Philip et al. Journal of Translational Medicine (2025) 23:549
https://doi.org/10.1186/s12967-025-06552-w

Journal of
Translational Medicine

RESEARCH

Open Access

Deciphering microbial and metabolic influences in gastrointestinal diseases-unveiling their roles in gastric cancer, colorectal cancer, and inflammatory bowel disease

Daryll Philip[1], Rebecca Hodgkiss[2], Swarnima Kollampallath Radhakrishnan[2], Akshat Sinha[2] and Animesh Acharjee[1,2,3,4*]

Abstract

Introduction Gastrointestinal disorders (GIDs) affect nearly 40% of the global population, with gut microbiome-metabolome interactions playing a crucial role in gastric cancer (GC), colorectal cancer (CRC), and inflammatory bowel disease (IBD). This study aims to investigate how microbial and metabolic alterations contribute to disease development and assess whether biomarkers identified in one disease could potentially be used to predict another, highlighting cross-disease applicability.

Methods Microbiome and metabolome datasets from Erawijantari et al. (GC: n = 42, Healthy: n = 54), Franzosa et al. (IBD: n = 164, Healthy: n = 56), and Yachida et al. (CRC: n = 150, Healthy: n = 127) were subjected to three machine learning algorithms, eXtreme gradient boosting (XGBoost), Random Forest, and Least Absolute Shrinkage and Selection Operator (LASSO). Feature selection identified microbial and metabolite biomarkers unique to each disease and shared across conditions. A microbial community (MICOM) model simulated gut microbial growth and metabolite fluxes, revealing metabolic differences between healthy and diseased states. Finally, network analysis uncovered metabolite clusters associated with disease traits.

Results Combined machine learning models demonstrated strong predictive performance, with Random Forest achieving the highest Area Under the Curve(AUC) scores for GC(0.94[0.83–1.00]), CRC (0.75[0.62–0.86]), and IBD (0.93[0.86–0.98]). These models were then employed for cross-disease analysis, revealing that models trained on GC data successfully predicted IBD biomarkers, while CRC models predicted GC biomarkers with optimal performance scores.

Conclusion These findings emphasize the potential of microbial and metabolic profiling in cross-disease characterization particularly for GIDs, advancing biomarker discovery for improved diagnostics and targeted therapies.

Keywords Gastric cancer, Inflammatory bowel disease, Colorectal cancer, Microbiome, Metabolome, Biomarkers, Machine learning

*Correspondence:
Animesh Acharjee
a.acharjee@bham.ac.uk
Full list of author information is available at the end of the article

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Introduction

Gastrointestinal diseases (GIDs) are disorders that impact the gastrointestinal tract, which extends from the esophagus to the rectum and includes the pancreas, liver, and gallbladder [1]. GIDs can be broadly categorized into those resulting from malignancies, such as gastric cancer (GC), pancreatic cancer, esophageal cancer, liver cancer, and colorectal cancer, also known as colon cancer (CRC) [2], and those driven by inflammatory responses, including inflammatory bowel disease (IBD) and irritable bowel syndrome (IBS) [3].

Among malignancy related GIDs, GC represents a significant global health burden. The primary risk factor for GC is infection by *Helicobacter pylori* (*H. pylori*), which causes chronic inflammation and significantly elevates the risk of malignant tumor formation in the gastric lining [4]. According to the Global Cancer Observatory (GLOBOCAN) 2022 data, GC is the fifth most common cancer worldwide, with 968,350 new cases and 659,853 deaths reported [5]. It ranks among the leading causes of mortality in 42 countries, with higher incidence rates in males compared to females, particularly in regions of Eastern Asia [6].

Likewise, CRC is characterized by malignant growths or polyps in the colon or rectum [7]. According to GLOBOCAN, CRC ranks as the third most commonly diagnosed cancer and the second leading cause of cancer related deaths globally, with 1.9 million new cases and 903,859 deaths in 2022. The disease is more prevalent in men than women, with the highest incidence rates found in European countries such as Norway and Denmark [5].

In contrast, inflammation driven GIDs, such as IBD, are characterized by chronic morbidity resulting from immune mediated inflammatory processes. IBD serves as an umbrella term that includes Crohn's disease (CD) and ulcerative colitis (UC) [8, 9]. According to the Global Burden of Disease (GBD) 2019, there were approximately 4.9 million IBD cases worldwide, resulting in 41,000 deaths. Prevalence rates were highest in Norway, followed by Canada, with both prevalence and mortality rates being higher in females compared to males [10].

