OPEN ACCESS

Original research

Gut microbiome signatures in colorectal neoplasia: a cross-sectional study across neoplasia stages and subtypes

Panayiotis Louca ,[1] Sarah Manning ,[2,3] Eleanor Hackney,[2,3] Linda Sharp ,[3] Mark A Hull ,[4] Sara Koo,[2] Gregory R Young ,[5] Guy S Taylor,[3] Yashvee Dunneram,[1] Suparna Mitra ,[4] James S Hampton,[2] Christina Dobson,[1,3] Laura J Neilson,[2,3] Caroline Addison,[6] Emad M El-Omar ,[7] the COLO-COHORT research team, Christopher J Stewart ,[5] Colin J Rees [2,3]

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/gutjnl-2025-337478).

For numbered affiliations see end of article.

Correspondence to
Professor Colin J Rees;
Colin.Rees@newcastle.ac.uk

PL and SM contributed equally.

Received 5 November 2025
Accepted 20 January 2026
Published Online First
6 March 2026

ABSTRACT

Background While colorectal cancer (CRC) has been linked to the gut microbiome, it remains unclear whether specific microbial signatures are detectable in precursor lesions such as adenomatous polyps, serrated lesions or sessile serrated lesions.

Objective To assess gut microbiome taxonomic and functional associations with colorectal neoplasia presence, severity (non-advanced, advanced and CRC) and subtype and evaluate predictive potential in high-risk neoplasia.

Design Analysed cross-sectional stool metagenomes (pre-colonoscopy) from 1762 participants (97% White British) undergoing colonoscopy in the multicentre COLO-COHORT study. Neoplasia was classified per British Society of Gastroenterology surveillance guidelines. Linear mixed-effects models and random forest classifiers assessed taxonomic and functional associations, adjusting for dietary, clinical and lifestyle covariates.

Results Gut microbiome composition differences between individuals with and without neoplasia were statistically significant but minimal (R^2=0.0008, p=0.03). A small number of species, including *Mediterraneibacter faecis* and *Pseudoruminococcus massiliensis*, and microbial pathways, including amino acid biosynthesis and β-lactam resistance, were modestly linked to neoplasia, particularly early lesions (q value <0.05). Associations were generally weak and attenuated after covariate adjustment. Predictive models combining the microbiome with clinical/demographic features modestly improved high-risk neoplasia classification (area under the curve=0.64 vs 0.58 for clinical/demographic features alone).

Conclusion This large prospective cross-sectional study found weak and inconsistent associations between the gut microbiome and premalignant colorectal neoplasia, with no robust microbial signatures. Findings suggest that previously reported microbial shifts may emerge later in disease progression, potentially as a consequence rather than a cause of CRC. Longitudinal, multiomic studies disentangling temporal and causal pathways between the gut microbiome and neoplasia are required.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The gut microbiome of patients with colorectal cancer (CRC) differs from that of healthy individuals, with taxa such as *Fusobacterium nucleatum* and *Parvimonas micra* enriched in CRC in multiple studies.

⇒ Limited evidence raises the question as to whether microbial profiles may vary between those with and without colorectal polyps and across polyp types (adenomatous, serrated and sessile serrated), but studies are limited by small sample sizes, inconsistent polyp classifications and inadequate adjustment for confounders.

⇒ The role of the microbiome in precancerous lesions therefore remains unclear.

INTRODUCTION

Colorectal cancer (CRC) remains a leading cause of cancer morbidity and mortality, with ~1.9 million new cases annually.[1] Most CRCs develop through two main pathways. The most common is the adenoma-carcinoma sequence, which involves progression from adenomatous polyps (APs). A second pathway involves progression from sessile serrated lesions (SSLs) and other serrated lesions (SLs) and is estimated to account for up to 30% of CRC cases.[2–4] Advanced neoplasia, which includes larger polyps (≥10 mm) and/or those with dysplasia, is associated with higher malignant potential and increased CRC risk, distinguishing it from non-advanced neoplasia.[5]

The molecular and environmental drivers of colorectal neoplasia remain incompletely understood, limiting potential for early detection and precise risk stratification. The gut microbiome, shaped by diet, lifestyle and medication, modulates host immunity, metabolism and inflammation, all relevant to CRC pathogenesis.[6]

In our recent systematic review of shotgun metagenomic studies across 22 populations, we identified reproducible microbial signatures in CRC, particularly enrichment of *Fusobacterium nucleatum* and *Parvimonas micra*, reported in 35%–38% of the

Check for updates

© Author(s) (or their employer(s)) 2026. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

To cite: Louca P, Manning S, Hackney E, *et al*. *Gut* 2026;**75**:2118–2129.

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

WHAT THIS STUDY ADDS

⇒ This is the largest metagenomic study to date (n=1762) examining microbial associations of precancerous lesions across the spectrum of neoplasia, using standardised British Society of Gastroenterology surveillance classifications and comprehensive adjustment for dietary, clinical and lifestyle confounders.

⇒ Despite rigorous methodology and a large sample size, no robust microbial signature of precancerous lesions was identified, with most associations weak and attenuated after confounder adjustment.

⇒ Novel associations with *Mediterraneibacter faecis* and *Pseudoruminococcus massiliensis* were identified but showed small effect sizes and limited clinical significance.

⇒ Predictive models incorporating microbiome data showed only marginal improvement over clinical/demographic factors alone, suggesting limited utility of microbiome profiles in detecting high-risk neoplasia.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These results challenge the idea that the gut microbiome can reliably indicate early colorectal neoplastic changes and caution against overinterpreting microbiome signals in polyp detection or risk stratification.

⇒ Future work should prioritise longitudinal, multiomics studies with harmonised protocols to determine whether consistent microbial shifts emerge in lesions that do progress to malignancy.

26 studies, respectively.[7] Findings for precancerous polyps were inconsistent, and only one of the 26 studies stratified microbial associations by both adenomatous and serrated subtypes.[8] These studies had small sample sizes and substantial methodological heterogeneity, including inconsistent polyp definitions, inadequate confounder adjustment and non-standardised protocols.[7] It therefore remains unclear when and where in the pathway from normal mucosa to CRC changes in the microbiome may occur, thus, hampering the development of molecular approaches to early detection.

To address this gap, we generated and analysed gut microbiome profiles from 1762 participants recruited to the COLO-COHORT study,[9] applying standardised British Society of Gastroenterology (BSG) surveillance classifications for neoplastic lesions[4] and adjusting robustly for dietary, clinical and lifestyle confounders. Using shotgun metagenomic sequencing, we aimed to (1) characterise microbial differences between individuals with and without colorectal neoplasia, (2) explore variation across neoplasia stages (non-advanced polyps, advanced polyps and CRC) and neoplastic subtypes (APs, SLs and SSLs) and (3) evaluate the predictive potential of microbiome profiles for identifying high-risk neoplasia. This study sought to uncover taxonomic and functional markers relevant to early CRC development and assess their potential for improving early detection.

METHODS

Study sample

This analysis included the first 1762 participants aged ≥30 years from the ongoing COLO-COHORT study,[9] who provided pre-bowel preparation stool samples suitable for shotgun metagenomic sequencing along with concurrent clinical, dietary and lifestyle data. COLO-COHORT recruits adults undergoing colonoscopy across 32 UK National Health Service (NHS) sites, with referrals via the Bowel Cancer Screening Programme (BCSP), symptomatic pathways (investigation of rectal bleeding, change in bowel habit, iron deficiency anaemia, rectal or abdominal mass, positive faecal immunochemical test (FIT), unexplained weight loss, abdominal pain or abnormal imaging) or routine surveillance (familial history, post-CRC surveillance or polyp surveillance). Participants may choose different levels of consent: some provide only permission for future contact, others contribute limited clinical data, and a subset (Group A1) consent to full phenotyping and biospecimen collection, including pre-bowel preparation stool and blood samples. The present microbiome analysis is restricted to Group A1 participants who provided suitable stool samples; those consenting to less intensive levels were not included. Further details on consent options and recruitment are provided in the published protocol.[9]

Further details on clinical data collection and dietary assessment are provided in the online supplemental methods. A schematic of the study design and analysis workflow is presented in figure 1.

