Author/s:
Chu, YL;Georgeson, P;Clendenning, M;Mahmood, K;Walker, R;Como, J;Joseland, S;Preston, SG;Rice, T;Lynch, BM;Milne, RL;Southey, MC;Giles, GG;Phipps, AI;Hopper, JL;Win, AK;Rosty, C;Macrae, FA;Winship, I;Jenkins, MA;Buchanan, DD;Joo, JE

Title:
Intratumoural pks+ Escherichia coli is associated with risk of metachronous colorectal cancer and adenoma development in people with Lynch syndrome

Date:
2025-04

Citation:
Chu, Y. L., Georgeson, P., Clendenning, M., Mahmood, K., Walker, R., Como, J., Joseland, S., Preston, S. G., Rice, T., Lynch, B. M., Milne, R. L., Southey, M. C., Giles, G. G., Phipps, A. I., Hopper, J. L., Win, A. K., Rosty, C., Macrae, F. A., Winship, I. ,... Joo, J. E. (2025). Intratumoural pks+ Escherichia coli is associated with risk of metachronous colorectal cancer and adenoma development in people with Lynch syndrome. EBioMedicine, 114, https://doi.org/10.1016/j.ebiom.2025.105661.

Persistent Link:
https://hdl.handle.net/11343/360375

Intratumoural *pks*$^{+}$ *Escherichia coli* is associated with risk of metachronous colorectal cancer and adenoma development in people with Lynch syndrome

Yen Lin Chu,[a,b] Peter Georgeson,[a,b] Mark Clendenning,[a,b] Khalid Mahmood,[a,b,c] Romy Walker,[a,b] Julia Como,[a,b] Sharelle Joseland,[a,b] Susan G. Preston,[a,b] Toni Rice,[a,b] Brigid M. Lynch,[d,e] Roger L. Milne,[d,e,f] Melissa C. Southey,[d,f,g] Graham G. Giles,[d,e,f] Amanda I. Phipps,[h] John L. Hopper,[e] Aung K. Win,[e] Christophe Rosty,[a,b,i,j] Finlay A. Macrae,[k,l,m] Ingrid Winship,[l,m] Mark A. Jenkins,[e] Daniel D. Buchanan,[a,b,l,n] and Jihoon E. Joo[a,b,n,*]

[a]Colorectal Oncogenomics Group, Department of Clinical Pathology, Victorian Comprehensive Cancer Centre, The University of Melbourne, Parkville, Victoria, Australia
[b]Collaborative Centre for Genomic Cancer Medicine, Victorian Comprehensive Cancer Centre, The University of Melbourne, Parkville, Victoria, Australia
[c]Melbourne Bioinformatics, The University of Melbourne, Melbourne, Victoria, Australia
[d]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia
[e]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia
[f]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia
[g]Department of Clinical Pathology, The University of Melbourne, Parkville, Victoria, Australia
[h]Department of Epidemiology, University of Washington, Seattle, Washington, USA
[i]University of Queensland, Brisbane, Queensland, Australia
[j]Envoi Specialist Pathologists, Brisbane, Queensland, Australia
[k]Colorectal Medicine and Genetics, The Royal Melbourne Hospital, Parkville, Victoria, Australia
[l]Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, Melbourne, Victoria, Australia
[m]Department of Medicine, The University of Melbourne, Parkville, Victoria, Australia

eBioMedicine
2025;114: 105661
Published Online xxx
https://doi.org/10.1016/j.ebiom.2025.105661

Summary

Background The adverse gut microbiome may underlie the variability in risks of colorectal cancer (CRC) and metachronous CRC in people with Lynch syndrome (LS). The role of *pks*$^{+/-}$ *Escherichia coli* (*pks*$^{+/-}$ *E. coli*), Enterotoxigenic *Bacteroides fragilis* (ETBF), and *Fusobacterium nucleatum* (*Fn*) in CRCs and adenomas in people with LS is unknown.

Methods A total of 358 LS cases, including 386 CRCs, 90 adenomas, 195 normal colonic mucosa DNA from the Australasian Colon Cancer Family Registry were tested using multiplex TaqMan qPCR. Logistic regression was used to compare the intratumoural prevalence of each bacteria in Lynch CRCs with 1336 sporadic CRCs. Cox proportional-hazards regression estimated the associations of each bacteria with the risk of metachronous CRC and neoplasia.

Findings *Pks*$^{+}$ *E. coli* (odds ratio [95% confidence interval] = 1.60 [1.08–2.35], $P = 0.017$), *pks*$^{-}$ *E. coli* (3.87 [2.58–5.80], $P < 0.001$) and *Fn* (19.47 [13.32–28.87], $P < 0.001$) were significantly enriched in LS CRCs when compared with sporadic CRCs. *Pks*$^{+}$ *E. coli* in the initial CRC was associated with an increased risk of metachronous CRC (hazard ratio [95% confidence interval] = 2.32 [1.29–4.17], $P = 0.005$) and metachronous colorectal neoplasia (1.51 [1.02–2.23], $P = 0.040$) when compared with CRCs without *pks*$^{+}$ *E. coli*.

Interpretation *Pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli*, and *Fn* are enriched within LS CRCs, suggesting possible roles in CRC development in LS. Having intratumoural *pks*$^{+}$ *E. coli* is associated with increased risk of metachronous CRC, suggesting that, if validated, people with LS might benefit from *pks*$^{+}$ *E. coli* screening and eradication.

Funding This work was funded by an NHMRC Investigator grant (GNT1194896) and a Cancer Australia/Cancer Council NSW co-funded grant (GNT2012914).

*Corresponding author. Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Victorian Comprehensive Cancer Centre, 305 Grattan Street, Parkville, Victoria, 3010, Australia.
E-mail address: ji.joo@unimelb.edu.au (J.E. Joo).
[n]Joint senior authorship.

Keywords: Lynch syndrome; Metachronous CRCs; *pks*$^{+}$ *E. coli*; Enterotoxigenic *Bacteroides fragilis*; *Fusobacterium nucleatum*

Research in context

Evidence before this study

People with Lynch syndrome have an increased risk of developing colorectal cancer (CRC) including multiple primary CRCs during their life. However, the risk of CRC in Lynch syndrome is shown to be variable, where some individuals are almost certain to develop CRC whilst others are at population-level risk. The reason for this variation in CRC risk is unclear but postulated to be multifactorial. We hypothesise that dysbiosis of the gut microbiome underlies this variability in CRC risk in Lynch syndrome. Previous studies have shown differences in the gut microbiome composition between people with Lynch syndrome with and without a cancer history. Three genotoxic gut bacteria, *Escherichia coli* harbouring the *pks* island (*pks*$^{+}$ *E. coli*), Enterotoxigenic *Bacteroides fragilis* (ETBF), and *Fusobacterium nucleatum* (*Fn*), are variably associated with different molecular CRC subtypes including tumour mismatch repair deficiency. However, it is unknown if the intratumoural presence of these specific genotoxic gut bacteria are associated with risk of metachronous CRC in people with Lynch syndrome.

Added value of this study

This study showed that *pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli*, and *Fn* were significantly enriched in Lynch syndrome CRCs when compared with sporadic CRCs. We observed the intratumoural presence of *pks*$^{+}$ *E. coli* in the initial primary CRC is associated with increased risk of both metachronous CRCs and metachronous neoplasia in people with Lynch syndrome.

Implications of all the available evidence

The association between intratumoural *pks*$^{+}$ *E. coli* and increased risk of metachronous CRC, if validated by other studies, suggests that *pks*$^{+}$ *E. coli* may represent a biomarker to screen for in people with Lynch syndrome that could potentially improve risk assessment and their clinical management and surveillance.

Introduction

Lynch syndrome (LS) is the most common hereditary cancer syndrome and accounts for ~3% of all colorectal cancers (CRCs).[1,2] LS is caused by germline pathogenic variants in one of the DNA mismatch repair (MMR) genes, *MLH1*, *MSH2*, *MSH6*, and *PMS2*,[1] or by deletions in the 3′ end of the *EPCAM* gene that leads to silencing of *MSH2* expression.[3]

People with LS have an increased risk of developing CRC, among other cancers.[4] To reduce their risk of CRC, individuals are advised to undergo biennial colonoscopy surveillance with polypectomy to remove adenomatous precursor lesions.[5,6] There is substantial variation in risk of CRC for people with LS, where some individuals are almost certain to develop CRC while others are at population-level risk.[4,7] The reason for this variation in CRC risk is unknown, with both genetic[8–11] and environmental/lifestyle[12–14] factors investigated as modifiers of CRC risk.

Furthermore, people with LS show an elevated risk of developing multiple primary CRCs when compared with people who developed non-hereditary CRCs[15]: synchronous CRC is observed in ~5–7% of heterozygotes and metachronous CRC has been observed in ~24.1% of heterozygotes, with a lifetime risk of up to 70%.[15,16] Factors that impact the risk of metachronous CRC include the type of surgery (e.g., partial or total colectomy) and the affected MMR gene, with the *MLH1* and *MSH2* gene pathogenic variant heterozygotes having higher metachronous CRC risk.[17] Identifying other factors that modify the risk of metachronous CRC will improve the precision prevention of CRC in LS. We hypothesise that specific genotoxic gut bacteria may contribute to the variability in CRC development and metachronous CRC risk in LS.

