S.I. (2022). Fusobacterium nucleatum load in MSI colorectal cancer subtypes. Int. J. Clin. Oncol. *27*, 1580–1588. https://doi.org/10.1007/s10147-022-02218-5.
19. Mima, K., Nishihara, R., Qian, Z.R., Cao, Y., Sukawa, Y., Nowak, J.A., Yang, J., Dou, R., Masugi, Y., Song, M., et al. (2016). Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. Gut *65*, 1973–1980. https://doi.org/10.1136/gutjnl-2015-310101.
20. Yan, Y., Drew, D.A., Markowitz, A., Lloyd-Price, J., Abu-Ali, G., Nguyen, L.H., Tran, C., Chung, D.C., Gilpin, K.K., Meixell, D., et al. (2020). Structure of the Mucosal and Stool Microbiome in Lynch Syndrome. Cell Host Microbe *27*, 585–600.e4. https://doi.org/10.1016/j.chom.2020.03.005.
21. Mori, G., Orena, B.S., Cultrera, I., Barbieri, G., Albertini, A.M., Ranzani, G.N., Carnevali, I., Tibiletti, M.G., and Pasca, M.R. (2019). Gut Microbiota Analysis in Postoperative Lynch Syndrome Patients. Front. Microbiol. *10*, 1746. https://doi.org/10.3389/fmicb.2019.01746.
22. Ferrarese, R., Zuppardo, R.A., Puzzono, M., Mannucci, A., Amato, V., Ditonno, I., Patricelli, M.G., Raucci, A.R., Clementi, M., Elmore, U., et al. (2020). Oral and Fecal Microbiota in Lynch Syndrome. J. Clin. Med. *9*, 2735. https://doi.org/10.3390/jcm9092735.
23. Rifkin, S.B., Sze, M.A., Tuck, K., Koeppe, E., Stoffel, E.M., and Schloss, P.D. (2024). Gut Microbiome Composition in Lynch Syndrome With and Without History of Colorectal Neoplasia and Non-Lynch Controls. J. Gastrointest. Cancer *55*, 207–218. https://doi.org/10.1007/s12029-023-00925-4.
24. Naddaf, R., Carasso, S., Reznick-Levi, G., Hasnis, E., Qarawani, A., Maza, I., Gefen, T., Half, E.E., and Geva-Zatorsky, N. (2023). Gut microbial signatures are associated with Lynch syndrome (LS) and cancer history in Druze communities in Israel. Sci. Rep. *13*, 20677. https://doi.org/10.1038/s41598-023-47723-3.
25. Tsugane, S., and Sawada, N. (2014). The JPHC Study: Design and Some Findings on the Typical Japanese Diet. Jpn. J. Clin. Oncol. *44*, 777–782. https://doi.org/10.1093/jjco/hyu096.
26. Yachida, S., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Watanabe, H., Masuda, K., Nishimoto, Y., Kubo, M., et al. (2019). Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nat. Med. *25*, 968–976. https://doi.org/10.1038/s41591-019-0458-7.
27. Vieira-Silva, S., Sabino, J., Valles-Colomer, M., Falony, G., Kathagen, G., Caenepeel, C., Cleynen, I., van der Merwe, S., Vermeire, S., and Raes, J. (2019). Quantitative microbiome profiling disentangles inflammation- and bile duct obstruction-associated microbiota alterations across PSC/IBD diagnoses. Nat. Microbiol. *4*, 1826–1831. https://doi.org/10.1038/s41564-019-0483-9.
28. Ning, L., Zhou, Y.-L., Sun, H., Zhang, Y., Shen, C., Wang, Z., Xuan, B., Zhao, Y., Ma, Y., Yan, Y., et al. (2023). Microbiome and metabolome features in inflammatory bowel disease via multi-omics integration analyses across cohorts. Nat. Commun. *14*, 7135. https://doi.org/10.1038/s41467-023-42788-0.
29. Wirbel, J., Pyl, P.T., Kartal, E., Zych, K., Kashani, A., Milanese, A., Fleck, J.S., Voigt, A.Y., Palleja, A., Ponnudurai, R., et al. (2019). Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat. Med. *25*, 679–689. https://doi.org/10.1038/s41591-019-0406-6.
30. Thomas, A.M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., Beghini, F., Manara, S., Karcher, N., Pozzi, C., et al. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat. Med. *25*, 667–678. https://doi.org/10.1038/s41591-019-0405-7.
31. Rynazal, R., Fujisawa, K., Shiroma, H., Salim, F., Mizutani, S., Shiba, S., Yachida, S., and Yamada, T. (2023). Leveraging explainable AI for gut microbiome-based colorectal cancer classification. Genome Biol. *24*, 21. https://doi.org/10.1186/s13059-023-02858-4.
32. Rubinstein, M.R., Wang, X., Liu, W., Hao, Y., Cai, G., and Han, Y.W. (2013). Fusobacterium nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via its FadA Adhesin. Cell Host Microbe *14*, 195–206. https://doi.org/10.1016/j.chom.2013.07.012.
33. Abed, J., Emgård, J.E.M., Zamir, G., Faroja, M., Almogy, G., Grenov, A., Sol, A., Naor, R., Pikarsky, E., Atlan, K.A., et al. (2016). Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by

Binding to Tumor-Expressed Gal-GalNAc. Cell Host Microbe *20*, 215–225. https://doi.org/10.1016/j.chom.2016.07.006.