Neoplasia classification

Colorectal neoplasia was determined by histopathological evaluation of resected polyps performed as part of routine NHS clinical pathology. Lesions were categorised as APs (tubular, tubulovillous or villous adenomas) or SLs (SSLs, traditional serrated adenomas, mixed polyps or hyperplastic polyps, excluding small (<5 mm) rectal hyperplastic polyps). Advanced polyps were defined using BSG surveillance guidelines[4] as APs ≥10 mm or with high-grade dysplasia, or SLs ≥10 mm or with any dysplasia.

Outcome measures

We examined gut microbiome differences across three analytical frameworks (figure 1), with participant classifications described in online supplemental table 2:

Comparison group

Participants with no polyps, either neoplastic or non-neoplastic (eg, inflammatory) or cancer detected during colonoscopy, comprised the no neoplasia group, which served as the reference population for all comparative analyses across the three analytical frameworks.

Neoplasia presence

Participants with neoplastic polyps or CRC detected comprised the neoplasia group. The neoplasia group was further stratified by exclusively left-sided (splenic flexure—rectum) or right-sided lesions (caecum—transverse colon) colonic location.

Neoplasia subtype and stages

Microbiome differences were assessed by polyp subtype (APs, SLs (including SSLs) and SSLs) and neoplasia stage (cancer, advanced polyps excluding CRC and non-advanced polyps) using participants with exclusively one lesion type/stage to investigate association specificity.

Clinical utility

Participants were categorised as high-risk per BSG guidelines[4] (≥2 premalignant polyps with ≥1 advanced polyp, ≥5 premalignant polyps or ≥1 large non-pedunculated polyp (>20 mm with specific Paris morphology[10])). Non-high-risk participants were 1:1 matched on age, sex and body mass index (BMI) to test microbiome predictive capacity for high-risk classification.

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

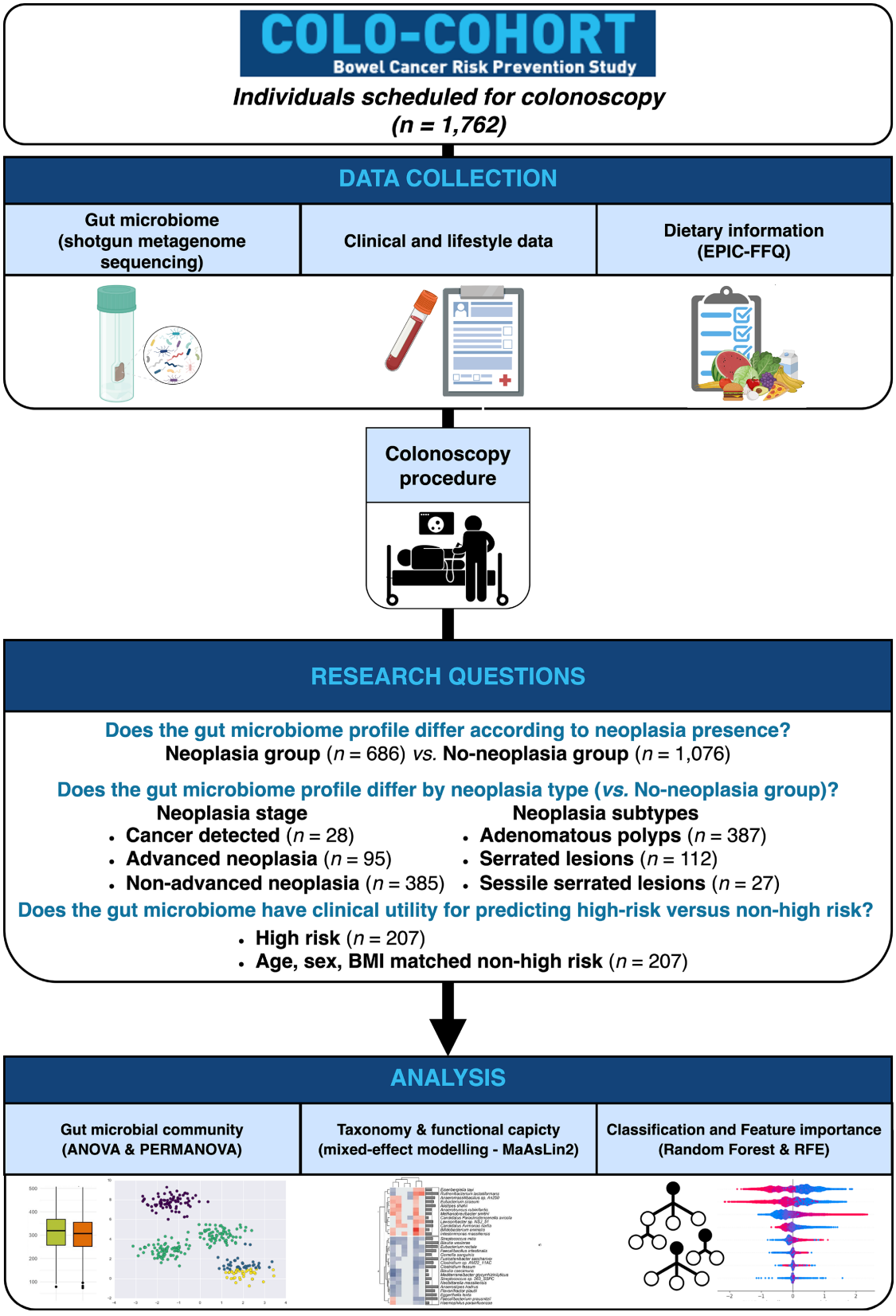


Figure 1 Schematic representation of the study design, data collection and analysis workflow. A total of 1762 participants undergoing colonoscopy, including patients from both BCSP and non-BCSP pathways, were selected. Available data included gut microbiome profiles generated through shotgun metagenome sequencing, dietary information via the EPIC-FFQ and clinical and lifestyle data. The analysis pipeline explored within- and between-participant differences in gut microbial diversity and their correlations with colorectal neoplasia outcomes and included mixed-effect modelling (MaAsLin2) and random forest modelling for both classification and feature importance. ANOVA, analysis of variance; BCSP, bowel cancer screening programme; EPIC-FFQ, European Prospective Investigation into Cancer and Nutrition-food frequency questionnaire; PERMANOVA, permutational multivariate analysis of variance

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Faecal sample collection and gut metagenome sequencing

Stool samples were collected prior to bowel preparation using a validated FIT (OC Sensor, Mast Diagnostics)[11 12] and subjected to shotgun metagenomic sequencing (Illumina NovaSeq6000) by Novogene. DNA extraction, sequencing and quality control followed standardised protocols, with details provided in online supplemental methods.

Alpha diversity

We calculated alpha diversity metrics at the species level using 'vegan'.[13] Observed taxonomic richness represented the count of unique species detected, and the Shannon Diversity Index[14] represented per-sample species dominance.

Beta diversity

To evaluate between-sample diversity, we computed Bray-Curtis dissimilarities[15] as a measure of dissimilarity between samples based on species relative abundance and Jaccard distances[16] as a measure of dissimilarity based on species presence or absence.

Statistical analysis

All statistical analyses were performed using R V.4.3.1.[17]

Imputation

Missing covariate data were imputed using k-nearest neighbours (k=9).[18] Missingness ranged from 0.06% (n=1; sex) to 2.4% (n=43; waist circumference). Imputation used only non-microbiome, non-outcome variables to avoid bias (online supplemental table 3).

Statistical modelling approaches

Three complementary strategies were used to investigate associations with neoplasia outcomes. The base and a priori models were fitted for all analyses (α-diversity, β-diversity, taxonomic abundances and functional pathways), whereas the data-driven model was applied only to taxonomic analyses and when additional non-a priori features were identified.

- The base model (age, sex, BMI and ethnicity, recruitment site and sequencing batch) was applied to all analyses for comparability, and results are presented in figures and online supplemental material.
- The a priori model extended the base model with prespecified dietary (energy intake (kcal/day), healthy diet score (HDS)), lifestyle (alcohol intake (g/day) and pack-years of smoking) and medication (proton pump inhibitor (PPI) and aspirin use).
- A data-driven model was applied only to taxonomic analyses when additional, stable non-microbial covariates were identified through random forest recursive feature elimination (RFE), extending the a priori model (see online supplemental methods for thresholds and parameters).

Diversity analyses

α-diversity differences were assessed using analysis of variance (ANOVA) and mixed-effects regression (accounting for batch and site as fixed effects).

β-diversity was evaluated using permutational multivariate analysis of variance (PERMANOVA), implemented in *adonis2*,[13] with 1000 permutations, treating recruitment site as a blocking factor. Homogeneity of dispersion was tested using permdisp.