Dysbiosis of the gut microbiome has been associated with CRC development.[18] Previous studies showed an enrichment of *pks*$^{+}$ *Escherichia coli* (*pks*$^{+}$ *E. coli*), Enterotoxigenic *Bacteroides fragilis* (ETBF), and *Fusobacterium nucleatum* (*Fn*) in the colonic mucosa of CRC-affected individuals compared to healthy individuals.[19–21] Our recent study has demonstrated that intratumoural presence of ETBF and *Fn* were associated with MMR-deficiency (MMRd) in CRC tumours, suggesting these gut bacteria may play a role in the development of LS-related CRC.[22] No studies to date have investigated the role of these specific genotoxic gut bacteria in CRCs and adenomas from people with LS.

The aim of this study was to determine the prevalence of *pks*$^{+/-}$ *E. coli*, *Fn*, and ETBF in CRC, adenoma, and normal colonic mucosa (NCM) tissue from people with LS, and to investigate the association between the intratumoural presence of these bacteria and molecular and clinicopathological characteristics, and risk of metachronous CRC and adenomas in people with LS.

Methods

Study participants

The study participants were people with a germline pathogenic variant in *MLH1, MSH2, MSH6, PMS2* or a 3′ deletion of *EPCAM* (Lynch Syndrome, LS) with CRC and/or adenoma tissue DNA available for testing from the Australasian Colon Cancer Family Registry (ACCFR; https://www.coloncfr.org).[23,24] LS cases included in this study were from both population- and clinic-based recruitment arms of the ACCFR, with details of ascertainment provided in Jenkins et al. and Newcomb et al.[23,24] ACCFR participants were followed up every ~5 years after completing the baseline questionnaire. Information on participants' history of polypectomy, colonoscopy, and surgery (partial or total colectomy) was collected using standardised questionnaires and from medical records. CRC and adenoma incidences were verified using obtainable pathology reports, cancer registry reports, medical records, and/or death certificates.[23,24] The study participants were 97.2% white, 1.8% Asian, 0.3% Pacific islander, 0.3% black, and 0.6% unknown race.

A group of 1336 participants from the ACCFR and from the Melbourne Collaborative Cohort Study[25] who developed sporadic, MMR-proficient (MMRp) CRC and had been previously tested for intratumoural presence of $pks^{+/-}$ *E. coli*, *Fn*, and ETBF[22] were included as a reference group. LS cases and heterozygotes of a variant of uncertain significance in the MMR genes or with *MLH1*-methylated MMRd CRCs were excluded from this reference group.

Written informed consent was obtained from all study participants or the next-of-kin if the participant was deceased to collect CRC, adenoma, and NCM materials. All participants included in this study had been de-identified and assigned a study ID number. The study protocol was approved by Human Research Ethics Committees at the University of Melbourne (Ethics approval #13094). The ethics is reviewed on a yearly basis by the institutional review board.

Germline MMR and CRC molecular and pathology characteristics

Details of the germline and CRC tumour testing are provided in Buchanan et al.[2] The MMR status of CRC and adenoma was determined by immunohistochemical staining for *MLH1, MSH2, MSH6*, and *PMS2* protein expression as previously described.[22]

A standardised pathological review was performed by expert gastrointestinal anatomical pathologist (CR) as previously described.[26] CRCs, adenomas and NCM samples from the ileo-cecal junction, caecum, ascending colon, hepatic flexure, and transverse colon were categorised as proximal, whereas samples from the splenic flexure, descending, and sigmoid colon were categorised as distal. Lesions from the recto-sigmoid junction and rectum were categorised as rectal.

Synchronous CRCs were defined as more than one primary CRC diagnosed in the same individual at initial staging. Metachronous CRC was defined as a new primary CRC diagnosed at least one year after a prior primary CRC diagnosis. In the 316 individuals with CRC tissues available, multiple CRC samples were tested in 35 and 25 individuals with metachronous and synchronous CRCs, respectively. In cases of multiple primary CRCs, one *index* CRC per participant was included in the analysis. The *index* CRCs were defined as: 1) the earliest primary CRC that was tested for bacteria in cases of metachronous CRC events, or 2) the case with the most complete histopathological data and testing results in cases of synchronous CRC events. In total, 106 participants who developed metachronous CRCs were identified. Of these, tissue from the first primary CRC was tested in 70 participants, while tissue from subsequent primary CRCs was available in 36 participants. Although the subsequent CRCs were considered index CRCs for these 36 cases, these participants were excluded from the analysis of metachronous cancer risk, as no other metachronous CRCs were identified. For the metachronous neoplasm development analysis, we chose the earliest event of adenoma where testing for the three bacteria was completed as the index adenoma.

Multiplex quantitative polymerase chain reaction assays for detecting *pks E. coli*, ETBF, and *Fn*

DNA was extracted from formalin-fixed paraffin embedded (FFPE) tissue from CRC (n = 386), adenoma (n = 90), and NCM from the resection margin (n = 195) or NCM adjacent to tumour (n = 56) samples from 358 LS participants from 215 families. The tumour rich regions (i.e., CRC), adenomas and NCM of the FFPE tissue were macrodissected as previously described.[22] The intratumoural presence of $pks^{+/-}$ *E. coli*, ETBF, and *Fn* was assessed by performing quantitative polymerase chain reaction (qPCR) targeting four specific bacterial genes using two multiplexed assays. The *ClbB* (*pks* island)[27] and *UidA* (*E. coli*)[28,29] genes were combined into one duplex assay and optimised with MGB probe used and fluorophore modified to FAM and VIC, respectively (Supplementary Table S1). Each DNA sample was categorised based on the amplification of the *ClbB* and *UidA* genes into either: 1) pks^{+} *E. coli* showing the amplification of both *ClbB* and *UidA* genes, demonstrating the presence of *pks* island and *E. coli* or 2) pks^{-} *E. coli* showing the amplification of *UidA* gene only, demonstrating the presence of *E. coli* without *pks* island (Supplementary Figure S1).

The *bft* (ETBF)[30] and *nusG* (*Fn*)[31,32] genes were duplexed and optimised with QSY probe used and fluorophore modified to ABY and JUN respectively (Supplementary Table S1). These combinations were chosen for optimal specificity and to avoid spectra overlaps. The housekeeping gene *SLCO2A1* was used as an internal control as previously described.[21] Each TaqMan reaction was performed in technical duplicates using TaqMan Multiplex Master Mix (Thermo Fisher

Scientific, California, USA). qPCR was performed on Thermo QuantStudio 7 (Thermo Fisher Scientific, California, USA). Prior to the testing, the multiplex assays were cross-validated for accuracy and sensitivity using the single-plex assays[22] on a subset of randomly selected samples with 100% concordance (data not shown).

The *pks*$^+$ *E. coli* strain (ID 34351) that were used as the positive control were collected as part of the Controlling Superbugs study[33] and classified as Extra-intestinal Pathogenic *E. coli*.[34] The *pks*$^-$ *E. coli* strains (MG1655) were provided by Dr. Dianna M. Hocking and Prof. Roy Robins-Browne from the Department of Microbiology and Immunology of the University of Melbourne. The positive and negative controls for ETBF and *Fn* were purchased from Leibniz-Institute DSMZ (Braunschweig, Germany). The bacterial genomic DNA were used as internal controls for each qPCR experiment.

A custom-designed TaqMan genotyping assay was used to detect *APC*: c.835-8A>G somatic mutation in tumour and adenoma samples as previously described.[22]

Statistical analysis

All statistical tests were performed using R programming software (version 4.3.1). Logistic regression was used to assess differences in the prevalence of each bacteria in the LS CRCs and sporadic CRCs, and the association of each bacteria with clinicopathological and molecular features. Univariable and multivariable logistics models were performed to estimate the association and both results are shown where relevant. In the multivariable analysis, the age, sex and anatomical location variables were examined and variables with a $P < 0.15$ from the univariable analysis were included in the multivariable analyses. $P < 0.05$ were considered statistically significant.

Cox proportional hazards regression models were used to assess associations between the presence of each bacterial species and the risk of metachronous CRC or metachronous CRC and adenoma (neoplasia) development. For analyses of the risk of metachronous CRCs, the data were censored at the development of metachronous CRC or at last known date or death (no metachronous CRCs reported), whichever occurred earlier. For analyses of the risk of metachronous neoplasia, the data were censored at the earliest event of polypectomy or metachronous CRCs after the index CRC or at last known date or death (no metachronous neoplasms reported), whichever occurred earlier. The cox modelling was performed using *survival* R package (v 3.6). The proportional hazards assumption for the cox regression model fit was tested by examining the scaled Schoenfeld residuals using the "cox.zph" function.

Role of funders

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Results

pks$^+$ *E. coli*, *pks*$^-$ *E. coli*, and *Fn* are enriched in CRC tumours from people with Lynch syndrome

A total of 358 participants with LS (49.7% female), comprising heterozygotes of germline pathogenic variants in *MLH1* (n = 161), *MSH2* (n = 143), *MSH6* (n = 28), and *PMS2* (n = 26), with DNA from ≥1 tissue samples for multiplex qPCR testing (Supplementary Table S2). Of these, 292 (82%) participants had only CRC tissue available for testing while 42 (12%) participants had only adenoma tissue, with 24 (7%) participants with both CRC and adenoma tissue tested (Supplementary Figure S2). A total of 386 CRCs were tested from 316 LS cases, which included 60 cases with DNA available from multiple primary CRCs (Supplementary Figure S3). The analysis also included 90 adenomas from 66 LS participants and 195 NCMs from 179 LS participants. The mean (±IQR) age at CRC diagnosis was 50.4 ± 18.1 years (Supplementary Table S2).