34. Kaplan, C.W., Ma, X., Paranjpe, A., Jewett, A., Lux, R., Kinder-Haake, S., and Shi, W. (2010). Fusobacterium nucleatum Outer Membrane Proteins Fap2 and RadD Induce Cell Death in Human Lymphocytes. Infect. Immun. *78*, 4773–4778. https://doi.org/10.1128/IAI.00567-10.
35. Galaski, J., Shhadeh, A., Umaña, A., Yoo, C.C., Arpinati, L., Isaacson, B., Berhani, O., Singer, B.B., Slade, D.J., Bachrach, G., et al. (2021). Fusobacterium nucleatum CbpF Mediates Inhibition of T Cell Function Through CEACAM1 Activation. Front. Cell. Infect. Microbiol. *11*, 692544.
36. Wu, Y., Guo, S., Chen, F., Li, Y., Huang, Y., Liu, W., and Zhang, G. (2023). Fn-Dps, a novel virulence factor of Fusobacterium nucleatum, disrupts erythrocytes and promotes metastasis in colorectal cancer. PLoS Pathog. *19*, e1011096. https://doi.org/10.1371/journal.ppat.1011096.
37. Umaña, A., Sanders, B.E., Yoo, C.C., Casasanta, M.A., Udayasuryan, B., Verbridge, S.S., and Slade, D.J. (2019). Utilizing Whole Fusobacterium Genomes To Identify, Correct, and Characterize Potential Virulence Protein Families. J. Bacteriol. *201*, 10–1128. https://doi.org/10.1128/JB.00273-19.
38. Gur, C., Ibrahim, Y., Isaacson, B., Yamin, R., Abed, J., Gamliel, M., Enk, J., Bar-On, Y., Stanietsky-Kaynan, N., Coppenhagen-Glazer, S., et al. (2015). Binding of the Fap2 Protein of Fusobacterium nucleatum to Human Inhibitory Receptor TIGIT Protects Tumors from Immune Cell Attack. Immunity *42*, 344–355. https://doi.org/10.1016/j.immuni.2015.01.010.
39. Li, H., Zhou, Y., Zhao, A., Qiu, Y., Xie, G., Jiang, Q., Zheng, X., Zhong, W., Sun, X., Zhou, Z., and Jia, W. (2013). Asymmetric dimethylarginine attenuates serum starvation-induced apoptosis via suppression of the Fas (APO-1/CD95)/JNK (SAPK) pathway. Cell Death Dis. *4*, e830. https://doi.org/10.1038/cddis.2013.345.
40. Gobert, A.P., Latour, Y.L., Asim, M., Barry, D.P., Allaman, M.M., Finley, J.L., Smith, T.M., McNamara, K.M., Singh, K., Sierra, J.C., et al. (2022). Protective Role of Spermidine in Colitis and Colon Carcinogenesis. Gastroenterology *162*, 813–827.e8. https://doi.org/10.1053/j.gastro.2021.11.005.
41. Ding, N., Cheng, Y., Liu, H., Wu, Y., Weng, Y., Cui, H., Cheng, C., Zhang, W., and Cui, Y. (2023). Fusobacterium nucleatum Infection Induces Malignant Proliferation of Esophageal Squamous Cell Carcinoma Cell by Putrescine Production. Microbiol. Spectr. *11*, e0275922. https://doi.org/10.1128/spectrum.02759-22.
42. Kuo, C.-C., Wu, J.-Y., and Wu, K.K. (2022). Cancer-derived extracellular succinate: a driver of cancer metastasis. J. Biomed. Sci. *29*, 93. https://doi.org/10.1186/s12929-022-00878-z.
43. Costea, P.I., Hildebrand, F., Arumugam, M., Bäckhed, F., Blaser, M.J., Bushman, F.D., de Vos, W.M., Ehrlich, S.D., Fraser, C.M., Hattori, M., et al. (2018). Enterotypes in the landscape of gut microbial community composition. Nat. Microbiol. *3*, 8–16. https://doi.org/10.1038/s41564-017-0072-8.
44. Holmes, I., Harris, K., and Quince, C. (2012). Dirichlet Multinomial Mixtures: Generative Models for Microbial Metagenomics. PLoS One *7*, e30126. https://doi.org/10.1371/journal.pone.0030126.
45. Bohaumilitzky, L., Kluck, K., Hüneburg, R., Gallon, R., Nattermann, J., Kirchner, M., Kristiansen, G., Hommerding, O., Pfuderer, P.L., Wagner, L., et al. (2022). The Different Immune Profiles of Normal Colonic Mucosa in Cancer-Free Lynch Syndrome Carriers and Lynch Syndrome Colorectal Cancer Patients. Gastroenterology *162*, 907–919.e10. https://doi.org/10.1053/j.gastro.2021.11.029.
46. Chang, K., Taggart, M.W., Reyes-Uribe, L., Borras, E., Riquelme, E., Barnett, R.M., Leoni, G., San Lucas, F.A., Catanese, M.T., Mori, F., et al. (2018). Immune Profiling of Premalignant Lesions in Patients With Lynch Syndrome. JAMA Oncol. *4*, 1085–1092. https://doi.org/10.1001/jamaoncol.2018.1482.
47. Schwitalle, Y., Kloor, M., Eiermann, S., Linnebacher, M., Kienle, P., Knaebel, H.P., Tariverdian, M., Benner, A., and von Knebel Doeberitz, M. (2008). Immune Response Against Frameshift-Induced Neopeptides in HNPCC Patients and Healthy HNPCC Mutation Carriers. Gastroenterology *134*, 988–997. https://doi.org/10.1053/j.gastro.2008.01.015.
48. Walkowska, J., Kallemose, T., Jönsson, G., Jönsson, M., Andersen, O., Andersen, M.H., Svane, I.M., Langkilde, A., Nilbert, M., and Therkildsen, C. (2019). Immunoprofiles of colorectal cancer from Lynch syndrome. OncoImmunology *8*, e1515612. https://doi.org/10.1080/2162402X.2018.1515612.
49. Xia, L., Oyang, L., Lin, J., Tan, S., Han, Y., Wu, N., Yi, P., Tang, L., Pan, Q., Rao, S., et al. (2021). The cancer metabolic reprogramming and immune response. Mol. Cancer *20*, 28. https://doi.org/10.1186/s12943-021-01316-8.