Microbial taxa associations

Taxa present in ≥5% of samples were retained, and centre-log ratio normalised. Mixed-effects regression models were fitted using *MaAsLin2*,[19] at species, genus and phylum levels with Benjamini-Hochberg false discovery rate (FDR) <0.1 considered significant.

To evaluate robustness to modelling approaches, we further performed ANCOM-BC2 (V.2.6.0[20]) for the main outcome, neoplasia presence and on count data of species.

Functional microbiome analysis

Functional profiling was conducted using HUMAnN V.3.0. Gene family abundance tables were filtered to microbial enzymes annotated with enzyme commission numbers, based on UniRef90[21] mappings from UniProt.[22] We analysed 132 794 microbial gene families using linear mixed-effects models.

Metabolic pathway abundance, expressed as the number of complete copies of each pathway in the community, was filtered to exclude microbially stratified abundance values. We analysed 462 microbial metabolic pathways using the same approach as gene families.

Predictive modelling for clinical utility

Random forest models with fivefold cross-validation distinguished high-risk cases from matched controls, using all available covariates, all gut microbial species or a combined model. Matching, model parameters and feature selection details are provided in the online supplemental methods.

Subgroup analyses and sensitivity analyses are described in the online supplemental methods.

RESULTS

Study population characteristics

This large study recruited all patients using the same protocol and included 1762 participants with gut microbiome profiles assessed by shotgun metagenomics (mean (SD): age 60.9 (12.1) years, BMI 28.7 (6.0) kg/m^2, 51.4% female) (table 1). Participants were representative of the overall UK colonoscopy population (50.4% female, median age 61 years (IQR 50–71)).[23] Of the participants, 61.1% (n=1076) had no polyps or cancer detected (the no-neoplasia group). Participants were predominantly of White British ethnicity (97%).

Colorectal neoplasia was detected in 686 individuals (38.9%, neoplasia group). Among these, 28 individuals had CRC only, 95 had only advanced neoplastic polyps and 385 had non-advanced neoplastic polyps only. Others had a mixture of polyps. By subtype, 387 participants had APs only, 112 SLs only and 27 SSLs only. Overall, 207 participants met BSG high-risk criteria.[4] See online supplemental tables 4 and 5 for demographics by neoplasia stage and subtype.

Compared to the no-neoplasia group, the neoplasia group was significantly (p<0.01) older (63.5 vs 59.2 years), more often male (57.7% vs 42.8%) and had a higher BMI (29.2 vs 28.3 kg/m^2), greater smoking exposure (50.3 vs 36.4 pack-years), lower fibre intake (14.6 vs 15.3 g), lower HDS (5.08 vs 5.46) and was more likely to have a history of polyps (29.7% vs 20.7%) (table 1). No significant differences were observed for family history of CRC, PPI use or aspirin use.

Microbial differences between individuals with and without any neoplasia

We first examined whether colorectal neoplasia was associated with broad microbial diversity.

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 Demographic and clinical characteristics of study participants

	Overall (n=1762)		No-neoplasia group (n=1076)		Neoplasia group (n=686)		
	n	%	n	%	n	%	
Sex							*
Male	856	48.58	460	42.75	396	57.73	
Ethnicity							
White British	1709	97.0	1038	96.5	671	97.8	
Black British	5	0.3	5	0.5	0	0	
Asian/Asian British	17	1.0	10	0.9	7	1.0	
Mixed	11	0.6	8	0.7	3	0.4	
Other	15	0.9	13	1.2	2	0.3	
Family history of colorectal cancer	478	27.1	300	27.9	178	25.95	
Referral pathway							*
BCSP	120	6.8	32	2.97	88	12.8	
Non-BCSP	1620	91.9	1028	95.5	592	86.3	
Medications							
Proton pump inhibitor user	640	36.3	407	37.8	233	34.0	
Aspirin user	207	11.8	117	10.9	90	13.1	
Previous polyps detected	427	24.2	223	20.7	204	29.7	*
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	
Age (years)	60.89	12.1	59.2	12.7	63.5	10.6	*
Body mass index (kg/m^2)	28.66	6.0	28.3	6.2	29.2	5.8	*
Healthy diet score	5.31	1.8	5.46	1.8	5.1	1.8	*
Dietary fibre intake (g)	15.1	7.0	15.3	6.7	14.6	7.3	*
Pack years of smoking	41.8	60.1	36.4	57.4	50.3	64.5	*
Alpha diversity							
Shannon diversity	3.8	0.4	3.75	0.4	3.8	0.4	
Taxonomic richness	199.5	57.0	197.5	58.7	202.7	54.0	

Characteristics were measured prior to imputation and shown overall and stratified by neoplasia presence. Missingness is reported in online supplemental table 3.
Baseline differences between the neoplasia group and the no-neoplasia group were assessed using χ^2 or Fisher's exact tests for categorical variables, and Mann-Whitney U tests for continuous variables, as appropriate.
*Indicates significant (p<0.01) differences.
BCSP, bowel cancer screening programme; PPI, proton pump inhibitor.;

Alpha diversity metrics (Shannon: 3.79 vs 3.75, p=0.58; richness: 202.7 vs 197.5, p=0.63) were comparable between the neoplasia group and the no-neoplasia group across all models (online supplemental table 6), indicating that neoplasia presence was not associated with major differences in alpha diversity.

We next assessed overall microbial community composition. PERMANOVA revealed weak but statistically significant differences with neoplasia presence (Bray-Curtis R^2=0.0008, p=0.03; Jaccard R^2=0.0007, p=0.03) after adjustment for the a priori model (figure 2A; online supplemental table 7). However, dispersion testing demonstrated significantly greater within-group variability among participants with neoplasia (Bray-Curtis: F=11.1, p=0.002, Jaccard: F=10.5, p=9.99×10^{-4}). Principal coordinates analysis based on Bray-Curtis distances revealed no distinct clustering by neoplasia status, with the first two axes explaining 7.2% and 6.5% of variance, respectively.

To identify stable predictors of neoplasia presence, we applied RFE using random forest classifiers across 100 iterations with random seeds. Of 430 features selected at least once, 30 were consistently identified in ≥70% of models (figure 2B). These included 18 microbial species and 12 non-microbial covariates, seven of which (waist circumference, body weight, red meat intake, statin use, history of polyps, employment status and height) were not part of our a priori models. These were incorporated into the a priori models to evaluate their influence on taxa-neoplasia associations. The most important features identified at the genus level are shown in online supplemental figure 1.

We then used linear mixed-effects models to identify taxa associated with neoplasia presence. Three bacterial species were associated with neoplasia presence after adjusting for a priori covariates and multiple testing (FDR<0.1): *Mediterraneibacter faecis* (p=1.18×10^{-3}) and *Pseudoruminococcus massiliensis* (p=2.19×10^{-3}) were positively associated, while an uncultured species, GGB3475 SGB4638, was negatively associated (p=1.87×10^{-3}) (figure 2C, online supplemental table 8). However, these associations were no longer significant after adjusting for additional RFE-identified covariates.

At the genus level, *Pseudoruminococcus* was positively associated with neoplasia in the a priori model (p=2.15×10^{-3}, FDR=0.08), but not in the data-driven model (p=2.11×10^{-3}, FDR=0.13) (online supplemental table 8). No phylum-level associations met the FDR threshold in any model. Plots illustrating relative abundances and prevalence rates for key species/genera are shown in online supplemental figure 2.