In 316 index LS CRCs, the prevalences of *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, ETBF, and *Fn* were 13.6%, 15.5%, 5.7%, and 57%, respectively (Table 1). When compared with the reference group of 1336 sporadic MMRp CRCs from Joo et al.,[22] the prevalence of *pks*$^+$ *E. coli* (odds ratio (OR) [95% confidence interval (CI)] = 1.60 [1.08–2.35], $P = 0.017$), *pks*$^-$ *E. coli* (OR = 3.87 [2.58–5.80], $P < 0.001$), and *Fn* (OR = 19.47 [13.32–28.87], $P < 0.001$) were significantly higher in those with LS CRCs (Table 1). The prevalence of ETBF did not differ between LS CRCs and sporadic MMRp CRCs (OR = 0.98 [0.56–1.63], $P = 0.933$).

In 66 index LS adenomas, the prevalences of *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, ETBF, and *Fn*, were 6.1%, 21.2%, 1.5%, and 21.2%, respectively (Supplementary Table S3). The prevalence of *Fn* was significantly lower in adenomas (OR = 0.22 [0.11–0.42], $P < 0.001$) and resection margin normal mucosa (OR = 0.51 [0.35–0.74], $P < 0.001$) when compared with LS CRCs. The prevalence of *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, and ETBF was not significantly different in adenoma or normal mucosa when compared with CRC (Supplementary Table S3). The *APC*: c.835-8A>G somatic mutation was found in only one of the 316 CRCs (0.32%) and none of 66 adenomas tested.

pks$^{+/-}$ *E. coli*, *Fn*, and ETBF are not associated with CRC diagnosis age, anatomical location and affected MMR gene

The association between the bacterial presence in CRCs and adenomas and age at CRC diagnosis, anatomical site and affected MMR gene is shown in Table 2. The presence of *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, *Fn* and ETBF in CRCs was not associated with CRC diagnosis age, anatomical site or affected MMR gene (all $P > 0.05$). Similarly, the presence of *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, *Fn*, and ETBF in adenomas was not significantly associated

	Lynch syndrome CRCs[a]	Sporadic MMRp CRCs[b]	*P* (Uni)	*P* (Multi)[c]	OR (±95% CI)
Total	316	1336	–	–	–
pks^{+} *E.coli*					
Negative, n (%)	233 (73.7%)	1001 (74.9%)	*Ref*	*Ref*	*Ref*
Positive, n (%)	**43 (13.6%)**	136 (10.2%)	0.106	**0.017**$^{sex\ +\ age}$	**1.60 (1.08–2.35)**
Unknown, n (%)	40 (12.7%)	199 (14.9%)	–	–	–
pks^{-} *E.coli*					
Negative, n (%)	227 (71.8%)	1077 (80.6%)	*Ref*	*Ref*	*Ref*
Positive, n (%)	**49 (15.5%)**	60 (4.5%)	**<0.001**	**<0.001**^{site}	**3.87 (2.58–5.80)**
Unknown, n (%)	40 (12.7%)	199 (14.9%)	–		–
ETBF			–	–	
Negative, n (%)	283 (89.6%)	1091 (81.7%)	*Ref*	–	*Ref*
Positive, n (%)	18 (5.7%)	71 (5.3%)	0.933	–	0.98 (0.56–1.63)
Unknown, n (%)	15 (4.7%)	174 (13%)	–	–	–
Fn					
Negative, n (%)	109 (34.5%)	1114 (83.4%)	*Ref*	*Ref*	*Ref*
Positive, n (%)	**180 (57%)**	81 (6.1%)	**<0.001**	**<0.001**$^{age\ +\ site}$	**19.47 (13.32–28.87)**
Unknown, n (%)	27 (8.5%)	141 (10.6%)	–	–	–

Bold was used to highlight the statistically significant associations. [a]Index CRCs (n = 316) excluding multiple CRCs from same person were included in the analysis. [b]MMRp CRCs excluding lynch syndrome cases, heterozygotes of a variant of uncertain significance in the MMR genes or with MLH1-methylated MMRd CRCs from Joo & Chu et al., 2024. Index CRCs (n = 1336) excluding multiple CRCs from same person were included in the analysis. [c]Age, gender and/or anatomical location was included in the multivariable model if it had a *P*-value of <0.15 from the univariable tests.

Table 1: **Intratumoural bacteria in Lynch syndrome and non-hereditary MMR-proficient CRCs.**

with age at adenoma diagnosis, anatomical site, affected MMR gene or MMR protein expression (all $P > 0.05$).

Intratumoural *pks*$^{+}$ *E. coli* is associated with increased risk of metachronous CRCs and metachronous neoplasms

Of the 316 participants, 106 LS carriers developed a metachronous CRC and 208 LS carriers were never reported to have developed a metachronous CRC during the follow-up period. Two cases with missing information on metachronous CRCs and 36 metachronous cases where DNA from the initial CRC events was unavailable for screening were excluded from this analysis. The presence of *pks*$^{+}$ *E. coli* in the index LS CRC was associated with increased risk of metachronous CRC (hazard ratio (HR) [95% CI] = 2.32 [1.29–4.17], $P = 0.005$) when compared with index LS CRCs without *pks*$^{+}$ *E. coli* (Fig. 1a). *Pks*$^{-}$ *E. coli*, *Fn*, and ETBF were not associated with risk of metachronous CRC (all $P > 0.05$, Fig. 1b–d).

The association between *pks*$^{+}$ *E. coli* and metachronous CRC risk was assessed separately by the affected MMR genes demonstrating a significant increased risk in *MLH1* pathogenic variant heterozygotes (HR = 3.68 [1.75–7.74], $P < 0.001$) and not in *MSH2* pathogenic variant heterozygotes (HR = 1.27 [0.37–4.31], $P = 0.7$) (Fig. 1e and f). *MSH6* and *PMS2* heterozygotes were not tested due to insufficient sample size.

Of the 278 Lynch cases with an index CRC tested that were included in the analysis, 188 cases had metachronous neoplasms (adenoma or CRC), and 90 cases had no neoplasms after the initial CRC. *Pks*$^{+}$ *E. coli* in the index CRC was associated with increased risk of metachronous neoplasms (HR = 1.51 [1.02–2.23], $P = 0.04$) (Fig. 2a). *Pks*$^{-}$ *E. coli*, *Fn*, and ETBF showed no evidence of an association with the risk of metachronous neoplasms (all $P > 0.05$, Fig. 2b–d). *Pks*$^{+}$ *E. coli* was associated with increased risk of metachronous neoplasms in *MLH1* pathogenic variant heterozygotes (HR = 1.99 [1.21–3.28], $P = 0.007$) but not for *MSH2* pathogenic variant heterozygotes (HR = 1.19 [0.58–2.44], $P = 0.6$) (Fig. 2e and f).

Concordance of bacterial presence between synchronous CRCs and between CRC and normal colonic mucosa

The concordance of either *pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli*, *Fn*, or ETBF within the CRC and NCM adjacent to the tumour from the same person was 100% (26/26), 96.2% (25/26), 80.6% (25/31), and 94.1% (32/34), respectively (Supplementary Figure S4). Where the CRC was positive for either *pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli*, *Fn*, or ETBF, the bacteria was also present in 100% (6/6), 100% (3/3), 72% (13/18), and 67% (2/3) of the adjacent NCM, respectively. For the CRC and matched NCM from the resection margin (non-adjacent), the concordance was 91.2% (124/136), 81.6% (111/136), 71.9% (105/146), and 94% (141/150) for *pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli*, *Fn*, and ETBF, respectively (Supplementary Figure S5). Where the CRC was positive for either *pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli*, *Fn*, or ETBF, these bacteria were also detected in 78% (17/22), 50% (11/22), 50% (5/10), and 63% (52/82) of the matched non-adjacent NCM.