50. Mima, K., Sukawa, Y., Nishihara, R., Qian, Z.R., Yamauchi, M., Inamura, K., Kim, S.A., Masuda, A., Nowak, J.A., Nosho, K., et al. (2015). *Fusobacterium nucleatum* and T Cells in Colorectal Carcinoma. JAMA Oncol. *1*, 653–661. https://doi.org/10.1001/jamaoncol.2015.1377.
51. Hamada, T., Zhang, X., Mima, K., Bullman, S., Sukawa, Y., Nowak, J.A., Kosumi, K., Masugi, Y., Twombly, T.S., Cao, Y., et al. (2018). Fusobacterium nucleatum in Colorectal Cancer Relates to Immune Response Differentially by Tumor Microsatellite Instability Status. Cancer Immunol. Res. *6*, 1327–1336. https://doi.org/10.1158/2326-6066.CIR-18-0174.
52. Roelands, J., Kuppen, P.J.K., Ahmed, E.I., Mall, R., Masoodi, T., Singh, P., Monaco, G., Raynaud, C., de Miranda, N.F.C.C., Ferraro, L., et al. (2023). An integrated tumor, immune and microbiome atlas of colon cancer. Nat. Med. *29*, 1273–1286. https://doi.org/10.1038/s41591-023-02324-5.
53. Galeano Niño, J.L., Wu, H., LaCourse, K.D., Kempchinsky, A.G., Baryiames, A., Barber, B., Futran, N., Houlton, J., Sather, C., Sicinska, E., et al. (2022). Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. Nature *611*, 810–817. https://doi.org/10.1038/s41586-022-05435-0.
54. Yu, T., Guo, F., Yu, Y., Sun, T., Ma, D., Han, J., Qian, Y., Kryczek, I., Sun, D., Nagarsheth, N., et al. (2017). Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. Cell *170*, 548–563.e16. https://doi.org/10.1016/j.cell.2017.07.008.
55. Gao, Y., Bi, D., Xie, R., Li, M., Guo, J., Liu, H., Guo, X., Fang, J., Ding, T., Zhu, H., et al. (2021). Fusobacterium nucleatum enhances the efficacy of PD-L1 blockade in colorectal cancer. Signal Transduct. Targeted Ther. *6*, 398. https://doi.org/10.1038/s41392-021-00795-x.
56. Gao, Y., Zou, T., Xu, P., Wang, Y., Jiang, Y., Chen, Y.-X., Chen, H., Hong, J., and Fang, J.-Y. (2023). Fusobacterium nucleatum stimulates cell proliferation and promotes PD-L1 expression via IFIT1-related signal in colorectal cancer. Neoplasia *35*, 100850. https://doi.org/10.1016/j.neo.2022.100850.
57. Jiang, S.-S., Xie, Y.-L., Xiao, X.-Y., Kang, Z.-R., Lin, X.-L., Zhang, L., Li, C.-S., Qian, Y., Xu, P.-P., Leng, X.-X., et al. (2023). Fusobacterium nucleatum-derived succinic acid induces tumor resistance to immunotherapy in colorectal cancer. Cell Host Microbe *31*, 781–797.e9. https://doi.org/10.1016/j.chom.2023.04.010.
58. Guo, P., Tian, Z., Kong, X., Yang, L., Shan, X., Dong, B., Ding, X., Jing, X., Jiang, C., Jiang, N., and Yu, Y. (2020). FadA promotes DNA damage and progression of Fusobacterium nucleatum-induced colorectal cancer through up-regulation of chk2. J. Exp. Clin. Cancer Res. *39*, 202. https://doi.org/10.1186/s13046-020-01677-w.
59. Huang, D., Wang, Y., Thompson, J.W., Yin, T., Alexander, P.B., Qin, D., Mudgal, P., Wu, H., Liang, Y., Tan, L., et al. (2022). Cancer-cell-derived GABA promotes β-catenin-mediated tumour growth and immunosuppression. Nat. Cell Biol. *24*, 230–241. https://doi.org/10.1038/s41556-021-00820-9.
60. Papes, F., Surpili, M.J., Langone, F., Trigo, J.R., and Arruda, P. (2001). The essential amino acid lysine acts as precursor of glutamate in the mammalian central nervous system. FEBS Lett. *488*, 34–38. https://doi.org/10.1016/S0014-5793(00)02401-7.
61. Yang, Y., Wang, Z., Li, X., Lv, J., Zhong, R., Gao, S., Zhang, F., and Chen, W. (2023). Profiling the metabolic disorder and detection of colorectal cancer based on targeted amino acids metabolomics. J. Transl. Med. *21*, 824. https://doi.org/10.1186/s12967-023-04604-7.
62. Proietti, E., Rossini, S., Grohmann, U., and Mondanelli, G. (2020). Polyamines and Kynurenines at the Intersection of Immune Modulation. Trends Immunol. *41*, 1037–1050. https://doi.org/10.1016/j.it.2020.09.007.
63. Murray Stewart, T., Dunston, T.T., Woster, P.M., and Casero, R.A. (2018). Polyamine catabolism and oxidative damage. J. Biol. Chem. *293*, 18736–18745. https://doi.org/10.1074/jbc.TM118.003337.
64. Lamaudière, M.T.F., Arasaradnam, R., Weedall, G.D., and Morozov, I.Y. (2023). The Colorectal Cancer Microbiota Alter Their Transcriptome To Adapt to the Acidity, Reactive Oxygen Species, and Metabolite Availability of Gut Microenvironments. mSphere *8*, e00627-22. https://doi.org/10.1128/msphere.00627-22.
65. Gharbia, S.E., and Shah, H.N. (1991). Pathways of glutamate catabolism among Fusobacterium species. Microbiology *137*, 1201–1206. https://doi.org/10.1099/00221287-137-5-1201.
66. Sakanaka, A., Kuboniwa, M., Shimma, S., Alghamdi, S.A., Mayumi, S., Lamont, R.J., Fukusaki, E., and Amano, A. (2022). Fusobacterium nucleatum Metabolically Integrates Commensals and Pathogens in