In addition to MaAsLin2, we applied ANCOM-BC2, a compositional bias-correcting linear mixed model that estimates log-fold changes (LFC), for the primary comparison (neoplasia vs no neoplasia). Using the same covariate sets and random effects (site, batch), ANCOM-BC2 identified *P. massiliensis* as significantly enriched in participants with neoplasia (a priori model: LogFoldChange (LFC)=1.22, SE=0.29, p=3.29×10^{-5}, FDR=0.023; data-driven model: LFC=1.28, SE=0.29, p=2.21×10^{-5}, FDR=0.015). Additionally, the *Pseudoruminococcus* genus was also significantly upregulated in the neoplasia

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

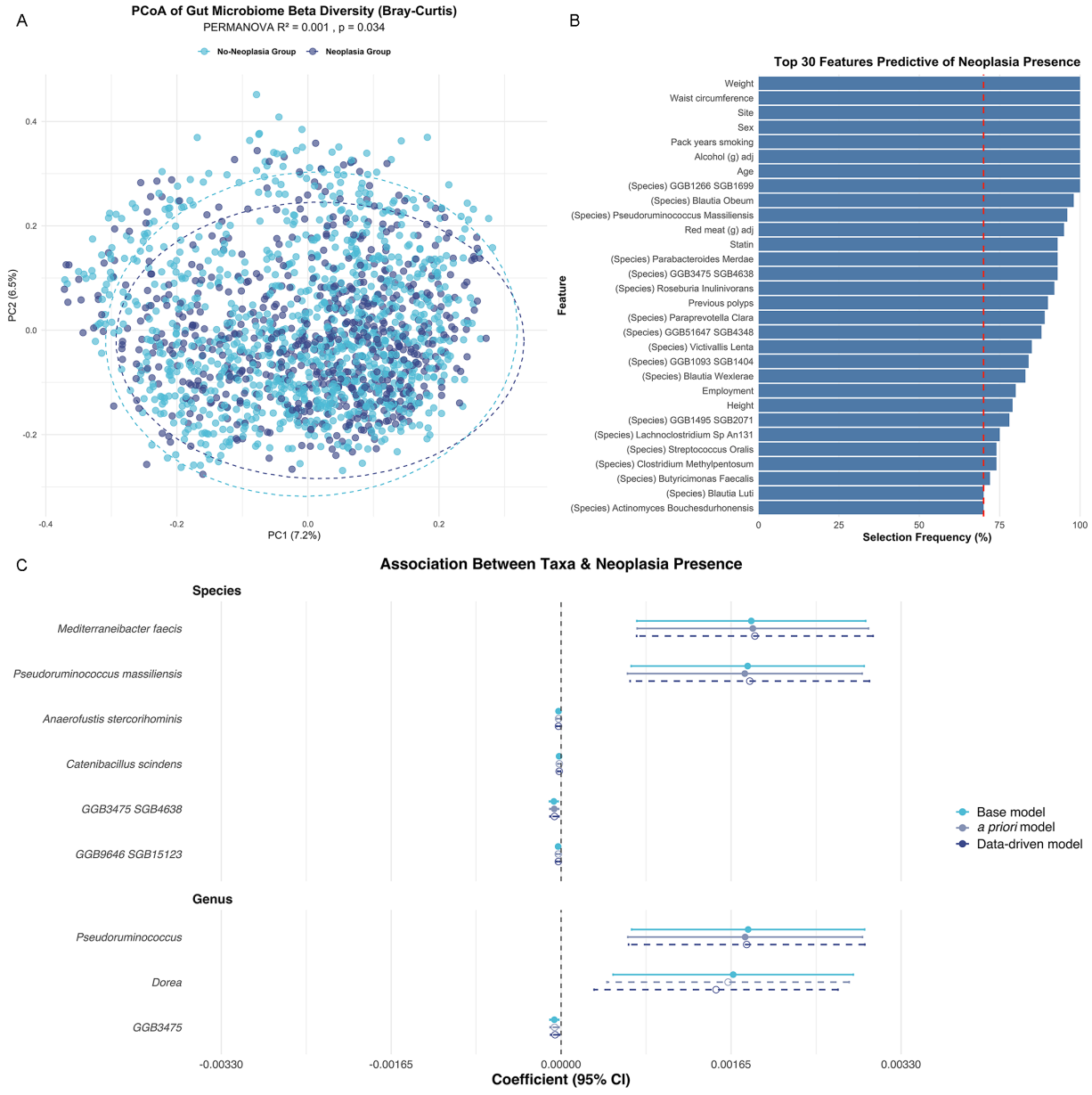


Figure 2 Microbiome community structure and taxonomic features associated with colorectal neoplasia. (A) PCoA based on Bray-Curtis dissimilarity, coloured by neoplasia status. Each point represents an individual sample; subtle separation reflects minor global differences in community composition. PERMANOVA R^2 and p value reflect the a priori model. Reported R^2 is rounded to three decimal places. Points are coloured by neoplasia status (dark blue=neoplasia group; light blue=no-neoplasia group). (B) Bar plot of the top 30 most stable features discriminating neoplasia presence identified using recursive feature elimination across 100 random seeds. Feature stability represents the proportion of times a feature was selected across iterations; the red dashed line indicates the 70% stability threshold. (C) Forest plot of neoplasia-associated species and genera identified using linear mixed-effects models. Points represent model-estimated beta coefficients, and error bars indicate 95% CIs. Features not passing the false discovery rate <0.1 threshold are displayed with reduced opacity. PCoA, principal coordinates analysis; PC, principal component; PERMANOVA, permutational multivariate analysis of variance.

group. No phylum-level associations passed multiple testing. ANCOM-BC2, therefore, supports the association of *P. massiliensis* and *Pseudoruminococcus* with neoplasia under a compositional count-based modelling framework; ANCOM-BC2 results are provided in online supplemental table 9.

Model estimates (base model) for microbial species with robust prior evidence linking them to colorectal neoplasia[7] are presented in online supplemental table 10. Although, of note, many were observed at low prevalence, consistent with prior observations that these species are enriched primarily in later-stage or invasive disease.

Analysis of microbial functional potential identified five metabolic pathways significantly associated with neoplasia presence (online supplemental figure 3). Three pathways were positively associated: L-glutamate and L-glutamine biosynthesis (PWY-5505), tRNA processing (PWY0-1479) and peptidoglycan biosynthesis V (β-lactam resistance; PWY-6470). Two pathways were negatively associated: UTP and CTP dephosphorylation I (PWY-7185) and pyrimidine deoxyribonucleotide biosynthesis from CTP (PWY-7210). Additionally, 2674 UniRef90 gene families with enzymatic capacity were significantly associated with neoplasia (FDR<0.1), suggesting potential differences in

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

the functional potential of the microbiome in the presence of neoplasia (online supplemental figure 4).

Subgroup analysis by neoplasia location

To explore the potential influence of polyp location, we conducted a subgroup analysis comparing left-sided (n=251) and right-sided (n=226) neoplasia.

No differences in alpha or beta diversity were observed between these groups (online supplemental tables 11 and 12); all p>0.05).

At the taxonomic level, three species, including *P. massiliensis*, were associated with right-sided neoplasia, and two, including *Streptococcus cristatus*, with left-sided neoplasia (a priori model, all FDR<0.1). At the genus level, *Pseudoruminococcus* and an uncultured genus (GGB3000) were associated with right- and left-sided neoplasia, respectively (online supplemental table 13; both FDR<0.1). Similar to the overall neoplasia analysis, no phyla passed the multiple testing threshold.

Subgroup analysis by patient referral route

We next examined whether microbial associations varied by clinical presentation by performing a stratified analysis based on referral pathway: BCSP versus non-BCSP.

Alpha diversity did not differ significantly between the neoplasia group and the no-neoplasia group within either stratum (online supplemental table 14). When investigating beta diversity differences, modest but statistically significant differences were identified in BCSP participants using both Bray-Curtis and Jaccard distance metrics (online supplemental table 15). In contrast, no differences were found in non-BCSP participants.

Taxonomic analyses revealed stratum-specific associations. Among BCSP participants, no microbial species reached significance in the a priori model after multiple testing correction. In contrast, two species, *P. massiliensis and M. faecis*, were positively associated in the non-BCSP stratum (online supplemental figure 5). At the genus level, no genera remained significant in either stratum.

Differences across neoplasia stages and specific neoplasia subtypes

No differences in Shannon diversity or taxonomic richness were detected across disease stages (non-advanced, advanced or CRC) or polyp subtypes (APs, SLs and SSLs) in either ANOVA or mixed-effects models (online supplemental tables 16 and 17), indicating that alpha diversity does not differ by neoplasia subtype or disease stage.

However, beta diversity differences were identified among individuals with non-advanced neoplasia. PERMANOVA analyses showed significant but small compositional differences between individuals with non-advanced neoplasia and the no-neoplasia group using Bray-Curtis (R^2=0.00092, p=0.046) and Jaccard (R^2=0.00083, p=0.049) distance metrics (online supplemental table 18). No significant beta diversity differences were detected for cancer, advanced neoplasia or specific polyp subtypes. These results suggest that microbiome compositional changes are more pronounced in early, non-advanced lesions, with limited evidence of consistent community-level differences in advanced or specific neoplasia subtypes.

Features for each neoplasia stage and polyp subtype are presented in online supplemental figures 6 and 7, respectively.