		pks+ E. coli						*pks- E. coli*						ETBF						*Fn*					
	Total[ł]	Positive	Negative	NA	*P* (Uni)	*P* (Multi)[ǂ]	OR (±95 CI)	Positive	Negative	NA	*P* (Uni)	*P* (Multi)[ǂ]	OR (±95 CI)	Positive	Negative	NA	*P* (Uni)	*P* (Multi)[ǂ]	OR (±95 CI)	Positive	Negative	NA	*P* (Uni)	*P* (Multi)[ǂ]	OR (±95 CI)
Total CRCs, n (%)	**316**	43 (13.6)	233 (73.7)	40 (12.7)	-	-	-	49 (15.5)	227 (71.8)	40 (12.7)	-	-	-	18 (5.7)	283 (89.6)	15 (4.7)	-	-	-	180 (57)	109 (34.5)	27 (8.5)	-	-	-
CRC Diagnosis Age (mean year±IQR)		51.2± 21.7	48.3± 13.9	50.3± 19.9	0.187	0.090[title]	0.98 (0.95-1.00)	51.6± 16.5	48.1± 14.5	50.3± 19.9	0.095	-	0.98 (0.96-1.00)	48.2± 14.7	48.9± 16.0	49.4± 12.1	0.809	-	1.00 (0.97-1.04)	48.6± 17.6	49.6± 15.1	48.3± 14.5	0.503	-	1.01 (0.99-1.02)
CRC Site, n (%)																									
Proximal	188	26 (13.8)	141 (75)	21 (11.2)	*Ref*	-	-	25 (13.3)	142 (75.5)	21 (11.2)	*Ref*	-	-	15 (8)	167 (88.8)	6 (3.2)	*Ref*	-	-	113 (60.1)	62 (33)	13 (6.9)	*Ref*	-	-
Distal	48	5 (10.4)	37 (77.1)	6 (12.5)	0.552	-	1.36 (0.53-4.25)	8 (16.7)	34 (70.8)	6 (12.5)	0.518	0.666[aps]	1.26 (0.48-3.93)	0 (0)	45 (93.8)	3 (6.3)	0.991	-	not estimated	25 (52.1)	18 (37.5)	5 (10.4)	0.434	-	1.31 (0.66-2.58)
Rectum	37	9 (24.3)	23 (62.2)	5 (13.5)	0.093	-	0.47 (0.20-1.18)	3 (8.1)	29 (78.4)	5 (13.5)	0.409	0.069[aps]	0.44 (0.18-1.11)	1 (2.7)	35 (94.6)	1 (2.7)	0.275	-	not estimated	22 (59.5)	14 (37.8)	1 (2.7)	0.694	-	1.16 (0.54-2.41)
Unknown	43	3 (7)	32 (74.4)	8 (18.6)	-	-	-	13 (30.2)	22 (51.2)	8 (18.6)	-	-	-	2 (4.7)	36 (83.7)	5 (11.6)	-	-	-	20 (46.5)	15 (34.9)	8 (18.6)	-	-	-
Affected MMR gene, n (%)																									
MLH1	149	24 (16.1)	107 (71.8)	18 (12.1)	*Ref*	-	-	23 (15.4)	108 (72.5)	18 (12.1)	*Ref*	-	-	9 (6)	130 (87.2)	10 (6.7)	*Ref*	-	-	84 (56.4)	48 (32.2)	17 (11.4)	*Ref*	-	-
MSH2	121	16 (13.2)	88 (72.7)	17 (14)	0.552	0.179[title]	1.68 (0.80-3.68)	18 (14.9)	86 (71.1)	17 (14)	0.960	0.526[aps]	1.25 (0.63-2.56)	7 (5.8)	110 (90.9)	4 (3.3)	0.871	-	1.09 (0.39-3.13)	74 (61.2)	40 (33.1)	7 (5.8)	0.835	-	0.95 (0.56-1.60)
MSH6	21	1 (4.8)	18 (85.7)	2 (9.5)	0.185	0.171[title]	4.25 (0.80-78.84)	4 (19)	15 (71.4)	2 (9.5)	0.711	0.140[aps]	4.79 (0.90-89.23)	1 (4.8)	20 (95.2)	0 (0)	0.763	-	1.38 (0.24-26.22)	11 (52.4)	9 (42.9)	1 (4.8)	0.459	-	1.43 (0.54-3.70)
PMS2	25	2 (8)	20 (80)	3 (12)	0.297	0.207[title]	2.69 (0.70-17.79)	4 (16)	18 (72)	3 (12)	0.943	0.188[aps]	2.84 (0.73-18.96)	1 (4)	23 (92)	1 (4)	0.666	-	1.59 (0.28-30.06)	11 (44)	12 (48)	2 (8)	0.155	-	1.91 (0.78-4.72)
	Total[ł]	***pks+ E. coli***						***pks- E. coli***						**ETBF**						***Fn***					
Total Adenomas, n (%)	**66**	4 (6.1)	46 (69.7)	16 (24.2)	-	-	-	14 (21.2)	36 (54.5)	16 (24.2)	-	-	-	1 (1.5)	57 (86.4)	8 (12.1)	-	-	-	14 (21.2)	38 (57.6)	14 (21.2)	-	-	-
Adenoma Diagnosis Age (mean year±IQR)		41.7± 14.2	49.8± 14.9	47.9± 13.1	0.231	-	0.95 (0.87-1.03)	51.3± 21.5	48.3± 14.1	47.9± 13.1	0.461	-	1.02 (0.97-1.07)	58.6± n.a	48.5± 14.6	50.2± 7.83	0.435	-	1.06 (0.91-1.22)	49.6± 13.3	49.2± 15.9	47.3± 9.12	0.914	-	1.00 (0.96-1.05)
Adenoma Site, n (%)																									
Proximal	18	0 (0)	14 (77.8)	4 (22.2)	*Ref*	-	-	4 (22.2)	10 (55.6)	4 (22.2)	*Ref*	-	-	1 (5.6)	15 (83.3)	2 (11.1)	*Ref*	-	-	5 (27.8)	11 (61.1)	2 (11.1)	*Ref*	-	*Ref*
Distal	10	0 (0)	7 (70)	3 (30)	not estimated	-	not estimated	1 (10)	6 (60)	3 (30)	0.477	-	0.42 (0.02-3.72)	0 (0)	8 (80)	2 (20)	not estimated	-	not estimated	2 (20)	3 (30)	5 (50)	0.718	-	1.47 (0.16-11.92)
Rectum	6	0 (0)	5 (83.3)	1 (16.7)	not estimated	-	not estimated	3 (50)	2 (33.3)	1 (16.7)	0.224	-	3.75 (0.46-38.26)	0 (0)	5 (83.3)	1 (16.7)	not estimated	-	not estimated	1 (16.7)	4 (66.6)	1 (16.7)	0.630	-	0.55 (0.02-5.08)
Unknown	**32**	4 (12.5)	20 (62.5)	8 (25)	-	-	-	6 (18.7)	18 (56.3)	8 (25)	-	-	-	0 (0)	29 (90.6)	3 (9.4)	-	-	-	6 (18.7)	20 (62.5)	6 (18.7)	-	-	-
Affected MMR gene, n (%)																									
MLH1	29	3 (10.3)	20 (69)	6 (20.7)	*Ref*	-	-	7 (24.1)	16 (55.2)	6 (20.7)	*Ref*	-	-	0 (0)	27 (93.1)	2 (6.9)	*Ref*	-	-	18 (62.1)	6 (20.7)	5 (17.2)	*Ref*	-	*Ref*
MSH2	28	1 (3.6)	18 (64.3)	9 (32.1)	not estimated	-	not estimated	4 (14.3)	15 (53.6)	9 (32.1)	0.493	-	0.61 (0.14-2.45)	1 (3.6)	21 (75)	6 (21.4)	not estimated	-	not estimated	14 (50)	6 (21.4)	8 (28.6)	0.711	-	1.29 (0.33-4.97)
MSH6	7	0 (0)	6 (85.7)	1 (14.3)	not estimated	-	not estimated	2 (28.6)	4 (57.1)	1 (14.3)	0.891	-	1.14 (0.14-7.43)	0 (0)	7 (100)	0 (0)	not estimated	-	not estimated	6 (85.7)	1 (14.3)	0 (0)	0.556	-	0.50 (0.02-3.84)
PMS2	2	0 (0)	2 (100)	0 (0)	not estimated	-	not estimated	1 (50)	1 (50)	0 (0)	0.578	-	2.29 (0.08-63.65)	0 (0)	2 (100)	0 (0)	not estimated	-	not estimated	0 (0)	1 (50)	1 (50)	not estimated	-	not estimated
MMR IHC, n (%)																									
MMR-deficient	40	2 (5)	28 (70)	10 (25)	*Ref*	-	-	10 (25)	20 (50)	10 (25)	*Ref*	-	-	1 (2.5)	34 (85)	5 (12.5)	*Ref*	-	-	10 (25)	22 (55)	8 (20)	*Ref*	-	*Ref*
MMR-proficient	12	1 (8.3)	8 (66.7)	3 (25)	0.664	-	1.75 (0.08-20.72)	1 (8.3)	8 (66.7)	3 (25)	0.220	-	0.25 (0.01-1.65)	0 (0)	10 (83.3)	2 (16.7)	not estimated	-	not estimated	1 (8.3)	8 (66.7)	3 (25)	0.252	-	0.28 (0.01-1.81)
Unknown	14	1 (7.1)	10 (71.5)	3 (21.4)	-	-	-	3 (21.4)	8 (57.2)	3 (21.4)	-	-	-	0 (0)	13 (92.9)	1 (7.1)	-	-	-	3 (21.4)	8 (57.2)	3 (21.4)	-	-	-

ł Index CRCs (n = 316) excluding multiple CRCs from same person were included in the analysis.
ǂ Age, gender and/or anatomical location was included in the multivariable model if it had a P-value of < 0.15 from the univariable tests.