Established risk factors such as dietary habits [11–13], genetic predispositions [14–16], and lifestyle choices [17–19] are more or less associated with GC, CRC, and IBD. Another significant risk factor is dysbiosis, which refers to an imbalance in the gut microbiome, the vast and diverse community of bacteria and other microorganisms residing in our digestive system. Research shows that the abundance or depletion of certain microbes in the gut, which is often referred to as the 'second brain', can play a pivotal role in GIDs [20]. For instance, Zeng et al. [21] demonstrated that in addition to *H.pylori,* microbes like *Prevotella* and *Streptococcus* were abundant, while beneficial microbes such as *Bifidobacterium* were depleted in the fecal samples from GC patients.

Similarly, Villéger et al. [22] reported increased levels of *Bacteroides* and *Prevotella* in CRC, alongside reduced levels of *Lactobacillus* and *Faecalibacterium.* In IBD, *Streptococcus* levels were elevated in UC patients, and *Lachnoclostridium* and *Fusobacterium* were significantly increased in CD patients [23].

Beyond microbial composition, GIDs are influenced by the metabolites produced by the gut microbiome. Disruptions in microbial metabolite production can lead to metabolic reprogramming, contributing to GC, IBD, and CRC pathogenesis. For example, metabolic pathways involving lipids, nucleotides, and amino acids such as alanine and valine were found to be dysregulated in patients with GC [24, 25]. Zhang et al. [26] highlighted that microbiota derived metabolites such as trimethylamine-N-oxide, secondary bile acids, hydrogen sulfide, and N-nitroso compounds could induce inflammation and modulate tumor immunity in the colon. In IBD, bile acids, such as deoxycholic acid, activate inflammatory signaling pathways while dysregulated tryptophan metabolism, particularly in UC patients, further heightens intestinal inflammation [27, 28].

GIDs are serious conditions with high mortality and morbidity rates, often going undiagnosed in their early stages [29]. This is partly because the symptoms can be subtle or easily overlooked. Unfortunately, delayed diagnosis allows these diseases to progress rapidly, with one condition often leading to another through shared pathological mechanisms.

For instance, Fretwell et al. [30], in their systemic review of case reports containing histopathology data, stated that although rare, GC can metastasize to the colorectum through lymphatic, vascular, or mesenteric routes.

Similarly, Tak et al. [31] analyzed clinicopathological characteristics of patients in Korea and reported that individuals with CRC are at an increased risk of developing intestinal metaplasia and gastric adenomas, which are precursors to GC, within the first four years of CRC diagnosis. Furthermore, Sato et al. [32] emphasized in their study, which evaluated clinical studies, meta-analyses, and systematic reviews, that chronic mucosal inflammation in UC can increase the chances of developing colorectal neoplasia by progressing from low-grade to high-grade dysplasia, eventually developing into CRC.

Given the complexity of microbiome-metabolite interactions and their critical role in GIDs, it has become steadily more beneficial to train machine learning models to produce highly accurate, reproducible, and interpretable insights from large and complex datasets

[33]. Recent studies highlight the effectiveness of machine learning algorithms in differentiating diseased patients from healthy individuals [34, 35], detecting important microbial and metabolic biomarkers [36–39], and uncovering risk factors [40] associated with diseases that affect the gastrointestinal tract.

Our goal is to implement machine learning models for GC, IBD, and CRC to identify the most significant and differential microorganisms and metabolites in fecal samples using publicly available datasets. These biomarkers are then used to stratify and predict cross-disease associations. Specifically, the GC model was utilised to predict IBD and CRC, the IBD model was used to predict GC and CRC, and the CRC model was applied to predict GC and IBD. This approach allows us to uncover both shared and unique patterns among GIDs. Subsequently, the identified biomarkers are integrated into in-silico modeling techniques to assess microbial contributions to metabolite production. Finally, network analysis techniques are conducted to uncover correlations between features and biological pathways linking microbes and metabolites in diseased and healthy patients.

Methods

Data preprocessing

To mitigate overfitting and improve model performance, we eliminated sparse features from the dataset, and the remaining data was normalized using min–max scaling [41], transforming values to a range between 0 and 1. This prevented the dominance of features with larger ranges or values, ensuring that all the features contribute equally to the model.