Oral Biofilms. mSystems *7*, e0017022. https://doi.org/10.1128/msystems.00170-22.

67. Ternes, D., Tsenkova, M., Pozdeev, V.I., Meyers, M., Koncina, E., Atatri, S., Schmitz, M., Karta, J., Schmoetten, M., Heinken, A., et al. (2022). The gut microbial metabolite formate exacerbates colorectal cancer progression. Nat. Metab. *4*, 458–475. https://doi.org/10.1038/s42255-022-00558-0.
68. Martí i Líndez, A.-A., and Reith, W. (2021). Arginine-dependent immune responses. Cell. Mol. Life Sci. *78*, 5303–5324. https://doi.org/10.1007/s00018-021-03828-4.
69. Godlewska, U., Bulanda, E., and Wypych, T.P. (2022). Bile acids in immunity: Bidirectional mediators between the host and the microbiota. Front. Immunol. *13*, 949033.
70. Liu, Y., Zhang, S., Zhou, W., Hu, D., Xu, H., and Ji, G. (2022). Secondary Bile Acids and Tumorigenesis in Colorectal Cancer. Front. Oncol. *12*, 813745.
71. Poetsch, A.R. (2020). The genomics of oxidative DNA damage, repair, and resulting mutagenesis. Comput. Struct. Biotechnol. J. *18*, 207–219. https://doi.org/10.1016/j.csbj.2019.12.013.
72. Jokela, T.A., Karppinen, J.E., Kärkkäinen, M., Mecklin, J.-P., Walker, S., Seppälä, T.T., and Laakkonen, E.K. (2024). Circulating metabolome landscape in Lynch syndrome. Cancer Metabol. *12*, 4. https://doi.org/10.1186/s40170-024-00331-9.
73. Pastor, D.M., and Schlom, J. (2021). Immunology of Lynch Syndrome. Curr. Oncol. Rep. *23*, 96. https://doi.org/10.1007/s11912-021-01085-z.
74. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
75. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.
76. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
77. Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics *28*, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
78. Besemer, J., and Borodovsky, M. (1999). Heuristic approach to deriving models for gene finding. Nucleic Acids Res. *27*, 3911–3920. https://doi.org/10.1093/nar/27.19.3911.
79. Buchfink, B., Reuter, K., and Drost, H.-G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods *18*, 366–368. https://doi.org/10.1038/s41592-021-01101-x.
80. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Mol. Biol. Evol. *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.
81. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). LightGBM: a highly efficient gradient boosting decision tree. In Proceedings of the 31st International Conference on Neural Information Processing Systems NIPS'17 (Curran Associates Inc.), pp. 3149–3157.
82. Head, T., Kumar, M., Nahrstaedt, H., Louppe, G., and Shcherbatyi, I. (2021). Scikit-Optimize/Scikit-Optimize. Version v0.9.0 (Zenodo). https://doi.org/10.5281/zenodo.5565057.
83. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
84. Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. *2*, 56–67. https://doi.org/10.1038/s42256-019-0138-9.
85. Mori, H., Maruyama, T., Yano, M., Yamada, T., and Kurokawa, K. (2018). VITCOMIC2: visualization tool for the phylogenetic composition of microbial communities based on 16S rRNA gene amplicons and metagenomic shotgun sequencing. BMC Syst. Biol. *12*, 30. https://doi.org/10.1186/s12918-018-0545-2.
86. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., and Glöckner, F.O. (2014). The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. *42*, D643–D648. https://doi.org/10.1093/nar/gkt1209.
87. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., and Ishiguro-Watanabe, M. (2023). KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. *51*, D587–D592. https://doi.org/10.1093/nar/gkac963.
88. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.
89. Nougayrède, J.-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006). *Escherichia coli* Induces DNA Double-Strand Breaks in Eukaryotic Cells. Science *313*, 848–851. https://doi.org/10.1126/science.1127059.
90. UniProt Consortium (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.
91. Shiroma, H., Darzi, Y., Terajima, E., Nakagawa, Z., Tsuchikura, H., Tsukuda, N., Moriya, Y., Okuda, S., Goto, S., and Yamada, T. (2023). Enteropathway: the metabolic pathway database for the human gut microbiota. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.28.546710v1.
92. Erawijantari, P.P., Mizutani, S., Shiroma, H., Shiba, S., Nakajima, T., Sakamoto, T., Saito, Y., Fukuda, S., Yachida, S., and Yamada, T. (2020). Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. Gut *69*, 1404–1415. https://doi.org/10.1136/gutjnl-2019-319188.
93. Soga, T., Ohashi, Y., Ueno, Y., Naraoka, H., Tomita, M., and Nishioka, T. (2003). Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry. J. Proteome Res. *2*, 488–494. https://doi.org/10.1021/pr034020m.
94. Mishima, E., Fukuda, S., Mukawa, C., Yuri, A., Kanemitsu, Y., Matsumoto, Y., Akiyama, Y., Fukuda, N.N., Tsukamoto, H., Asaji, K., et al. (2017). Evaluation of the impact of gut microbiota on uremic solute accumulation by a CE-TOFMS-based metabolomics approach. Kidney Int. *92*, 634–645. https://doi.org/10.1016/j.kint.2017.02.011.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Lynch Syndrome subject's feces samples	National Cancer Center Japan	NCC
Deposited data		
Whole-metagenome shotgun sequencing (Japan)	This study	DDBJ: DRA017459
Whole-metagenome shotgun sequencing (Japan)	Yachida et al.[26]	DDBJ: PRJDB4176
Whole-metagenome shotgun sequencing (Israel)	Naddaf et al.[24]	SRA: PRJNA939026
Chemicals, peptides, and recombinant proteins		
Methionine sulfone	Alfa Aesar	Cat# A17027
2-(N-morpholino)ethanesulfonic acid	DOJINDO LABORATORIES	Cat# 341-01622
D-Camphor-10-sulfonic acid	FUJIFILM Wako	Cat# 037-01032
Zirconia/silica beads	TOMY SEIKO	Cat# ZB-30, ZSB- 01
Methanol	FUJIFILM Wako	Cat# 134-14523
Chloroform	FUJIFILM Wako	Cat# 033-08631
5-kDa-cutoff filter column	Human Metabolome Technologies	Cat# UFC3LCCNB- HMT
Capillary column for negative mode	Nacalai Tesque	Cat# 07584-44
Capillary column for positive mode	Molex Incorporated	Cat# 1068150017
Critical commercial assays		
GNOME DNA Isolation Kit	MP Biomedicals	Cat# 11343139
Agilent 4200 TapeStation System	Agilent Technologies	Cat# G2991BA
Nextera XT DNA Library Preparation Kit	illumina	Cat# FC-131-1096
Nextera XT Index Kit v2	illumina	Cat# FC-131-2001
HiSeq PE Rapid Cluster Kit v2-HS	illumina	Cat# PE-402-4002
HiSeq Rapid Duo cBot v2 Sample Loading Kit	illumina	Cat# CT-403-2001
HiSeq 2500	illumina	Cat# SY-401-2501
cBot Cluster Generation System	illumina	Cat# SY-301-2002
HiSeq Rapid SBS Kit v2-HS	illumina	Cat# FC-402-4002
Zirconia Beads	Biospec	Cat# 11079123ss
Zirconia/Silica Beads	Biospec	Cat# 11079101z
Micro Smash	TOMY	Cat# MS-100R
UltrafreeMC-PLHCC 250/pk for Metabolome Analysis	Human Metabolome Technologies	Cat# UFC3LCCNB- HMT
Agilent 7100 Capillary Electrophoresis System	Agilent Technologies	Cat# 7100CE
Agilent 6224 TOF LC/MS	Agilent Technologies	Cat#6224 TOF LC/MS
Agilent 6230 TOF LC/MS	Agilent Technologies	Cat#6230 TOF LC/MS
Software and algorithms		
HiSeq Control Software (HCS) v2.2.68	illumina	
Real-Time Analysis (RTA) v1.18.66.3	illumina	
Bcl2fastq2 v2.17	illumina	