Linear mixed-effects models identified multiple species and genera associated with neoplasia types after covariate adjustment and FDR correction (figure 3). For cancer-only cases, nine bacterial species and five genera were significantly associated, with the strongest effects for *Faecalitalea cylindroides* and the genus *Faecalitalea* (figure 3A,B). No phylum-level associations passed FDR correction. Among those with advanced neoplasia (excluding cancer), six species were associated (figure 3C) and two, *Streptococcus thermophilus* and *Roseburia* sp AM59-24XD, remained significant in the data-driven model. At the genus level, *Prevotella* was consistently associated (figure 3D), although this association did not persist when cancer cases were included in the advanced neoplasia group (figure 3E,F). In non-advanced neoplasia, four species and two genera were identified, with *M. faecis* and *Pseudoruminococcus* remaining significant in the data-driven model (figure 3G,H).

Analyses of polyp subtypes also revealed distinct microbial signals. *P. massiliensis* and *Pseudoruminococcus* were positively associated with APs in the a priori model ($p=1.95\times10^{-3}$, respectively) but were not significantly associated when controlling for data-driven features (figure 4A,B). For SLs, *Gordonibacter urolithinfaciens* and *P. massiliensis* were associated in the a priori model ($p=5.17\times10^{-4}$, $p=2.17\times10^{-3}$), but not in the data-driven model (figure 4C,D). For SSLs, 10 species and 7 genera (six uncultured and one unclassified *Eubacteriaceae*) were associated (FDR<0.1) (figure 4E,F). No phylum-level associations met the FDR threshold. These findings highlight distinct microbial signatures across neoplasia stages and subtypes, with persistent associations for *Prevotella* in advanced neoplasia and multiple taxa in SSLs, suggesting microbial contributions to colorectal neoplasia progression.

Functional pathway analysis revealed distinct microbial profiles across disease trajectories and polyp subtypes (online supplemental figures 8 and 9). Nine pathways were differentially abundant in non-advanced neoplasia (online supplemental figure 8), with the strongest association for phospholipases. For polyp subtypes, three pathways were associated with APs (two negative and one positive), five with SLs (three negative and two positive), the strongest significant effect was observed for guanosine nucleotide degradation and two with SSLs (both negative (glucose and glucose 1 phosphate degradation and flavin biosynthesis III)) (online supplemental figure 9). Additionally, differentially abundant UniRef90 gene families were detected across neoplasia disease trajectories: 488 in cancer-only, 254 in advanced neoplasia and 1795 in non-advanced neoplasia (online supplemental figure 10), and 859, 125 and 248 were associated with APs, SLs and SSL-only, respectively (online supplemental figure 11). These findings suggest functional microbial alterations may play a greater role in early lesion development.

Predictive utility of the gut microbiome

We evaluated gut microbiome prediction of high-risk neoplasia using random forest models (figure 5). The microbiome-only model achieved a mean area under the curve (AUC) of 0.621 (range 0.582–0.663), a negligible improvement versus the covariate-only model, AUC of 0.582 (range 0.453–0.684). The combined model demonstrated the highest predictive performance, with a mean AUC of 0.639 (range 0.563–0.703). In the microbiome-only model, top-ranking predictive species included *Odoribacter splanchnicus*, *Roseburia inulinivorans* and *Ruminococcus torques*. Other key predictors included participant referral route, dietary protein, alcohol intake and red meat consumption. In the combined model, referral route (BCSP or non-BCSP) remained the most influential feature, but several microbial species, including *R. torques* and *O. splanchnicus*, also ranked highly, underscoring the marginal added predictive

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

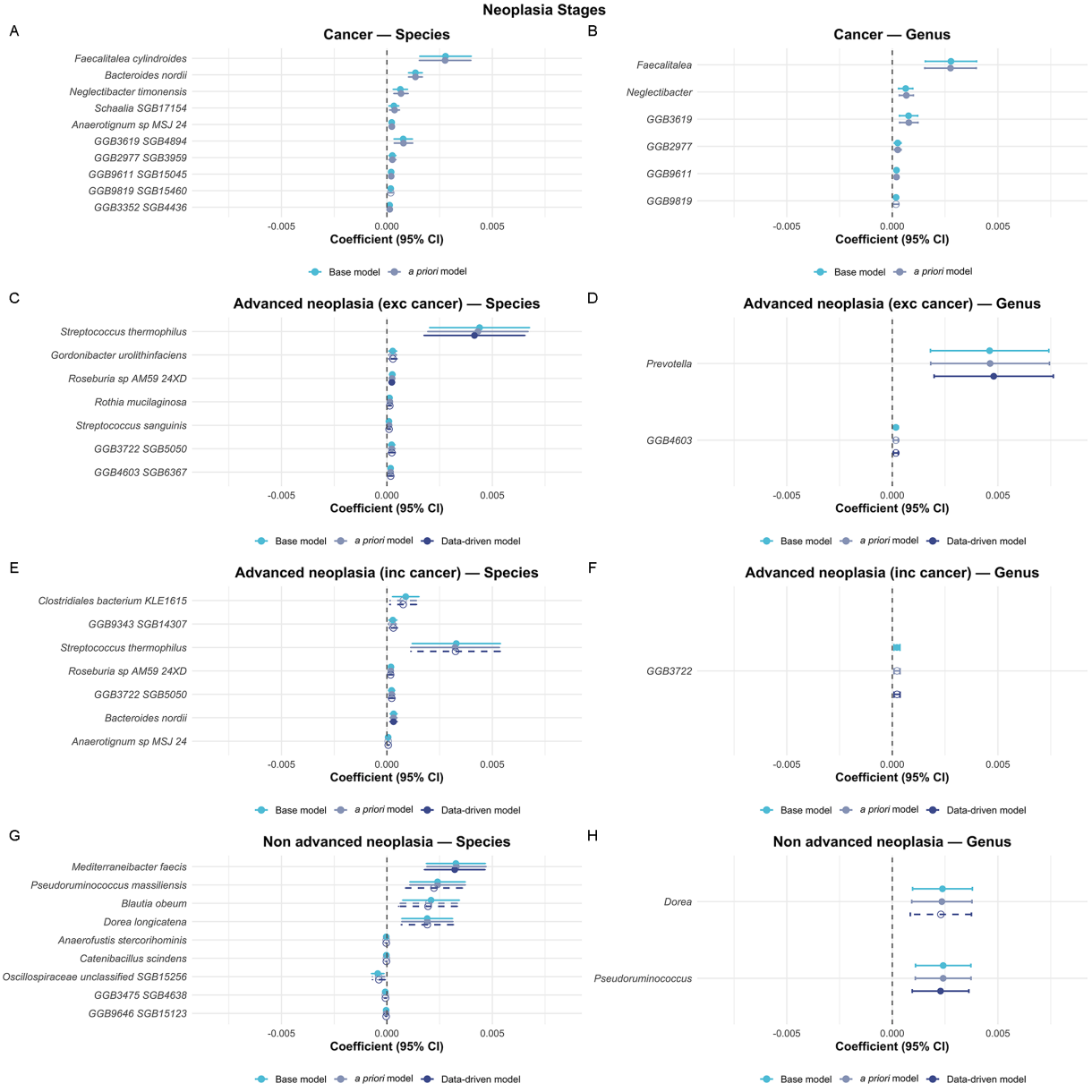


Figure 3 Taxonomic associations with colorectal neoplasia stages. Points represent beta coefficients; error bars indicate 95% CIs. Filled points and solid lines denote taxa passing the false discovery rate <0.1 threshold. Colours represent model adjustment depth: light blue=base model (age, sex, body mass index, ethnicity, site and batch); mid-blue=a priori model (base+dietary, lifestyle and medication covariates) and dark blue=data-driven model (a priori+additional stable covariates identified by recursive feature elimination).

value of microbiome features when integrated with conventional risk factors. Sensitivity analyses yielded similar findings, with the combined model again performing best (mean AUC: 0.633; range 0.535–0.685), and slight shifts in the importance of features, although *R. torques*, *O. splanchnicus and R. inulinivorans* remained key microbiome predictors alongside protein, alcohol, red meat intake and symptom or screening status (online supplemental figure 12).

DISCUSSION

This is the largest single study (using the same protocol for all patients) investigating the association between the gut microbiome and non-CRC colorectal neoplasia. Combining colonoscopy outcomes with deep phenotyping and high-resolution metagenomic profiling, we characterised microbiome composition and function across the spectrum of colorectal neoplasia. While we report several statistically significant microbial taxa and functional pathways, effect sizes were small, and many associations did not persist after adjustment for confounders. Improvements in predictive performance were negligible, indicating weak microbial signatures for non-cancerous colorectal lesions.