Unsupervised machine learning models

Principal components analysis (PCA)

We applied PCA [42] for dimensionality reduction, identifying principal components (PC1 and PC2) that capture maximum variance for the metabolite datasets. Outliers were detected by calculating the Mahalanobis distance [43], with a 95% confidence interval threshold derived from the chi-squared distribution.

Principal coordinates analysis (PCoA)

We performed PCoA [44], or metric multidimensional scaling, for outlier detection on the microbiome datasets. The Bray–Curtis [45] dissimilarity matrix was calculated based on abundance data, and the resulting PCoA plot projected the high-dimensional data into a lower-dimensional space. This visualisation captured similarities and differences between samples using the first two principal coordinates. The 95% confidence ellipses for each group provide a visual means of identifying potential outliers by highlighting variations within and between sample groups.

Univariate statistical analysis

To prioritize the most promising features for the computationally intensive machine learning models, we applied non-parametric tests such as the Mann–Whitney U test [46] for GC and IBD and the Kruskal–Wallis H test [47] for CRC to assess differences between two independent groups, particularly when the data does not follow normal distribution. To control the rate of false positives, p-values were adjusted using the Benjamini-Hochberg (BH) [48] method. To optimize computational efficiency, we considered features with adjusted p-values below 0.05 as significant, focusing our analysis on only the most differential microbes and metabolites.

Supervised machine learning models

Three machine learning models were used in this work to analyze the microbiome and metabolome datasets efficiently. The workflow can be seen in Supplementary Fig. 1. We employed eXtreme gradient boosting (XGBoost) [49] as an ensemble algorithm that excels at classification tasks by iteratively refining predictions through gradient boosting. It incorporates regularization techniques to minimize errors and prevent overfitting, ensuring a balance between accuracy and model complexity. Similarly, Random Forest [50] was employed due to its ability to handle high-dimensional microbiome data. It reduces overfitting and boosts overall performance by constructing numerous decision trees using random subsets of the input, thus reducing variance and improving model stability. In addition to the ensemble models, the Least Absolute Shrinkage & Selection Operator (LASSO) [51] was utilised due to its ability to combine classification and feature selection. In binary classification, LASSO uses the regularization parameter C to control regularization strength. Since C is the inverse of λ, which is the penalty term coefficient ($C = 1/\lambda$), a higher C value reduces regularization, allowing more features to stay in the model. Conversely, a lower C value strengthens regularization, shrinking more coefficients to zero. This eliminates the need for a separate feature selection method for LASSO.

Hyperparameter tuning through random search and Bayesian optimization

Hyperparameters are settings that control the behavior of machine learning algorithms, and hyperparameter tuning optimizes model performance by finding the best settings. In our study, this process begins with a random search [52], which explores random combinations of hyperparameters. However, this approach may miss optimal feature combinations if the search space is sparse.

To address this, Bayesian optimization (BO) [53] was employed. BO used a Gaussian process to build a probabilistic model from the best random search results, guiding the selection of subsequent hyperparameters. In this study, the data was split 75% for training and 25% for testing.

The random search was followed by fivefold cross-validation to refine the parameter grid, which was then optimized using BO with fivefold cross-validation to identify the best-performing hyperparameters, which were initially applied on the training data and then on the unseen test data.

Model evaluation

Performance metrics, like Receiver Operating Characteristic Area Under Curve (ROC-AUC) [54], were calculated across the models for all the diseases in our study because of their ability to consider trade-offs between specificity and recall, with higher values indicating better discrimination. Alongside ROC-AUC, other metrics such as accuracy, precision, recall, F1-score, and specificity were calculated, which provided a more comprehensive evaluation of the models. To further assess the reliability of these scores, the 95% confidence intervals (CI) were computed for each performance metric.

Feature selection

We applied Recursive Feature Elimination with Cross-Validation (RFECV) to XGBoost and Random Forest to refine the feature sets. This process iteratively removed less important features through cross-validation, selecting the optimal set based on the highest cross-validation scores. Overlapping features selected by RFECV and those automatically chosen by LASSO were identified. However, since the number of microbes and metabolites remained high, we further evaluated feature subsets by computing ROC-AUC scores for the best performing models. Subsets with the top 5, 10, 15, 20, 25, and 30 features were iteratively tested, and the model achieving the highest AUC with the fewest features was chosen for all diseases. Feature rankings were determined using either LASSO coefficients or Gini feature importance. Finally, a Spearman correlation cluster map with hierarchical clustering was generated to visualise clusters of microbes and metabolites that were strongly correlated, which provided more understanding into their relationships.