***Table 2*: Intra-tumoural and intra-adenoma presence of pks$^{+/-}$ *E. coli*, ETBF, and *F. nucleatum* and the association with clinicopathological characteristics.**

In 24 of the 316 cases, 28 synchronous CRCs were also tested for the bacteria. The overall concordance between synchronous CRCs within each person was 88% (15/17), 65% (11/17), 80% (16/20), 100% (22/22) for *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, *Fn*, and ETBF, respectively (Supplementary Figure S6). Where the index CRC was positive for *pks*$^+$ *E. coli*, *Fn*, and ETBF, the bacteria were also detected in 67% (4/6), 78% (14/18), and 100% (2/2) of the synchronous CRCs from the same person.

Of 316 index CRCs tested, 217 (68.7%) had ≥1 bacteria type detected per CRC and 99 (31.3%) had none of the four bacteria detected (Supplementary Figure S7a). In 217 CRCs that were detected with any of the bacteria, 51.2% presented solely with *Fn*. In 11.5% and 12.9% of the 217 CRCs showed co-presence of *Fn* with *pks*$^+$ *E. coli* or with *pks*$^-$ *E. coli*, respectively. The co-presence of *Fn* and ETBF was observed in 12 (5.53%) of the 217 CRCs. The co-presence of *Fn* and ETBF with either *pks*$^+$ or *pks*$^-$ *E. coli* was rarely observed, with only 2 (0.9%) and 2 (0.9%) showing the co-presence, respectively. In adenomas, only 24 (36.4%) were detected with ≥1 bacterial type where none showed co-presence of *Fn* and ETBF and 7 (29.2%) presented with both *Fn* and *pks*$^-$ *E. coli*. Only 1 (4.2%) co-presented with *Fn* and *pks*$^+$ *E. coli* (Supplementary Figure S7b).

Discussion

This study investigated the presence of genotoxic gut bacteria *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, ETBF, and *Fn* in CRCs and adenomas from people with LS. We found *pks*$^+$ *E. coli*, *pks*$^-$ *E. coli*, and *Fn* are enriched in CRCs from people with LS when compared with sporadic MMRp CRCs. Our findings showed that the initial (defined as "index") CRC with *pks*$^+$ *E. coli* present was associated with an increased risk of metachronous CRC and colorectal neoplasia in people with LS. These findings, for the first time, suggests a potential role of *pks*$^+$ *E. coli* which, if validated in independent studies, has the potential to improve our understanding of the risk of metachronous CRC in people with LS.

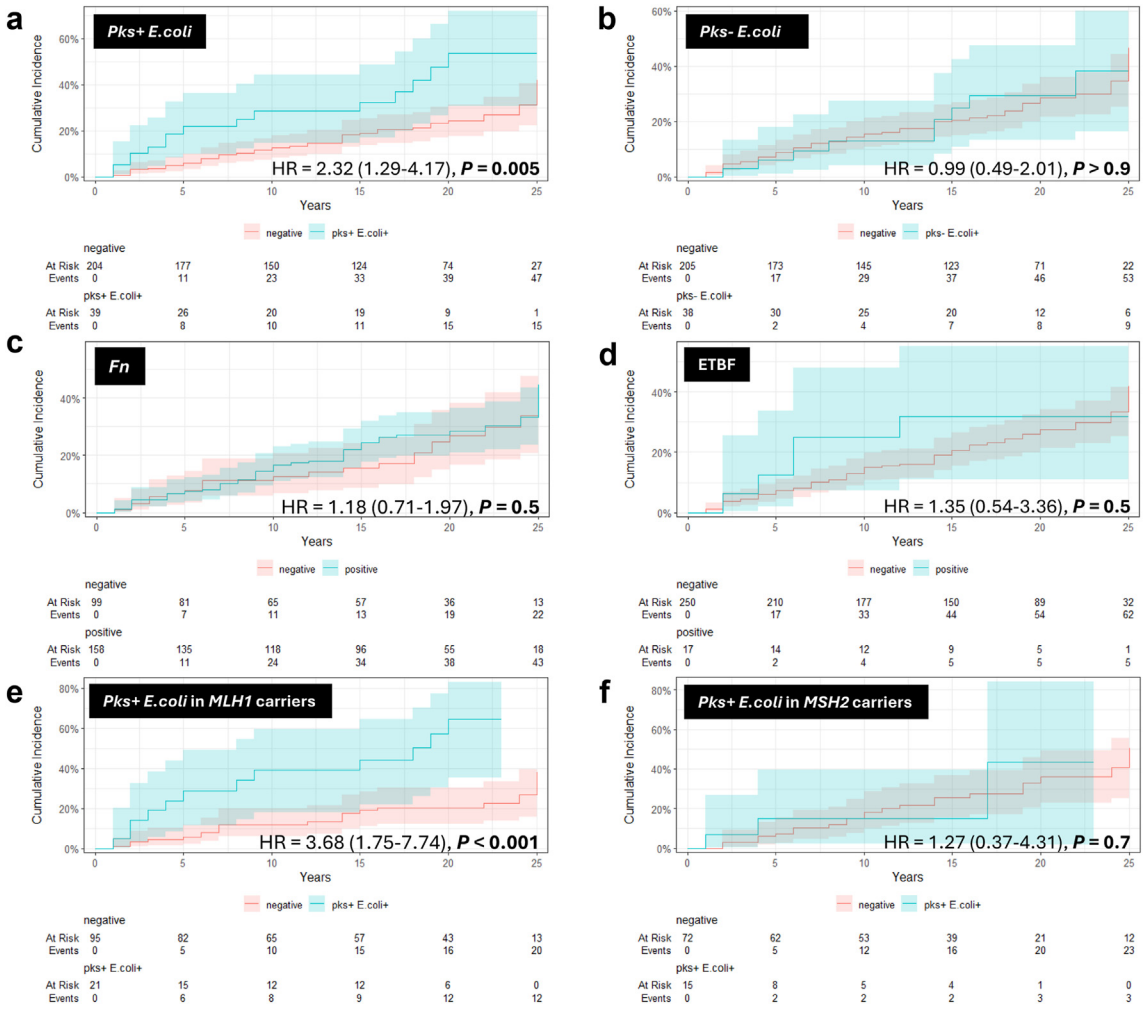

Fig. 1: The Kaplan–Meier plots showing the association between the presence of (a) *pks*$^{+}$ *E. coli*, (b) *pks*$^{-}$ *E. coli*, (c) *Fusobacterium nucleatum* (*Fn*), and (d) Enterotoxigenic *Bacteroides fragilis* (ETBF) with the cumulative risk of metachronous CRCs development. The association between *pks*$^{+}$ *E. coli* and metachronous cancer risk stratified by (e) *MLH1* and (f) *MSH2* genes is shown. The test was not adjusted.

Host genetics can influence the human gut microbiome composition.[35] In light of this, people with LS are shown to have increased immune cell infiltration in the colonic mucosa and this is suggested to contribute to the elevated risk of developing CRC.[36] The gut microbiome has an intricate relationship with the mucosal immune system, meaning that the microbiome composition can alter or be altered by the mucosa immune profiles.[37] People with LS have an increased but variable risk of developing CRC.[4,7] The affected MMR gene and sex are factors that contribute to the variability in CRC risks.[7,38] Modifiable risk factors including high body mass index at a young age,[39] being a current smoker[14] and excessive alcohol consumption[13] have also been shown to increase CRC risks in people with LS. However, there is limited understanding of the association between the gut microbiome and CRC development in LS. The findings from this study warrant further investigation of these genotoxic gut bacteria, and the broader microbiome, as potential modifiers of CRC and metachronous CRC risk in people with LS.

Our study revealed that CRCs from people with LS have significant enrichment of *Fn*, *pks*$^{+}$ *E. coli*, *pks*$^{-}$ *E. coli* but not ETBF when compared with sporadic MMRp CRCs. Although no prior studies have investigated intratumoural bacteria in LS, others have shown significant differences in the stool microbiome composition between people with and without LS who did not have cancer,[40–42] suggesting a potential connection between the genetic background and unique biology in LS and the stool microbiome composition. Other studies have reported differences in the gut microbiome between LS individuals with or without any cancer history,[43] suggesting a cancer diagnosis may directly or indirectly alter the gut microbiome.