In accordance with our systematic review[7] and previous smaller studies,[24–26] we observed no differences in alpha diversity across neoplasia presence, subtypes or disease stages, suggesting that microbial richness is largely preserved in early-stage disease. Subtle differences in beta diversity were observed in the neoplasia group, particularly those with non-advanced lesions. However, low R^2 values and lack of clustering suggest these differences are unlikely to be clinically meaningful.

We identified *M. faecis* and *P. massiliensis* as novel taxa associated with neoplasia presence, non-advanced lesions, APs and

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

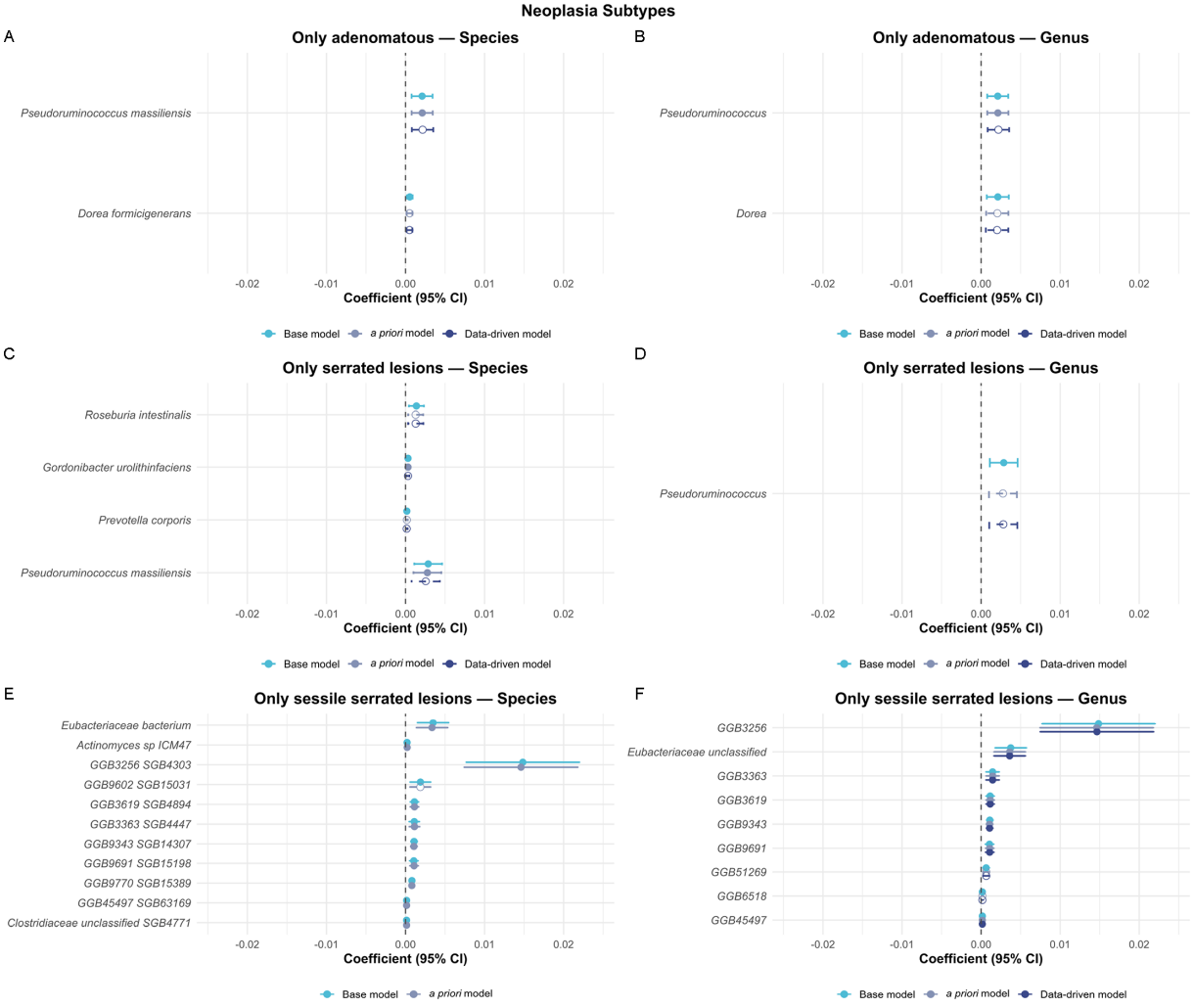


Figure 4 Taxonomic associations with colorectal neoplasia subtypes. Forest plots showing selected species and genera associations from linear mixed-effects models. Points represent beta coefficients; error bars indicate 95% CIs. Filled points and solid lines denote taxa passing the false discovery rate <0.1 threshold. Colours represent model adjustment depth: light blue=base model (age, sex, body mass index, ethnicity, site and batch); mid-blue=a priori model (base+dietary, lifestyle and medication covariates) and dark blue=data-driven model (a priori+additional stable covariates identified by recursive feature elimination).

ALs. *P. massiliensis* also showed significant LFCs between groups using an alternative modelling framework. Although the associations were consistent in direction across methods, effect sizes are not directly comparable because the models estimate different scales. Moreover, the Maaslin2 associations, while statistically significant, had small effect sizes and became non-significant after controlling for data-driven covariates such as red meat intake and waist circumference. This attenuation suggests observed microbial associations are explained by lifestyle and clinical factors.

Established CRC-associated taxa like *Fusobacterium nucleatum* and *Parvimonas micra* showed no significant associations in our neoplasia group (*F. nucleatum*, <5% prevalence; *P. micra*, 7.3% prevalence, p=0.36). This is consistent with Wirbel *et al*,[27] who noted variable associations across populations. Although non-cancerous neoplasia was the primary focus of this study, only a small number of participants had CRC (n=28), reflecting the unbiased design of this colonoscopy cohort, with participants recruited irrespective of findings. Nevertheless, *Bacteroides nordii* showed consistent association with CRC, but not precancerous lesions. This association remained significant in sensitivity analyses that grouped CRC with advanced neoplasia. These findings indicate that *B. nordii* is enriched in CRC, but not in precancerous lesions. While *B. nordii* was previously linked to high-risk, left-sided adenomas in the Gastrointestinal Disease Endoscopy Registry (GIDER) cohort,[8] we did not replicate this association when stratifying by colon subsite in this study.

Our findings are consistent with emerging evidence that early-stage colorectal neoplasia may not be associated with marked alterations in the faecal microbiome. A recent 16S ribosomal RNA (rRNA) sequencing study of low-grade adenomas[28] similarly reported no significant differences in alpha- or beta-diversity compared with controls. The study, which used mucosal biopsies collected within 20 mm from the lesion, identified limited associations confined to unclassified Lachnospiraceae amplicon sequence variants. In contrast, our shotgun metagenomic analysis did not detect Lachnospiraceae signals, which may reflect differences in assay type (16S vs shotgun metagenomics) or sampling (faecal vs mucosal). Taken together, these findings support increasing consensus that community-level taxonomic changes in early-stage neoplasia are subtle or absent.