Diversity analysis

To understand the complexity and diversity of the microbial communities in the gut within the healthy and diseased groups, we measured alpha diversity [55] indices, which quantify both richness (the number of distinct genera) and evenness (the uniformity of distribution among those genera). The diversity index (D) value increases with higher richness and evenness. Among these, the Shannon-Weiner index [56] is the most used metric due to its ease of interpretation. It measures the uncertainty of predicting a species from a community and is particularly sensitive to species richness.

The probability that two randomly chosen microbes belong to the same species is measured by Simpson's Index [57] with lower values indicating greater diversity. It is often expressed as the Gini-Simpson index (1 − D). Statistical significance within groups was evaluated by calculating p-values and applying FDR correction.

Beta diversity [58] captures the variation or dissimilarity in genera between the sample groups. We utilised the non-metric dimensional scaling (NMDS) to visualise the similarities or dissimilarities in a low-dimensional space, employing the Jaccard distance. The stress values obtained from NMDS indicate the accuracy of 2D representations with lower stress values indicating a better fit between the original dissimilarity matrix and the NMDS ordination (see Table 1).

Microbial community model (MICOM)

To explore connections between the microbes and metabolites identified, a microbial community model was created for each disease, stimulating gut communities for each sample. MICOM uses an L2 normalisation based model to calculate the community growth rate, denoted as μ_c, for all the microbes in a metagenomic sample [59]. This method enables what Diener et al. define as selfish individual growth maximization [59], allowing each microbe to reach its maximal growth (μ_i) rather than just a maximal overall community growth. Simulated growth rates are determined based on the microbes relative abundance, known metabolite fluxes, and growth rates from an input database, as well as user input minimal and maximal abundance values and growth rates. Utilizing the relative abundance of each genus selected by the machine learning process and its corresponding classification, a manifest for each disease was built using the build function from micom.workflows. The model database was set to the "agora103_genus.qza" [60] dataset, with the solver set to "osqp", a cutoff equal to zero, threads equal to two, and a phenotype column indicating the disease status for each sample. This manifest is a data frame created by the model that includes all the information on the microbes identified in the provided database, which is then used to construct the growth model.

To obtain maximal growth rates, a cooperative trade-off value must be determined. The model fixes the community growth rate to a fraction of its optimum and then

Table 1 Summary of datasets used for training and validation in GC, CRC, and IBD

Datasets	Source	Healthy	Diseased	Total features	Source of extraction	Technology used
Training data						
GC						
Microbiome	Erawijantari et al. [178]	54	42	10,528	Fecal	Shotgun metagenomics sequencing
Metabolome				525		*CE-TOFMS
CRC						
Microbiome	Yachida et al. [179]	127	150	11,942	Fecal	Whole genome sequencing
Metabolome				450		CE-TOFMS
IBD						
Microbiome	Franzosa et al. [180]	56	164	11,720	Fecal	Whole genome shotgun sequencing
Metabolome				466		*LC–MS
Validation data						
GC						
Microbiome	Jaeyun Sung et al. [181]	10	40	470	Gastric antrum	16S rRNA sequencing
Metabolome	UK BioBank	44,378	2,436	168	Plasma	NMR spectroscopy
CRC						
Microbiome	Kim et al. [182]	102	36	499	Fecal	16S rRNA sequencing
Metabolome				462		*UPLC-MS/MS
IBD						
Microbiome	iHMP/HMP2 [183]	104	278	9694	Fecal	Shotgun metagenomic sequencing
Metabolome				596		LC–MS

This table includes the total number of healthy and diseased patients, the number of microbiome and metabolome features, the source of extraction for microbes and metabolites, and the sequencing and analytical tools employed, respectively