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bowtie 2 (version 2.2.9)	Langmead and Salzberg[74]	https://sourceforge.net/projects/bowtie-bio/files/bowtie2/2.2.9/bowtie2-2.2.9-linux-x86_64.zip/download
cutadapt (version 1.9.1)	Martin et al.[75]	https://github.com/marcelm/cutadapt
BLAST+ (version 2.2.30)	Altschul et al.[76]	https://www.ncbi.nlm.nih.gov/books/NBK279690
IDBA_UD (version 1.1.1)	Peng et al.[77]	http://hku-idba.googlecode.com/files/idba-1.1.1.tar.gz
MetaGeneMark (v.3.26)	Bessemer and Bodorovsky[78]	http://exon.gatech.edu/GeneMark/license_download.cgi
DIAMOND (v2.1.2)	Buchfink et al.[79]	https://github.com/bbuchfink/diamond
eggnog-mapper (v2.1.9)	Cantalapiedra et al.[80]	https://github.com/eggnogdb/eggnog-mapper/tree/2.1.9
R 4.2.0	R Core Team	https://www.R-project.org/
R package stats	R Core Team	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html
R package MatchIt	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/MatchIt/index.html
R package vegan	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/vegan/index.html
R package rstatix	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/rstatix/index.html
R package ggpubr	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/ggpubr/index.html
R package clusterProfiler	The Comprehensive R Archive Network	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R package ape	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/ape/index.html
R package factoextra	The Comprehensive R Archive Network	https://cran.r-project.org/web/packages/factoextra/index.html
R package DirichletMultinomial	Bioconductor	https://bioconductor.org/packages/release/bioc/html/DirichletMultinomial.html
R package SIAMCAT	Bioconductor	https://github.com/zellerlab/siamcat
Python	Python Software Foundation	https://www.python.org/
LightGBM (version 3.3.3)	Ke et al.[81]	https://github.com/microsoft/LightGBM/tree/master/python-package
scikit-opt (version 0.9.0)	Head et al.[82]	https://scikit-optimize.github.io/stable/index.html
scikit-learn (version 1.1.3)	Pedregosa et al.[83]	https://scikit-learn.org/stable/index.html
SHAP (version 0.41.0)	Lundberg et al.[84]	https://github.com/shap/shap
Other		
phiX DNA sequences	RefSeq NC_001422.1	https://www.ncbi.nlm.nih.gov/nuccore/9626372
Human genome sequences (GRCh38)	RefSeq GCF_000001405.26	https://www.ncbi.nlm.nih.gov/nuccore/CM000663.2/ - https://www.ncbi.nlm.nih.gov/nuccore/CM000686.2/
VITCOMIC2	Mori et al.[85]	http://vitcomic.org/vitcomic2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LTP SILVA release 123	Yilmaz et al.[86]	https://www.arb-silva.de/documentation/release-123/
Kyoto Encyclopedia of Genes and Genomes (KEGG) (retrieved 2023/03/31)	Kanehisa et al.[87]	ftp://ftp.bioinformatics.jp/kegg/
Fusobacterium nucleatum virulent factors	Umana et al.[37]	http://fusoportal.org/fasta/protein/23726/23726_FusoPortal_Genes_AA.faa *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 23726 v1 *cbpF* (Gene 4) *radD* (Gene 32) *fadA2* (Gene 35) *fadA3* (Gene 357, Gene 368, Gene 864) *fadA* (Gene 878) *FnDps* (Gene 1296) *fap2* (Gene 2068)
eggNOG (v5.0.2)	Huerta-Cepas et al.[88]	http://eggnog5.embl.de/app/home
Colibactin biosynthesis cluster gene sequence	Nougayrede et al.[89]	https://www.uniprot.org/citations/16902142
UniProt (release 202205)	The UniProt Consortium[90]	https://ftp.uniprot.org/pub/databases/uniprot/previous_major_releases/release-2022_05/knowledgebase/
EnteroPathway (retrieved 2023/10/05)	Shiroma et al.[91]	https://enteropathway.org