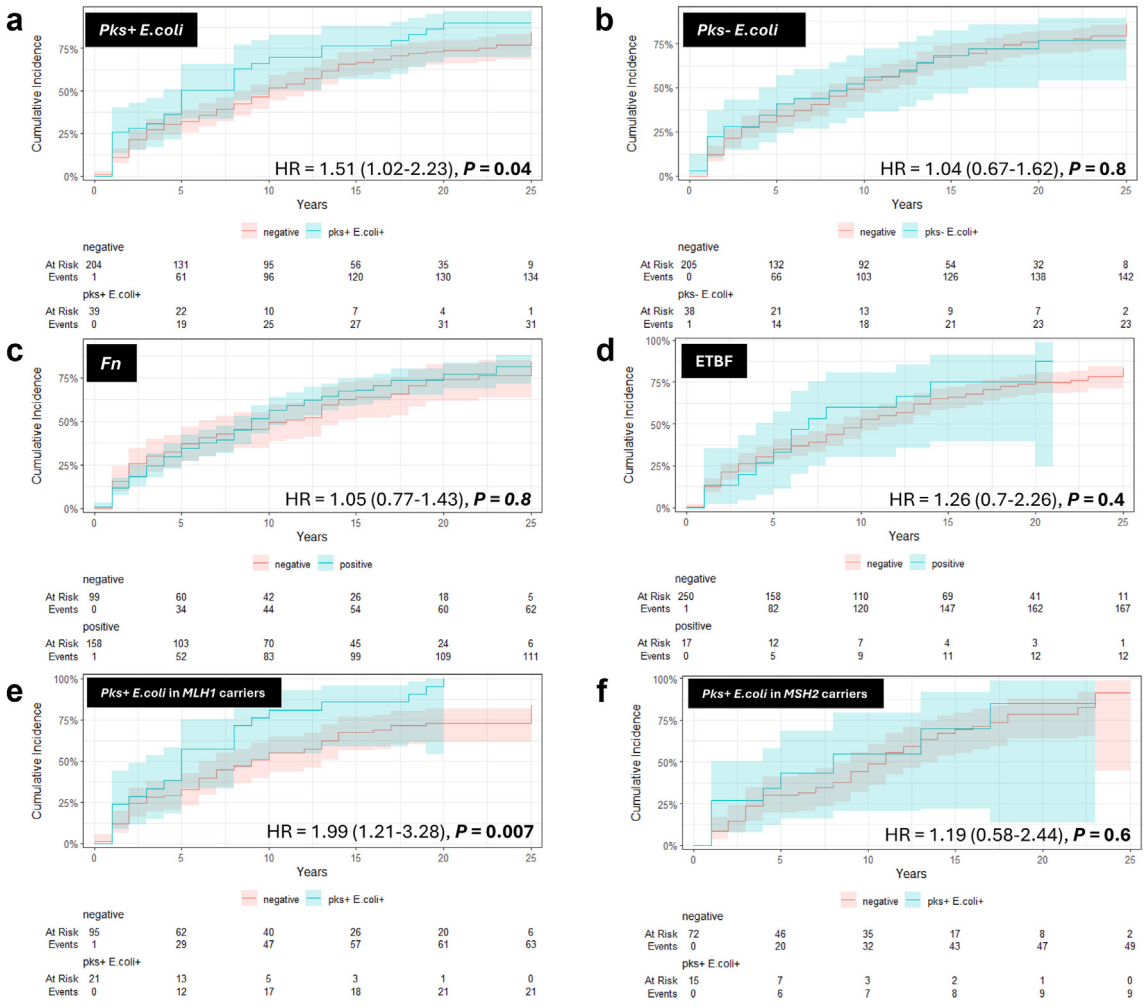

Fig. 2: The Kaplan–Meier plot showing the association between the presence of (a) *pks*$^{+}$ *E. coli*, (b) *pks*$^{-}$ *E. coli*, (c) *Fusobacterium nucleatum* (*Fn*), and (d) Enterotoxigenic *Bacteroides fragilis* (ETBF) with the cumulative risk of metachronous lesions development. We also showed the association between *pks*$^{+}$ *E. coli* and metachronous neoplasms risk stratified by (e) *MLH1* and (f) *MSH2* genes. The test was not adjusted.

We showed that CRCs in people with LS are >19 times more likely to harbour *Fn* than sporadic MMRp CRCs. In support of our findings, a previous study showed a significantly higher intratumoural *Fn* load in hereditary MSI CRCs (i.e., LS) compared to microsatellite stable CRCs.[44] Furthermore, the association between *Fn* and MMRd/microsatellite (MSI)-high CRC tumours of both hereditary and sporadic causes is well established.[22,45,46] Previous studies showed that *Fn* was associated with advanced disease stage[31] and metastasis[47] further expanding on a broader role of *Fn* in CRC tumourigenesis. Similar to previous studies,[48,49] we observed >4 times lower presence of *Fn* in adenomas compared with CRCs in people with LS. *Fn* has shown to be inversely correlated with CD3+ T-cell density[21] and has a dependent association with tumour infiltrating lymphocytes based on tumour MMR status,[22] further supporting that *Fn* could be an opportunistic bacterium found in specific highly immunogenic tumour microenvironment associated with MMRd and maybe less likely to play a role in initiating tumourigenesis.[50]

Our study showed that positive *pks*$^{+}$ *E. coli* in the initial/index CRC may increase the risks of metachronous CRC in people with LS. *Pks*$^{+}$ *E. coli* through the production of the genotoxin colibactin, can induce targeted DNA damage and double-stranded DNA breakage,[51,52] which is recognisable via tumour mutational signatures: single base substitution-88 (SBS88) and short insertions and deletions-18 (ID18).[53] We found that CRCs in people with LS are >1.5 times more likely to harbour *pks*$^{+}$ *E. coli* than sporadic MMRp CRCs and positive *pks*$^{+}$ *E. coli* in the initial CRC may increase the risk of metachronous CRCs and metachronous neoplasms. Infection with *pks*$^{+}$ *E. coli* or synthetic colibactin exacerbates MMRd-

associated mutations,[54] suggesting that colibactin from *pks*$^+$ *E. coli* may contribute to the mutation spectrum in MMRd cancers.

Our results suggest that *APC*: c.835-8A>G somatic mutation is not likely to be the main molecular mechanism for colibactin-induced CRC in LS. Previous studies showed that *APC*: c.835-8A>G fulfils the mutational context of colibactin-induced DNA damage, indicating that this hotspot mutation is a potential biomarker for CRCs caused by colibactin exposure.[55–58] Our previous study demonstrated a significant association between *pks*$^+$ *E. coli* and *APC*: c.835-8A>G somatic mutation in CRC samples.[22] This present study identified that *APC*: c.835-8A>G is rare (0.32%) in CRCs from people with LS, potentiating that pathogenesis may occur through other driver genes affected by colibactin DNA damage.

One plausible hypothesis is that colibactin-related DNA damage may increase the rate of a second somatic hit in the MMR gene affected by the constitutional pathogenic variant. Tumourigenesis through the MMRd pathway requires complete inactivation of both alleles of the MMR gene, where in LS, the normal allele is usually inactivated by a second somatic hit, leading to a complete silencing of the gene, which marks the initiation of MMRd tumour development. It is plausible that *pks*$^+$ *E. coli* through colibactin may cause DNA damage within the MMR gene and increase the rate of somatic mutations, leading to carcinogenesis via the MMRd pathway. The exposure to colibactin may create a "field defect" in the gut, therefore, positive *pks*$^+$ *E. coli* in the initial CRC may be the marker for this field defect and the risk of CRCs in the future. In support of this, our results showed that the NCM from the resection margin showed high concordance with the CRC for the presence of *pks*$^+$ *E. coli*. Previous studies utilising tumour prone animal models with a germline detect in *Apc* showed that colonic infection with *pks*$^+$ *E. coli* and ETBF accelerated the tumour onset and mortality.[59] Similarly, a study utilising an animal model with germline *Msh2* defect showed that colonic infection with ETBF induced loss of *Msh2* and cancer.[60] Future studies specifically assessing the colibactin-related somatic mutational context in the MMR genes could shed light on this. Whilst the findings from this study remain observational, future studies utilising colonic mucosa organoids derived from people with LS may be used to further elucidate the molecular mechanism of modulating colorectal cancer risks in LS.

It is widely hypothesised that the unique gut microbiome found in each of us is a likely outcome of the complex interplay between host genetic and environmental exposures.[61] Arima et al. identified that people exposed to diet that is high in red and processed meat and low in dietary fibre (so called "western diet") are more likely to develop CRCs with abundant *pks*$^+$ *E. coli*.[27] Similarly, recent studies highlight that the obesity-related gut microbiome could be the key carcinogenic mechanism of obesity-related CRC.[62] Although our study did not investigate the link between environmental exposures and the intratumoural gut bacteria in people with LS, we postulate that the presence of the genotoxic gut bacteria is tightly associated with environmental risk factors and these bacteria may serve as the biomarkers for early prevention of metachronous CRCs, especially in people with cancer predisposition syndromes. People with LS may benefit from the monitoring of *pks*$^+$ *E. coli* to modulate CRC risks.

The overall concordance of bacterial presence between CRC and both matching normal mucosa adjacent to the tumour and normal mucosa from resection margin suggests that the infection of these genotoxic bacteria is not specific to the tumour but a widespread colonic infection. This aligns with previous study where no significant differences were observed in the *pks*$^+$ *E. coli* copy numbers between tumour and normal mucosa.[63] This may support that widespread colonic infection of genotoxic gut bacteria such as *pks*$^+$ *E. coli* drives carcinogenesis and further creates the field defect to promote metachronous lesions development. Chen et al.[57] reported that the presence of *pks*$^+$ *E. coli* signatures were similar between matching tumour and NCM samples suggesting that *pks*$^+$ *E. coli* signatures are pervasive in the normal gut of CRC patients and the bacterial effect (such as *pks*$^+$ *E. coli*) from past exposure could be accumulated in cell genomes.