*CE-TOFMS: Capillary electrophoresis time-of-flight mass spectrometry, *LC–MS: liquid chromatography-mass spectrometry, *UPLC-MS/MS: ultra-performance liquid chromatography-mass spectrometry

calculates the minimum L2 normalisation of the individual growth rates. Individual growth rates are calculated as follows:

$$\mu_i = \frac{\alpha \mu_c}{a^T a} a_i$$

where α denotes the specified trade-off value (the fraction of community maximum to use), μ_c denotes the community growth rate, μ_i the individual growth rate for genus i and a_i is the relative abundance of that genus. Thus, the community growth rate is represented by the sum of individual growth rates and their abundance:

$$\mu_C = \sum\nolimits_i a_i \mu_i$$

Therefore, before creating the growth model, the optimal value for the trade-off was identified using the resulting manifest of each disease in the tradeoff function from micom.workflows, with the medium set to the "western_diet_gut.qza" [60] database and threads equal to two. The optimal value was defined as the highest trade-off value where the maximal number of taxa were enabled to grow.

Additionally, MICOM represents the flux balances of the microbes and provides the estimated production and consumption of metabolites by these recognised communities. A linear model based on the COBRApy Python package is utilised, with an assumption of a steady state of all fluxes within the microbes system required. The fluxes v_i are provided in millimoles per gram per hour (mmol/[gDWh]) and follow the rules:

$$maxmiziev_{bm}$$

$$suchthat(s.t)S_v = 0$$

$$andlb_i \leq v_i \leq ub_i$$

such that v_{bm} is the biomass reaction, which is normalised to produce 1 g of biomass in a unit 1/h, to correspond to the growth rate of the organism. Lower and upper bounds lb_i and ub_i are used to impose thermodynamic constraints. To allow a community of fluxes, the following must be considered:

$$maximize\mu_c = \sum_i a_i \mu_i$$

$$s.t."i : Sv = 0$$

$$\mu_i = {v_i}^{bm} \geq {\mu_i}^{min}$$

$$lb_i \leq v_i \leq ub_i$$

$$lb_i^{ex} \leq a_i v_i^{ex} \leq ub_i^{ex}$$

$$lb_i^{m} \leq v_i^{m} \leq ub_i^{m}$$

where ${v_i}^{bm}$ is the biomass flux, ${\mu_i}^{min}$ is the user-specified minimum growth rate (0 is used in this study), ${v_i}^{ex}$ is the exchange fluxes with the specified external environment, and lb and ub are the lower and upper bounds. ${v_i}^{m}$ are the exchanges between the entire community and the gut lumen, so a set metabolite environment representing the gut lumen must also be provided to the model. The “western_diet_gut.qza” database is used in this study. Overall production fluxes are calculated via:

$${v_{tot}}^{m} = \sum\nolimits_{i, {v_i}^{m} > 0} a_i {v_i}^{m}$$

with ${v_i}^{m}$ representing an exchange flux for the metabolite m in taxon i and ${v_{tot}}^{m}$ the total metabolite fluxes.

Consequently, to obtain the growth model and predict metabolite production and consumption by each genus, the grow function was enforced with the input trade-off set to the determined optimum, the manifest for the disease, and the same “western_diet_gut.qza” medium.

From these estimated fluxes, we utilise the phenotype provided for each sample to examine how these fluxes vary across disease groups, using MICOM’s built in non-parametric tests for each metabolite against the phenotype. To identify metabolites differentially produced between case and control samples of each disease, the plot_association function from micom.viz was populated with the growth results, variable_type set to binary, phenotype set to the disease status (case vs control), and fdr_threshold set to 0.5. Any metabolites or their derivatives identified by the MICOM model and in the predictive analytics were noted as important. Finally, the selected genus names were entered into the MicrobiomeAnalyst taxon set analysis tool to identify literature validated interactions between microbes and metabolites and were compared to the results provided by MICOM and machine learning analysis.

Weighted co-gene network analysis (WCGNA)

WCGNA [61] was performed on both the metabolite and microbiome datasets to explore co-expression patterns and their associations with case–control traits. First, the optimal power to create a scale-free topology network was determined by evaluating a range of soft-thresholding powers ranging from 1 to 20, with plots of mean connectedness and scale independence guiding the decision. Using the chosen power, hierarchical clustering and dynamic tree-cut techniques were used to identify modules of co-expressed features, each assigned a distinct colour for visualisation. For each module, the module eigengenes (MEs), were computed and compared to the case–control trait using the Pearson correlation coefficient. The statistical significance of the correlation coefficients, which varied from −1 (strong negative correlation) to +1 (strong positive correlation), was assessed using the appropriate p-values obtained from the correlation analysis. A heatmap of these correlations, displaying both the coefficients and the corresponding p-values, provided a comprehensive view of the module-trait relationships.