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Takuji Yamada (takuji@bio.titech.ac.jp).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The raw sequencing data reported in this study have been deposited in the DDBJ Sequence Read Archive (DRA) and are publicly available as of the date of publication. Public metagenome data of Japanese sporadic CRC cohort was published in a previous study.[26] Public metagenome data of Israel LS cohort[24] was obtained from SRA. Accession numbers are listed in the key resources table. De-identified and processed metagenomics, metabolomics, and lifestyle data are presented in Table S1.
- All software tools used in this paper are publicly available and are listed in the key resources table. This paper does not report original code as no new code was generated.
- Any additional information required to reanalyze the data reported in this study will be available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Regulatory compliance

Stool samples, questionnaires, and clinical information were obtained under informed consent and with the approval of the institutional review boards of each participating institute (National Cancer Center Hospital, 2013–244; Tokyo Institute of Technology, 2014018; Osaka University Hospital, 20064).

Subject recruitment

Seventy-one patients (30 males and 41 females, aged 12–79 (47.89 on average) years) were diagnosed with Lynch Syndrome by genetic testing. At the time of fecal sampling and questionnaire collection, all patients underwent a colonoscopy to check for cancerous conditions at the National Cancer Center Hospital, Tokyo (Table 1). Patients with pathogenic variants of germline mismatch repair (MMR) genes were included. Patients with record of gastrectomy or whose stool samples were insufficient for data collection were excluded.

METHOD DETAILS

Fecal sample collection and DNA extraction

In this study, 71 stool samples were collected from 71 patients with LS. All samples were collected on the same day before total colonoscopy. Detailed protocols for stool sample collection have been previously described.[26,92] Briefly, collected stool samples were immediately stored on dry ice at the hospitals, then stored at -80°C before further experiments.

Shotgun metagenomic sequencing

This study used 71 LS fecal samples for shotgun metagenomic sequencing. Total DNA was extracted from fecal samples using the bead-beating method with a GNOME DNA Isolation Kit (MP Biomedicals). The DNA quality was assessed using an Agilent 4200 TapeStation (Agilent Technologies). After final precipitation, the DNA samples were resuspended in TE buffer and stored at -80°C before analysis. Sequencing libraries were generated using the Nextera XT DNA Sample Kit (Illumina). The library quality was confirmed using an Agilent 4200 TapeStation instrument. Whole-genome shotgun sequencing of fecal samples was performed using the HiSeq 2500 platform (Illumina). All samples were paired-end sequenced with a 150-bp read length to a targeted data size of 5.0 Gb. A total of 3,541,042,358 (49,873,836 on average) paired-end reads covering 531,735,402,166 (7,489,231,016 on average) base pairs were obtained (Table S8).

Sequencing reads quality control

The raw sequence reads underwent a series of quality control steps, as follows: Reads containing the letter 'N' (unidentified base pair) were discarded. Reads containing the bacteriophage *phiX* DNA sequences were identified by mapping them against the reads using Bowtie2 (version 2.2.9)[74] with preset options in 'fast-local' and discarded. Reads were trimmed for adapter and primer sequences using cutadapt (version 1.9.1),[75] for which the following options were used ("-a CTGTCTCTTATACACATCTCCGAGCCCACGAGAC -O 33 -q 17" for the forward primer sequence; and "-a CTGTCTCTTATACACATCTGACGCTGCCGACGA -O 32 -q 17" for the reverse primer sequence). Reads with quality values of 17 or less consecutively were tailed-cut at the 3′ termini within the cutadapt program. Next, reads shorter than 50-bp long were discarded. Reads of average quality values of 25 or less were discarded. Next, reads were mapped against the human genome (GRCh38) using Bowtie2 (version 2.2.9)[74] (no-hd, no-sq, fast-local). Those that were mapped were considered derived from the human genome and discarded. Finally, the unpaired reads were discarded. Consequently, 3,180,329,758 (44,793,377 on average) quality-controlled reads were used for the subsequent analyses (Table S8).

Taxonomy annotation and abundance estimation

Taxonomy annotation and abundance estimation were performed on both the LS and sporadic CRC cohorts. The high-quality reads were aligned with a pre-calculated operational taxonomic unit (OTU) dataset stored in VITCOMIC2[85] using BLAST+ (version 2.2.30)[76] (cut-off: E-value < 1×10^{-8}) to obtain bacterial and archaeal 16S rRNA reads and exclude tRNA, 23S rRNA or the internal transcribed spacer (ITS) sequences.

The filtered reads were aligned to the 16S rRNA sequence database provided by The All-Species Living Tree (LTP) project of the SILVA database (version 123)[86] using BLASTn[76] (cut-offs: E-value < 1×10^{-8}, sequence identity > 97%, alignment coverage > 80%, bit score > 70). Alignment(s) with the highest sequence identity and bit score were used as aligned reads to estimate downstream abundance. In case of multiple alignments, the aligned reads were divided by the number of aligned taxa so that they could be "shared." Species-level relative abundance was computed per sample and was defined as the number of reads assigned to the species divided by the total number of LTP-aligned reads in the sample. Relative abundances at the genus level and higher taxonomic ranks were calculated as the sum of the relative abundances of all member species. The generated profiles are referred to as "species profile" and "genus profile" hereafter. The species profile consisted of 8,071 species aggregated into 1,898 genera (Table S9).

Genome assembly, gene prediction, and gene abundance estimation

Prior to gene prediction, high-quality reads were assembled per sample using IDBA_UD (version 1.1.1)[77] with the following parameters: mink = 20, maxk = 120, and step = 10. Post-assembly, open reading frames (ORFs) were predicted on the obtained scaffolds using MetaGeneMark (version 3.26)[78] with parameter "–g 11," giving a total of 15,527,604 predicted ORFs. Prior to downstream annotation, the ORFs were filtered to include only those with lengths of 50 amino acids or longer; 14,345,386 ORFs remained after filtering.