Our study has several strengths including a large sample size of people with LS from ACCFR that have extensive characterisation. The longitudinal design of ACCFR allowed us to obtain cancer and adenoma history, follow-up information on polypectomy, colonoscopy, and surgery in 5-yearly intervals. This enabled this first study investigating whether genotoxic gut bacteria as a risk factor for metachronous neoplasia in people with LS, which is clinically significant in managing the surveillance and CRC prevention program for people with LS. Preventing CRC and metachronous CRC development in people with LS is a recognised clinical challenge, that requires frequent (annual or biennial) colonoscopy surveillance. The potential to reduce the CRC risk in people with LS and offer less frequent colonoscopy surveillance by potentially eradicating *pks*$^+$ *E. coli* or altering the gut microbiome would be an effective way to reduce the burden on colonoscopy services and for the patient. Previous studies have mostly focused on the microbiome community within fecal samples and have been limited by a small number of people with LS studied ($n < 50$).[41,64] A further strength of our study was the use of multiplex assays that duplexed the *pks* island and *E. coli*, as well as ETBF and *Fn* together. This ensures the accurate differentiation of *pks*$^+$ *E. coli* from other *pks* harbouring bacteria whilst time- and cost-effective.

The limitations of this study include the lack of a control group of LS who did not develop CRCs, which limits our understanding on the presence of gut microbiome prior to tumourigenesis. To partly mitigate this, we utilised the bacterial results from our previous sporadic MMRp CRCs focused study and investigated the differences in bacteria prevalence between hereditary and sporadic caused CRCs. Further, we were unable to differentiate advanced from non-advanced adenomas for the metachronous neoplasia analysis due to the limited availability of this information. Despite this, the association between *pks*$^{+}$ *E. coli* and metachronous CRCs would have implications for future surveillance for both adenomas and CRCs if our findings are replicated. We also utilised FFPE specimens for detecting the presence of bacteria in people with LS. Although FFPE biospecimen has been widely used in large-scale cancer studies,[21,27] a meta-analysis showed that tissue type could influence the efficacy of the results.[65] However, Yamamura et al. systematically assessed the detection accuracies in FFPE and fresh frozen tissue samples and showed concordant results.[66] To also mitigate this, all assays used in this study had previously been developed for testing specifically in FFPE DNA. This study specifically targeted three genotoxic gut bacterial species with prior well-established proto-oncogenic role. Future studies investigating the broader microbiome community using 16S rRNA sequencing or metagenomics could offer a more comprehensive gut microbiome profile or identification of novel species linked to CRC development in people with LS.

In conclusion, this study provides novel findings of the association between the intratumoural presence of *pks*$^{+}$ *E. coli* in the initial/index CRC and the risk of metachronous CRCs and with colorectal neoplasia in people with LS. If these findings are validated, they would have potential clinical implications as *pks*$^{+}$ *E. coli* may represent a biomarker to test for in people with LS to understand their risk of metachronous neoplasms development. Future studies may focus on developing non-invasive biomarkers that could be useful for CRC risk stratification in people with Lynch syndrome. Further validation studies in independent CRC cohorts and/or functional studies (e.g., organoids models derived from people with LS) investigating the molecular mechanisms for explaining a potential causal relationship between *pks*$^{+}$ *E. coli* and CRC development in LS is needed before translating this into the clinical setting. The findings from this study represents an important step toward understanding risk modifiers of CRC in LS.

Contributors

YLC, JEJ, and DDB conceptualised the study. YLC, MC, RW, JC, and SGP performed laboratory experiments. YLC, JEJ, KM, and PG performed statistical analysis. SJ, TR, BML, RLM, MCS, GGG, AIP, JLH, AKW, FAM, IW, and MAJ contributed to the acquisition of study participants and data. CR oversaw the pathological reviews of tumour and polyp specimens. YLC, JEJ, and DDB wrote the manuscript. All authors read and provided critical revisions to the manuscript and approved the final manuscript.

Data sharing statement

The ACCFR data generated and/or analysed during the current study may be requested following approval of an Application for Collaboration to the Colon CFR (https://coloncfr.org/for-researchers/collaborate-with-the-ccfr/, last accessed 21/10/24).

Declaration of interests

The authors declare no competing interests.

Acknowledgements

We thank Dr. Christian Davey (Melbourne Statistical Consulting Platform and Statistical Consulting Center at the University of Melbourne) for guidance with the statistical aspects of this study. We thank Maggie Angelakos, Samantha Fox, Allyson Templeton for facilitating information and resources from the ACCFR. We thank Drs Danielle Ingle and Norelle Sherry from the Peter Doherty Institute for Infection and Immunity for providing the *pks*$^{+}$ *E. coli* strains. We thank Dr. Dianna M. Hocking and Prof. Roy Robins-Browne from the Department of Microbiology and Immunology of the University of Melbourne for providing the *pks*$^{-}$ *E. coli* strains.

Funding from an NHMRC Investigator grant (GNT1194896, Daniel D. Buchanan) and from a Cancer Australia/Cancer Council NSW co-funded grant (GNT2012914), supported the design, testing, analysis, and interpretation of data. Daniel D. Buchanan is further supported by a University of Melbourne Dame Kate Campbell Fellowship. Peter Georgeson is supported by an NHMRC Investigator Grant (GNT2026331). Romy Walker is supported by the University of Melbourne Early Career Researcher Grant. John L. Hopper is supported by the University of Melbourne Dame Kate Campbell Fellowship. Mark J. Jenkins is supported by an NHMRC Investigator grant (GNT1195099). Yen Lin Chu is supported by the Australian Government Research Training Program Scholarship, the Margaret and Irene Stewardson Fund Scholarship and the W. P. Greene Scholarship.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The content of this manuscript does not necessarily reflect the views or policies of the NIH or any of the collaborating centres in the CCFR, nor does mention of trade names, commercial products, or organisations imply endorsement by the US Government, any cancer registry, or the CCFR.

Melbourne Collaborative Cohort Study (MCCS) cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further augmented by Australian National Health and Medical Research Council grants 209057, 396414 and 1074383 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry and the Australian Institute of Health and Welfare, including the Australian Cancer Database.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.105661.

References

1 Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet*. 2009;76(1):1–18.
2 Buchanan DD, Clendenning M, Rosty C, et al. Tumor testing to identify lynch syndrome in two Australian colorectal cancer cohorts. *J Gastroenterol Hepatol*. 2017;32(2):427–438.
3 Ligtenberg MJ, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch

syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet*. 2009;41(1):112–117.
4 Dominguez-Valentin M, Sampson JR, Seppala TT, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. *Genet Med*. 2020;22(1):15–25.
5 Rubenstein JH, Enns R, Heidelbaugh J, Barkun A. American Gastroenterological Association Institute guideline on the diagnosis and management of Lynch syndrome. *Gastroenterology*. 2015;149(3):777–782. quiz e16-7.
6 van Leerdam ME, Roos VH, van Hooft JE, et al. Endoscopic management of Lynch syndrome and of familial risk of colorectal cancer: European Society of Gastrointestinal Endoscopy (ESGE) Guideline. *Endoscopy*. 2019;51(11):1082–1093.
7 International Mismatch Repair Consortium. Variation in the risk of colorectal cancer in families with Lynch syndrome: a retrospective cohort study. *Lancet Oncol*. 2021;22(7):1014–1022.
8 Bellido F, Guinó E, Jagmohan-Changur S, et al. Genetic variant in the telomerase gene modifies cancer risk in Lynch syndrome. *Eur J Hum Genet*. 2013;21(5):511–516.
9 Jenkins MA, Buchanan DD, Lai J, et al. Assessment of a polygenic risk score for colorectal cancer to predict risk of lynch syndrome colorectal cancer. *JNCI Cancer Spectr*. 2021;5(2):pkab022.
10 Dueñas N, Klinkhammer H, Bonifaci N, et al. Ability of a polygenic risk score to refine colorectal cancer risk in Lynch syndrome. *J Med Genet*. 2023;60(11):1044–1051.
11 Win AK, Clendenning M, Crawford W, et al. Genetic variants within the hTERT gene and the risk of colorectal cancer in Lynch syndrome. *Genes Cancer*. 2015;6(11–12):445–451.
12 Dashti SG, Win AK, Hardikar SS, et al. Physical activity and the risk of colorectal cancer in Lynch syndrome. *Int J Cancer*. 2018;143(9):2250–2260.
13 Dashti SG, Buchanan DD, Jayasekara H, et al. Alcohol consumption and the risk of colorectal cancer for mismatch repair gene mutation carriers. *Cancer Epidemiol Biomarkers Prev*. 2017;26(3):366–375.
14 Pande M, Lynch PM, Hopper JL, et al. Smoking and colorectal cancer in Lynch syndrome: results from the Colon Cancer Family Registry and the University of Texas M.D. Anderson Cancer Center. *Clin Cancer Res*. 2010;16(4):1331–1339.
15 Lindberg LJ, Wegen-Haitsma W, Ladelund S, et al. Risk of multiple colorectal cancer development depends on age and subgroup in individuals with hereditary predisposition. *Fam Cancer*. 2019;18(2):183–191.
16 Parry S, Win AK, Parry B, et al. Metachronous colorectal cancer risk for mismatch repair gene mutation carriers: the advantage of more extensive colon surgery. *Gut*. 2011;60(7):950–957.
17 Eikenboom EL, Moen S, van Leerdam ME, et al. Metachronous colorectal cancer risk according to Lynch syndrome pathogenic variant after extensive versus partial colectomy in the Netherlands: a retrospective cohort study. *Lancet Gastroenterol Hepatol*. 2023;8(12):1106–1117.
18 Nakatsu G, Li X, Zhou H, et al. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun*. 2015;6:8727.
19 Iyadorai T, Mariappan V, Vellasamy KM, et al. Prevalence and association of pks+ Escherichia coli with colorectal cancer in patients at the University Malaya Medical Centre, Malaysia. *PLoS One*. 2020;15(1):e0228217.
20 Zhou Y, He H, Xu H, et al. Association of oncogenic bacteria with colorectal cancer in South China. *Oncotarget*. 2016;7(49):80794–80802.
21 Mima K, Sukawa Y, Nishihara R, et al. Fusobacterium nucleatum and T cells in colorectal carcinoma. *JAMA Oncol*. 2015;1(5):653–661.
22 Joo JE, Chu YL, Georgeson P, et al. Intratumoral presence of the genotoxic gut bacteria pks(+) E. coli, Enterotoxigenic Bacteroides fragilis, and Fusobacterium nucleatum and their association with clinicopathological and molecular features of colorectal cancer. *Br J Cancer*. 2024;130(5):728–740.
23 Newcomb PA, Baron J, Cotterchio M, et al. Colon Cancer Family Registry: an international resource for studies of the genetic epidemiology of colon cancer. *Cancer Epidemiol Biomarkers Prev*. 2007;16(11):2331–2343.
24 Jenkins MA, Win AK, Templeton AS, et al. Cohort profile: the colon cancer family registry cohort (CCFRC). *Int J Epidemiol*. 2018;47(2):387–388i.
25 Milne RL, Fletcher AS, MacInnis RJ, et al. Cohort profile: the Melbourne collaborative cohort study (Health 2020). *Int J Epidemiol*. 2017;46(6):1757i.
26 Rosty C, Young JP, Walsh MD, et al. Colorectal carcinomas with KRAS mutation are associated with distinctive morphological and molecular features. *Mod Pathol*. 2013;26(6):825–834.
27 Arima K, Zhong R, Ugai T, et al. Western-style diet, pks island-carrying Escherichia coli, and colorectal cancer: analyses from two large prospective cohort studies. *Gastroenterology*. 2022;163(4):862–874.
28 Chern EC, Brenner KP, Wymer L, Haugland RA. Comparison of fecal indicator bacteria densities in marine recreational waters by QPCR. *Water Qual Expo Health*. 2009;1(3):203–214.
29 Abberton CL, Bereschenko L, van der Wielen PW, Smith CJ. Survival, biofilm formation, and growth potential of environmental and enteric Escherichia coli strains in drinking water microcosms. *Appl Environ Microbiol*. 2016;82(17):5320–5331.
30 Zamani S, Hesam Shariati S, Zali MR, et al. Detection of enterotoxigenic Bacteroides fragilis in patients with ulcerative colitis. *Gut Pathog*. 2017;9:53.
31 Castellarin M, Warren RL, Freeman JD, et al. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res*. 2012;22(2):299–306.
32 Li YY, Ge QX, Cao J, et al. Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. *World J Gastroenterol*. 2016;22(11):3227–3233.
33 Sherry NL, Lee RS, Gorrie CL, et al. Pilot study of a combined genomic and epidemiologic surveillance program for hospital-acquired multidrug-resistant pathogens across multiple hospital networks in Australia. *Infect Control Hosp Epidemiol*. 2021;42(5):573–581.
34 Sherry NL, Lee RS, Gorrie CL, et al. Genomic interrogation of the burden and transmission of multidrug-resistant pathogens within and across hospital networks. *bioRxiv*. 2019:764787.
35 Goodrich JK, Waters JL, Poole AC, et al. Human genetics shape the gut microbiome. *Cell*. 2014;159(4):789–799.
36 Bohaumilitzky L, Kluck K, Hüneburg R, et al. The different immune profiles of normal colonic mucosa in cancer-free lynch syndrome carriers and lynch syndrome colorectal cancer patients. *Gastroenterology*. 2022;162(3):907–919.e10.
37 Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science*. 2012;336(6086):1268–1273.
38 Møller P, Seppälä T, Bernstein I, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut*. 2017;66(3):464–472.
39 Win AK, Dowty JG, English DR, et al. Body mass index in early adulthood and colorectal cancer risk for carriers and non-carriers of germline mutations in DNA mismatch repair genes. *Br J Cancer*. 2011;105(1):162–169.
40 Ferrarese R, Zuppardo RA, Puzzono M, et al. Oral and fecal microbiota in lynch syndrome. *J Clin Med*. 2020;9(9):2735.
41 Rifkin SB, Sze MA, Tuck K, Koeppe E, Stoffel EM, Schloss PD. Gut microbiome composition in lynch syndrome with and without history of colorectal neoplasia and non-lynch controls. *J Gastrointest Cancer*. 2023;55(1):207–218.
42 Mori G, Orena BS, Cultrera I, et al. Gut microbiota analysis in postoperative lynch syndrome patients. *Front Microbiol*. 2019;10:1746.
43 Naddaf R, Carasso S, Reznick-Levi G, et al. Gut microbial signatures are associated with Lynch syndrome (LS) and cancer history in Druze communities in Israel. *Sci Rep*. 2023;13(1):20677.
44 Ono T, Yamaguchi T, Takao M, Kojika E, Iijima T, Horiguchi SI. Fusobacterium nucleatum load in MSI colorectal cancer subtypes. *Int J Clin Oncol*. 2022;27(10):1580–1588.
45 Hale VL, Jeraldo P, Chen J, et al. Distinct microbes, metabolites, and ecologies define the microbiome in deficient and proficient mismatch repair colorectal cancers. *Genome Med*. 2018;10(1):78.
46 Mima K, Nishihara R, Qian ZR, et al. Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut*. 2016;65(12):1973–1980.
47 Bullman S, Pedamallu CS, Sicinska E, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. *Science*. 2017;358(6369):1443–1448.
48 Yan Y, Drew DA, Markowitz A, et al. Structure of the mucosal and stool microbiome in lynch syndrome. *Cell Host Microbe*. 2020;27(4):585–600.e4.

49 Salim F, Mizutani S, Shiba S, et al. Fusobacterium species are distinctly associated with patients with Lynch syndrome colorectal cancer. *iScience*. 2024;27(7):110181.
50 Brennan CA, Garrett WS. Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. *Nat Rev Microbiol*. 2019;17(3):156–166.
51 Buc E, Dubois D, Sauvanet P, et al. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. *PLoS One*. 2013;8(2):e56964.
52 Bonnet M, Buc E, Sauvanet P, et al. Colonization of the human gut by E. coli and colorectal cancer risk. *Clin Cancer Res*. 2014;20(4):859–867.
53 Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, et al. Mutational signature in colorectal cancer caused by genotoxic pks(+) E. coli. *Nature*. 2020;580(7802):269–273.
54 Dougherty MW, Valdés-Mas R, Wernke KM, et al. The microbial genotoxin colibactin exacerbates mismatch repair mutations in colorectal tumors. *Neoplasia*. 2023;43:100918.
55 Terlouw D, Suerink M, Boot A, van Wezel T, Nielsen M, Morreau H. Recurrent APC splice variant c.835-8A>G in patients with unexplained colorectal polyposis fulfilling the colibactin mutational signature. *Gastroenterology*. 2020;159(4):1612–1614.e5.
56 Terlouw D, Boot A, Ducarmon QR, et al. Enrichment of colibactin-associated mutational signatures in unexplained colorectal polyposis patients. *BMC Cancer*. 2024;24(1):104.
57 Chen B, Ramazzotti D, Heide T, et al. Contribution of pks(+) E. coli mutations to colorectal carcinogenesis. *Nat Commun*. 2023;14(1):7827.
58 Lee-Six H, Olafsson S, Ellis P, et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature*. 2019;574(7779):532–537.
59 Dejea CM, Fathi P, Craig JM, et al. Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science*. 2018;359(6375):592–597.
60 Maiuri AR, Peng M, Podicheti R, et al. Mismatch repair proteins initiate epigenetic alterations during inflammation-driven tumorigenesis. *Cancer Res*. 2017;77(13):3467–3478.
61 Hall AB, Tolonen AC, Xavier RJ. Human genetic variation and the gut microbiome in disease. *Nat Rev Genet*. 2017;18(11):690–699.
62 Cani PD, Jordan BF. Gut microbiota-mediated inflammation in obesity: a link with gastrointestinal cancer. *Nat Rev Gastroenterol Hepatol*. 2018;15(11):671–682.
63 Miyasaka T, Yamada T, Uehara K, et al. Pks-positive Escherichia coli in tumor tissue and surrounding normal mucosal tissue of colorectal cancer patients. *Cancer Sci*. 2024;115(4):1184–1195.
64 Gonzalez A, Kapila N, Melendez-Rosado J, Liang H, Castro-Pavia F. An evaluation of the fecal microbiome in lynch syndrome. *J Gastrointest Cancer*. 2021;52(1):365–368.
65 Kim Y, Cho NY, Kang GH. Prognostic and clinicopathological significance of Fusobacterium nucleatum in colorectal cancer: a systemic review and meta-analysis. *J Pathol Transl Med*. 2022;56(3):144–151.
66 Yamamura K, Baba Y, Miyake K, et al. Fusobacterium nucleatum in gastroenterological cancer: evaluation of measurement methods using quantitative polymerase chain reaction and a literature review. *Oncol Lett*. 2017;14(6):6373–6378.