Functional analyses revealed enrichment of pathways, including L-glutamate biosynthesis and peptidoglycan biosynthesis V (β-lactam resistance) in the neoplasia group. These

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

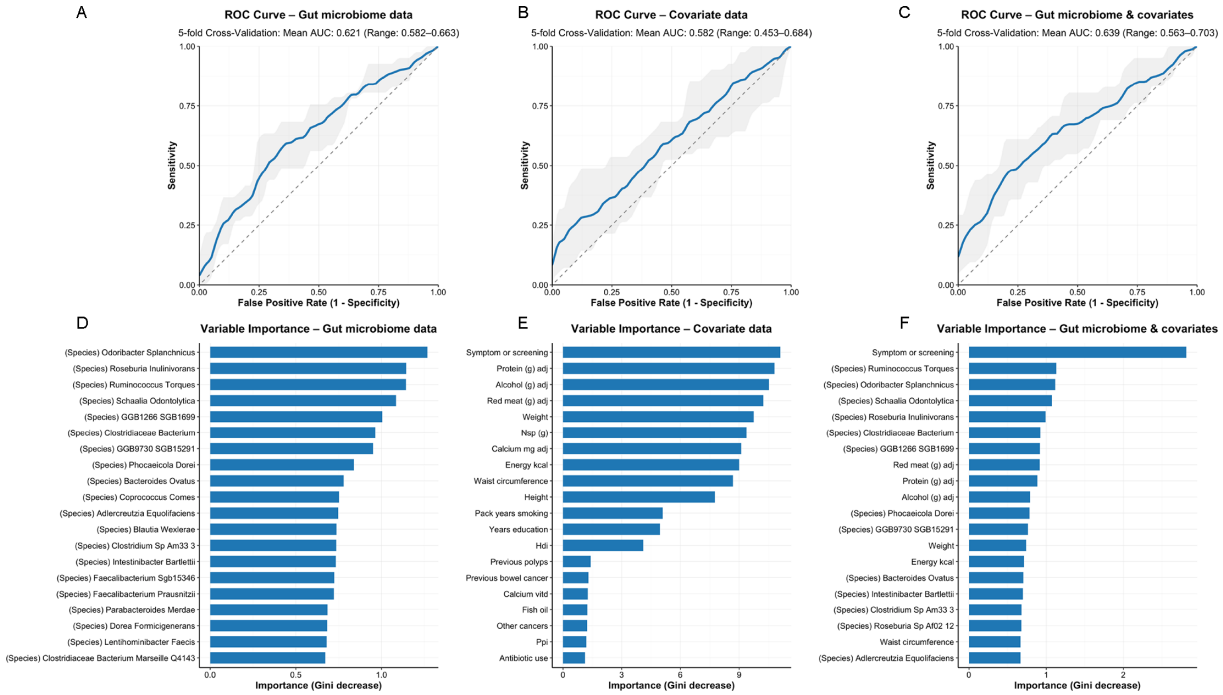


Figure 5 Random forest classification of high-risk neoplasia using gut microbiome species, covariates and their combination. (A–C) Receiver operating characteristic (ROC) curves showing the mean area under the curve (AUC) with minimum and maximum AUC values from fivefold cross-validation. Models were trained using (A) gut microbiome species alone, (B) covariates alone, and (C) a combination of gut microbial species and covariates. (D–F) Top 20 features ranked by mean decrease in Gini impurity for each respective model: (D) microbiome species, (E) covariates, and (F) combined model. Feature importance reflects each variable's contribution to model performance.

findings align with Yachida *et al*,[29] who reported amino acid metabolism to CRC progression and suggest that microbial metabolic shifts may support tumour metabolism or adaptation to altered gut environments. The association with β-lactam resistance pathways, also reported in metatranscriptomic studies of CRC,[30] may reflect microbial responses to host immune activation.

Random forest models demonstrated moderate predictive performance of the microbiome for high-risk neoplasia (combined model AUC=64%), outperforming, although modestly, covariate-only (58%) and microbiome-only (62%) models. Key microbial predictors, including *O. splanchnicus* and *R. torques*, alongside clinical factors like symptomatic/screening presentation and red meat intake, might suggest a complementary role for microbiome data. However, this performance is inferior to several other models that do not include the microbiome when applied in symptomatic populations (AUC=80%–85%),[31] indicating microbiome profiles are not viable standalone diagnostic tools for the presence of high-risk neoplasia. The prominence of the referral pathway as the top predictor in random forest models likely reflects underlying differences in symptom burden, case-mix severity and pre-test probability rather than a biological signal, highlighting the importance of accounting for clinical context when interpreting microbiome-based predictions.

Key strengths of this study include its large sample size (n=1762) nested within routine clinical practice, with consistent recruitment protocols across multiple NHS sites, standardised BSG classifications, comprehensive confounder adjustment and minimal missing data (0.06%–2.44%). High-resolution metagenomics provided greater taxonomic resolution than 16S rRNA methods, facilitating convergent evidence across methodologies. The study population's age and sex distribution closely align with the UK colonoscopy population, enhancing generalisability to routine clinical settings. RFE revealed key confounders, exposing the fragility of some associations while reducing spurious findings.

However, several limitations should be acknowledged. The cohort was predominantly composed of symptomatic referrals (approximately 93%), with only 7% from the BCSP, largely due to logistical constraints in retrieving pre-bowel preparation samples from screening hubs. Consequently, clinical presentation, symptom severity, pre-test probability and FIT positivity (UK referrals are largely FIT-driven) may have influenced microbiome composition and neoplasia risk, making the referral pathway a composite confounder. These findings therefore primarily reflect microbiome differences in symptomatic patients undergoing colonoscopy rather than asymptomatic screening populations. Stratified analyses by referral pathway revealed distinct patterns, but direct comparisons between groups are constrained by the smaller BCSP subgroup.

The no-neoplasia control group was restricted to a subset of individuals with a healthy colon (no polyps), providing a clean baseline for mechanistic inference but potentially reducing real-world relevance, as routine colonoscopy populations often include non-neoplastic polyps. There remains no consensus in the field on optimal control group definitions.[7 32] Similarly, strict outcome categorisation excluded participants with mixed polyp profiles from subtype analyses to enhance specificity, but this limits generalisability to clinical practice, where mixed or indeterminate lesions are common. The cross-sectional design precludes causal inference or evaluation of temporal dynamics.

A key consideration is that faecal sampling captures the luminal microbiome and may not fully reflect localised microbial interactions at the mucosal–neoplastic interface. However, stool sampling aligns with potential non-invasive screening applications, whereas mucosal sampling, while valuable for mechanistic

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

insights, lacks feasibility at scale. Self-reported dietary, medication and lifestyle data may introduce measurement error, though completion rates were excellent. Modest sample sizes for some subtypes, for example, n=28 cancer cases, may reduce statistical power for detecting subtype-specific associations; the absence of significant associations should therefore be interpreted cautiously, though precancerous lesions were the primary focus. Additionally, predictive performance models were evaluated using fivefold cross-validation without an independent test set due to the modest sample size for high-risk cases (n=207 cases) and matched comparators. Consequently, the reported AUCs, although modest, may be slightly optimistic and should be interpreted cautiously.

The cohort was overwhelmingly White British (97%), reflecting broader recruitment challenges in UK colonoscopy settings, where South Asian and Black communities are less likely to attend colonoscopy following FIT screening.[33] This limits applicability to more ethnically diverse populations. Importantly, UK colonoscopy referral is largely driven by FIT positivity (≥10 µg)[34]; accordingly, our cohort is FIT-enriched, and any relationship between FIT status and gut microbiome composition could represent a potential source of confounding. Finally, while metagenomic analyses provide valuable insights into genetic potential, they do not capture true microbial function, underscoring the need for future studies integrating metatranscriptomic or metabolomic data.

In conclusion, these findings, together with our prior review (see Manning *et al*[7]), suggest weak and inconsistent microbiome associations with precancerous colorectal lesions, questioning the diagnostic value of gut microbiome profiling in precancer settings. This contrasts with evidence for microbial signatures in CRC. These observations may stem from the heterogeneity of precancerous lesions, the fact that many may never become cancerous, and that microbiome alterations may manifest closer to or after malignant transformation. The fact that the microbial signature becomes apparent when CRC develops raises the possibility that the microbiome may reflect the presence of cancer rather than the microbiome driving the development of cancer.

Author affiliations
[1]Human Nutrition & Exercise Research Centre, Centre for Healthier Lives, Population Health Sciences Institute, Newcastle University, Newcastle University, Newcastle upon Tyne, UK
[2]Department of Gastroenterology, South Tyneside and Sunderland NHS Foundation Trust, South Tyneside District Hospital, South Tyneside and Sunderland NHS Foundation Trust, South Shields, UK
[3]Population Health Sciences Institute, Newcastle University Centre for Cancer, Faculty of Medical Sciences, Newcastle University, Newcastle University Centre for Cancer, Newcastle upon Tyne, UK
[4]Leeds Institute of Medical Research, University of Leeds, St James's University Hospital, Leeds, UK
[5]Translational and Clinical Research Institute, Newcastle University, Newcastle University Faculty of Medical Sciences, Newcastle upon Tyne, UK
[6]North East Hub, Bowel Cancer Screening Programme, Gateshead, UK
[7]Microbiome Research Centre, St George & Sutherland Clinical Campuses, School of Clinical Medicine, Faculty of Medicine & Health, University of New South Wales, Sydney, New South Wales, Australia

Correction notice This article has been corrected since it published Online First. The author affiliations have been corrected.