Gene abundance was estimated by mapping quality-controlled reads onto scaffolds using Bowtie2 (version 2.2.9).[74] ORF read coverage was calculated by dividing the number of base pairs mapped onto the corresponding scaffold regions by the predicted ORF length.

Gene annotation

Gene annotation was performed on both the LS and sporadic CRC cohorts. Length-filtered ORFs were aligned against a reference database of prokaryotic complete genomes recorded in the Kyoto Encyclopedia of Genes (KEGG) and Genomes[87] (obtained on 2023/03/31) using DIAMOND (version 2.1.2)[79] (cut-offs: sequence identity > 40, bit score > 70, coverage > 80). Alignment results were concatenated with an ORF read coverage table per sample to calculate functional gene abundance.

For cases in which one ORF corresponded to a single gene, read coverage was used as is. If an ORF correspond to multiple genes, the read coverage was "shared" over all corresponding genes by dividing the read coverage with the number of aligned genes. Finally, gene-level

relative abundances were calculated per sample by dividing the read coverage of the gene by the sum of all gene read coverages. Relative gene abundance was further aggregated to the KEGG Orthology (KO) level. The generated profile contained 8,198 KO and referred to as the "KO gene profiles" (Table S9).

Virulent factor annotation

To detect *Fusobacterium nucleatum* virulent factors, reference amino acid sequences of CRC-associated virulent factors (*fap2*, *fadA*, *fadA2*, *fadA3*, *FnDps*, *cbpF*, and *radD*) were obtained from a previous study.[37] To expand the search range for these virulence genes, reference genomes were annotated using eggnog-mapper[80] (version 2.1.9) to obtain eggNOG[88] IDs. These eggNOG IDs serve as references for putative *Fusobacterium nucleatum* CRC-associated virulence factors.

Length-filtered ORFs from the LS and sporadic CRC cohorts were annotated using eggnog-mapper[80] (version 2.1.9). Next, the eggNOG-annotated ORF alignment results were concatenated with an ORF read coverage table to calculate eggNOG read coverage. If an ORF was annotated with multiple eggNOGs, the read coverage was split evenly for all eggNOGs. Each eggNOG read coverage was then divided by the total sum of the eggNOG read coverages to obtain the relative abundance of each eggNOG. Finally, reference-matched eggNOG IDs were extracted and utilized for *Fusobacterium nucleatum* CRC-associated virulence factor abundance in downstream analysis.

UniProt IDs for the colibactin-producing genes were obtained from a previous study.[89] Next, length-filtered ORFs from the LS and sporadic CRC cohorts were aligned against the UniProt TrEMBL database (release 2022_05)[90] using DIAMOND (version 2.1.2)[79] (cut-offs: sequence identity > 40, bit score > 70, and coverage > 80). Alignment results were concatenated with an ORF read coverage table per sample to calculate UniProt gene abundance. Finally, reference-matched UniProt IDs were used for colibactin-producing gene abundance in the downstream analysis.

Metabolome analysis

The present study used 70 LS samples for the metabolomic analysis. Quantitative analysis of charged metabolites was performed using capillary electrophoresis time-of-flight mass spectrometry (CE-TOF-MS) as previously described.[93] Fecal metabolites were extracted by vigorous shaking with methanol containing 20 μM of methionine sulfone and d-camphol-10-sulfonic acid each as internal standards.[94] All the CE-TOF-MS experiments were performed using an Agilent CE system. CE-TOF-MS detected 449 compounds from the LS and sporadic CRC cohort samples, referred to as the "metabolite profile" (Table S9).

QUANTIFICATION AND STATISTICAL ANALYSIS

Cohort matching

For the comparison between LS and non-LS controls, non-LS controls were selected from our previous study.[26] To consider the effects of possible confounding covariates (e.g., sample numbers, age, and sex) on the gut microbiota and fecal metabolite composition, *in silico* propensity score-based cohort matching was performed. The propensity score for each subject was calculated based on age and sex. Then, non-LS controls with similar propensity score were selected as a match for the LS controls ("matchit" function, method = "optimal," ratio = 1, R MatchIt). Age and sex distributions were evaluated post-matching to validate cohort demographic similarity. Matched non-LS controls were used for downstream statistical tests against the LS controls.

Microbial community structure analysis

The Shannon–Wiener index was calculated on the LTP annotated taxonomic profiles at the species level ("diversity" function, index = "shannon," R vegan). The calculated index was compared between the LS and sporadic cohorts using the Mann–Whitney U test and within the LS cohort using the pairwise Mann–Whitney U test.

Bray–Curtis dissimilarity were calculated from genus level relative abundances ("vegdist" function, method = "bray," R vegan). Principal coordinate analysis (PCoA) was performed on the calculated Bray–Curtis dissimilarity to visualize gut microbiota composition ("pcoa" function, R ape).

Quantification of the variance within the gut microbiota composition was evaluated with permutational multivariate analysis of variance (PERMANOVA) ("adonis2" function, permutations = 10000, by = "margin," R vegan). Demographics (e.g., age and sex) and medical information (e.g., endoscopy findings, mutated MMR genes, and colectomy records) were selected as covariates to explain the taxonomic profile variances at genus level.

Subjects were clustered based on gut microbiota composition by fitting all xx genus-level abundances into a Dirichlet Multinomial Mixture (DMM) model ("dmn" function, R DirichletMultinomial). The optimal number of community types was determined based on the lowest Laplace approximation score (Figure S1C).

Statistical tests

Microbial features (species and KOs) with low read alignment or coverage and low prevalence were discarded (read alignment cutoffs for species and read coverage cutoffs for KO were 10.0, respectively; the prevalence cutoff was 5%). Similarly, fecal metabolites with low concentrations and prevalence were discarded (concentration cutoff: 10 nmol/g; prevalence cutoff: 20%). A higher prevalence threshold for fecal metabolites was used to remove batch-specific metabolites.

Pairwise Mann–Whitney *U* test ("wilcox_test" function, R rstatix) was performed within the LS cohort or sporadic cohort to identify taxonomic, metagenomic and metabolite features with differential abundances. The Mann–Whitney *U* test was also performed between the LS and non-LS controls. *p* value of 0.05 was considered at be statistically significant.

The generalized fold-change, a pseudo-fold-change calculated as the geometric mean of the differences between quantiles across both groups, was calculated for each comparison (adapted from R SIAMCAT). Effect size estimation and magnitude interpretation was done for each comparison ("wilcox_effsize" function, R rstatix). Benjamini–Hochberg correction, a false discovery rate estimation method, was performed on the *p* values to obtain the FDR corrected *q* value ("p.adjust" function, method = "BH," R stats).

Training CRC classifier

To validate the findings identified by the statistical tests, CRC classifiers were trained based on species profiles and then screened for microbial markers based on species contribution to model decisions. As previous studies have shown that gut microbiota profiles vary according to CRC progression stage,[26] five classification tasks were identified for the classifier: differentiation between non-LS controls and 1) sporadic adenoma, 2) stage 0 CRC, 3) stages I/II CRC, 4) stages III/IV CRC, and 5) stages I–IV CRC. Only non-LS subject-species profiles were used to train the classifiers.

Gradient boosted decision tree (GBDT), as implemented in Light Gradient Boosted Model (LightGBM),[81] was used as the classifier architecture ("LGBMClassifier" function, boosting = "gbdt," objective = "binary," metric = "auc," learning_rate = 0.1, bagging_freq = 1, random_state = 0, Python LightGBM). Abundance filtering was performed by selecting species with a mean abundance higher than a specific threshold for classifier training. Bayesian optimization was performed in a nested cross-validation (CV) framework for hyperparameter tuning. Five inner loops ("BayesSearchCV" function, n_iter = 30, Python scikit-optimize[82]) and five outer loops ("cross_validate" function, scoring= "roc_auc," Python scikit-learn[83]) were used for the nested CV. Data split was done in a stratified fashion for both inner and outer fold to ensure similar case and control balance ("StratifiedKFold" function, n_splits = 5, shuffle = True, random_state = 0, Python scikit-learn). Details of the filtering thresholds and hyperparameter search ranges are listed in Table S10.

Classifier performance was evaluated by area under the receiver-operating characteristic curve (AU-ROC) ("roc_curve" function, Python scikit-learn[83]). Classifier performance during training was evaluated using the mean AU-ROC of the five-fold CV. The classifier performance for the three LS subject classification tasks (LS control vs. LS-adenoma, LS control vs. LS-CRC, and LS control vs. LS-colectomy) was evaluated with no CV.

CRC probability of each subject was estimated using scikit-learn "predict_proba" function. For non-LS subjects, CRC probability estimation was performed in a leave-one-out manner ("LeaveOneOut" function, Python scikit-learn[83]), with classifier parameters set to be identical with the final classifier. For the patients with LS, CRC probability estimation was parallelly performed with classifier performance evaluation. Because the CRC probability may be affected by the classifier performance and class ratio, we performed probability normalization on the estimated probability values. The CRC probability was normalized with a min–max normalization within each classification task.

Global and local feature importance analysis

To screen for potential CRC microbial markers, the importance of the species in the classifier was estimated. Species importance was obtained using "feature_importance" function (parameter importance = "gain") from Python LightGBM.[81] Species contributions obtained in this manner were estimated for the entire model (global importance) rather than for specific subject prediction (local importance).

A previous study proposed that while global importance is sufficient to obtain a general overview of microbe-CRC associations, local importance estimation is essential to elucidate microbe–CRC associations for a specific subset of subjects.[31] Here, "TreeExplainer" function of Python SHAP[84] was used to estimate Shapley Additive Explanations (SHAP) values. The estimated SHAP values served as indicators of the local importance of each species in CRC prediction in individual subjects.

Principal coordinate analysis (PCoA) was performed on the estimated SHAP values for data exploration ("prcomp" function, R stats). Because the PCoA plot showed a clear separation between LS/non-LS controls and LS/non-LS CRC patients, we further explored the data for potential LS CRC-specific clusters. K-means clustering was performed on the first two principal components (PC1 and PC2) of correctly predicted LS and non-LS CRC subject's SHAP value ("kmeans" function, R 'stats'). The optimal number of clusters were decided by an elbow plot of the total within cluster square sums ("fviz_nbclust" function, FUNcluster = 'kmeans,' method = 'wss,' R factoextra).

To obtain a visual representation of the local feature importance for LS CRC classification and the feature importance trends within a specific cluster, the correctly predicted SHAP values of the LS CRC subjects SHAP values was summarized into a waterfall plot.

Overrepresentation analysis (ORA)

Overrepresentation analysis was performed to identify metabolic modules with an overrepresentation of KOs and metabolites that were differentially abundant between the case and control groups. The feature set used for ORA was created by aggregating both KOs and KEGG Compounds into KEGG MODULE level metabolic modules (KEGG database obtained on 2023/03/31) (Table S11). Furthermore, to encompass metabolic modules not available in KEGG, another custom gene set was created by aggregating both KOs and KEGG Compounds into EnteroPathway[91] MODULE level metabolic modules (EnteroPathway database obtained on 2023/10/05) (Table S11). ORA was

performed as implemented by the "enrichr" function of R clusterProfiler. Multiple testing corrections were performed using the Benjamini–Hochberg method, and modules with FDR $p < 0.1$ were considered as overrepresented modules.

Correlation analysis

Correlation analysis between microbe-gene, microbe-metabolite, and gene-metabolite profiles was performed to explore potential association between omics data. Spearman correlation coefficient and *p*-value was calculated with cor.test function of R stats (method = 'spearman').