Acknowledgements We express our sincere thanks to all participants of COLO-COHORT, and staff at each recruitment site. Additionally, we also express thanks to the study sponsor, South Tyneside and Sunderland NHS Trust.

Collaborators The COLO-COHORT research team was comprised of Richard Gallon, Ingrid Emmerson, Claire Livingstone, Amy Burns, Mark Graham, Nathania Bestwick, Amy Peasland, Aamar Ghouri, John Whelpton, Ajay Verma, Katharina Wallis, Sajal Rai, Tom Lee, Matthew Johnston, Sorena Afshar, John Greenaway, John Louis-Auguste, Andrew Day, Mark Canoy, Abby Conlin, Simon Hellier, Ian Farrell, Rosin Bevan, Susan Moug, Joe Geraghty, Noo Mohammed, Chris Mountford, Roger Prudham, Matthew Tutton, Ali Malik, Mo Thoufeeq, Shameer Mehta, S Sebastian, Kathy Dent.

Contributors Conceived and designed the experiments: LS, MH, CS and CJR. Contributed reagents/materials/analysis tools: GRY, CD, SMi and CS. Data curation: GRY, SMa, EH, YD and GST. Formal Analysis: PL. Guarantor: PL, CJR and CS. Wrote the manuscript: PL, LS, MH, CS and CJR. Revised the manuscript: all authors read and approved the final manuscript.

Funding This study is primarily funded by GUTS UK (Charity Registration No. 1137029, formerly known as Core), which administers an award by The Parabola Foundation, along with a funding award from The Sir Bobby Robson Foundation for Colo-Speed infrastructure and research delivery. CS is funded by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (221745/Z/20/Z) and the 2021 Lister Institute Prize Fellow Award. GRY is supported by the NIHR Newcastle Biomedical Research Centre.

Competing interests LN has received an unrestricted project grant from Medtronic and conference attendance support from Dr Falk Pharma and Tillotts. All other authors declare no competing interests.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants. The COLO-COHORT study was reviewed and approved by the West Midlands Edgbaston Research Ethics Committee (ref: 19/WM/0193, June 2019) and received favourable Bowel Cancer Screening Programme Research Advisory Committee approval (October 2019). The study was registered with ISRCTN (ISRCTN: 17473023) and Clinical Trials.gov (NCT04185779). The COLO-COHORT study was also adopted into the National Institute for Health Research Clinical Research Network portfolio. All participants provided informed consent to participate in the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The gut microbiome data used in this study are available on the EBI—European Nucleotide Archive (www.ebi.ac.uk/ena/) under accession number PRJEB91658.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: https://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Panayiotis Louca https://orcid.org/0000-0001-5956-1433
Sarah Manning https://orcid.org/0000-0002-7105-8109
Linda Sharp https://orcid.org/0000-0001-9515-1722
Mark A Hull https://orcid.org/0000-0001-7414-1576
Gregory R Young https://orcid.org/0000-0001-5342-1421
Suparna Mitra https://orcid.org/0000-0002-9378-1496
Emad M El-Omar https://orcid.org/0000-0002-0011-3924
Christopher J Stewart https://orcid.org/0000-0002-6033-338X
Colin J Rees https://orcid.org/0000-0003-3050-8473

REFERENCES

1 Morgan E, Arnold M, Gini A, *et al*. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* 2023;72:338–44.
2 Strum WB. Colorectal Adenomas. *N Engl J Med* 2016;374:1065–75.
3 Leslie A, Carey FA, Pratt NR, *et al*. The colorectal adenoma-carcinoma sequence. *Br J Surg* 2002;89:845–60.
4 Rutter MD, East J, Rees CJ, *et al*. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. *Gut* 2020;69:201–23.
5 East JE, Atkin WS, Bateman AC, *et al*. British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. *Gut* 2017;66:1181–96.

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

6 Song M, Chan AT, Sun J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology* 2020;158:322–40.
7 Manning S, Hackney E, Dunneram Y, *et al*. Systematic Review: The Relationship Between the Faecal Microbiome and Colorectal Neoplasia in Shotgun Metagenomic Studies. *Aliment Pharmacol Ther* 2025;62:568–84.
8 Lee JWJ, Plichta DR, Asher S, *et al*. Association of distinct microbial signatures with premalignant colorectal adenomas. *Cell Host Microbe* 2023;31:827–38.
9 Hampton JS, Koo S, Dobson C, *et al*. The COLO-COHORT (Colorectal Cancer Cohort) study: Protocol for a multi-centre, observational research study and development of a consent-for-contact research platform. *Colorectal Dis* 2022;24:1216–26.
10 Participants in the Paris Workshop. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon. *Gastrointest Endosc* 2003;58:S3–43.
11 Mitra S, J Stewart C, Nelson A, *et al*. Long-term stability of the faecal microbiome profile in faecal immunochemical test (FIT) samples. *Gut* 2025;74:e8.
12 Masi AC, Koo S, Lamb CA, *et al*. Using faecal immunochemical test (FIT) undertaken in a national screening programme for large-scale gut microbiota analysis. *Gut* 2021;70:429–31.
13 Oksanen J, Simpson GL, Blanchet FG, *et al*. vegan: community ecology package. 2024.
14 Shannon CE, Weaver W. The mathematical theory of communication. Urbana: The University of Illinois Press, 1949.
15 Bray JR, Curtis JT. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol Monogr* 1957;27:325–49.
16 Jaccard P. The distribution of the flora in the alpine zone. *New Phytol* 1912;11:37–50.
17 R Core Team. R: a language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing; 2020.
18 Kowarik A, Templ M. Imputation with the R Package VIM. *J Stat Softw* 2016;74:1–16.
19 Mallick H, Rahnavard A, McIver LJ, *et al*. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* 2021;17:e1009442.
20 Lin H, Peddada SD. Multigroup analysis of compositions of microbiomes with covariate adjustments and repeated measures. *Nat Methods* 2024;21:83–91.
21 Suzek BE, Wang Y, Huang H, *et al*. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 2015;31:926–32.
22 Consortium TU. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 2022;51:D523–31.
23 Beaton D, Sharp L, Trudgill N, *et al*. British Society of Gastroenterology national evaluation of colonoscopy quality: findings from the National Endoscopy Database. *Gastrointest Endosc* 2025;102:117–28.
24 Feng Q, Liang S, Jia H, *et al*. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* 2015;6:6528.
25 Peters BA, Dominianni C, Shapiro JA, *et al*. The gut microbiota in conventional and serrated precursors of colorectal cancer. *Microbiome* 2016;4:69.
26 Piccinno G, Thompson KN, Manghi P, *et al*. Pooled analysis of 3,741 stool metagenomes from 18 cohorts for cross-stage and strain-level reproducible microbial biomarkers of colorectal cancer. *Nat Med* 2025;31:2416–29.
27 Wirbel J, Pyl PT, Kartal E, *et al*. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019;25:679–89.
28 Welham Z, Li J, Tse B, *et al*. Gut Mucosal Microbiome of Patients With Low-Grade Adenomatous Bowel Polyps. *Gastro Hep Adv* 2025;4:100687.
29 Yachida S, Mizutani S, Shiroma H, *et al*. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* 2019;25:968–76.
30 Lamaudière MTF, Arasaradnam R, Weedall GD, *et al*. The Colorectal Cancer Gut Environment Regulates Activity of the Microbiome and Promotes the Multidrug Resistant Phenotype of ESKAPE and Other Pathogens. *mSphere* 2023;8:e00626–22.
31 Hampton JS, Kenny RPW, Rees CJ, *et al*. The performance of FIT-based and other risk prediction models for colorectal neoplasia in symptomatic patients: a systematic review. *EClinicalMedicine* 2023;64:102204.
32 Manning S, Sinha R, Rees CJ. Need for standardised approaches to human microbiome research using the example of colorectal neoplasia research. *Gut* 2025;74:875–7.
33 Kerrison RS, Gil N, Travis E, *et al*. Barriers to colonoscopy in UK colorectal cancer screening programmes: Qualitative interviews with ethnic minority groups. *Psychooncology* 2023;32:779–92.
34 von Wagner C, Stoffel S, Freeman M, *et al*. Attitudes towards faecal immunochemical testing in patients at increased risk of colorectal cancer: an online survey of GPs in England. *Br J Gen Pract* 2018;68:e757–64.

Gut: first published as 10.1136/gutjnl-2025-337478 on 6 March 2026. Downloaded from http://gut.bmj.com/ on October 23, 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.