Additionally, the Topological Overlap Matrix (TOM) was calculated to assess the similarity between features based on their network connectivity. Heatmaps and network dendrograms were created using this TOM to show co-expression patterns. The co-expression networks were shown using graph based visualisation approaches after the TOM was filtered to highlight the strongest linkages for network visualisation.

Features (metabolites or taxa) were represented by nodes in these networks, while co-expression relationships were represented by edges, whose attributes were proportionate to the strength of the connection.

Results

Demographic characteristics for all datasets

Given the large number of datasets, we focused on basic demographic characteristics (Table 2). The datasets from Erawijantari et al. (GC) and Yachida et al. (CRC) revealed a higher proportion of male participants compared to females. The median age was 66 years for GC patients and 64 years for CRC, suggesting that the risk of developing GC and CRC may increase with age. However, statistical analysis showed no significant differences in median age between GC patients ($p = 0.75$), CRC patients ($p = 0.14$), and healthy controls. BMI for gastrectomy patients was higher (23.2) compared to the healthy group (21), with a p-value of 0.0004, indicating a statistically significant difference between the two groups. However, even though the mean BMI for CRC patients (23) was higher than the healthy group (22.9), there was no statistical difference between the two groups ($p = 0.66$).

For the IBD dataset by Franzosa et al., gender related data was unavailable, but the median age of IBD patients was 41 years. A significant difference in age between healthy individuals and IBD patients suggests a potential association between age and IBD prevalence ($p = 0.005$).

Table 2 Baseline demographic characteristics

Population demographics	Gastric cancer (Erawijantari et al.)				Inflammatory bowel disease (Franzosa et al.)				Colerectral cancer (Yachida et al.)			
	Total	**Gastrectomy**	**Healthy**	**p-value**	**Total**	**IBD**	**Healthy**	**p-value**	**Total**	**CRC**	**Healthy**	**p-value**
Participants	96	42	54		220	164	56		277	150	127	
Sex												
Male (%)	66.7	33.33	33.33	0.12	NA	NA	NA	NA	57.8	32.1	25.6	0.65
Female (%)	33.3	10.42	22.92		NA	NA	NA		42.2	22	20.2	
Age (median)	65 (60–68)	66 (60–68)	64 (60–68)	0.75	38 (35–43)	41 (36–47)	32 (30–37)	0.005	65 (64–67)	64 (63–66)	67 (65–69)	0.14
BMI (mean)	22.2 (22–23)	23.2 (22–24)	21 (20–22)	0.0004	NA	NA	NA	NA	23 (23–24)	23 (23–24)	22.9 (22–24)	0.66

Table describing the calculations done on the test datasets

P-values for categorical variables, such as sex, were calculated using Fisher's exact test, while for continuous variables, such as body mass index(BMI) and age, the Mann–Whitney U test was used

NA, not available

Demographic characteristics for the validation datasets were also calculated, which can be found in (Supplementary Table 2).

Data preprocessing for gastric cancer

In the Erawijantari et al. dataset, we excluded features with 80% sparsity from the microbiome data and those with 20% from the metabolome data. The remaining features were subjected to min–max scaling. During PCoA, we identified one sample in the microbiome data as an extreme outlier based on its visual distance from the main cluster, which was subsequently removed. Similarly, two samples in the metabolome data were identified as outliers based on their visual deviation from the main cluster in the PCA graph. They were excluded, and this resulted in a final dataset of 95 microbiome samples and 94 metabolome samples. Using the Mann–Whitney U test (FDR-adjusted $p < 0.05$), we identified 140 significant features from the microbiome data and 146 from the metabolome data, which were ultimately used for further analysis (Fig. 1b).

Model performance across multiple models for gastric cancer and validation

We employed three models, XGBoost, Random Forest, and LASSO, separately on the GC microbiome and metabolome datasets, The models were hyper-tuned through random search and Bayesian optimization. For feature selection, we applied RFECV with tenfold cross-validation for XGBoost and Random Forest, while LASSO utilised its built-in feature selection. This process identified 59 microbes and 45 metabolites that were common across all models, which we used to train the classifiers. Performance metrics were calculated based on test scores.

For the microbiome data, the Random Forest model performed best with an AUC of 0.96 (0.86–1.00) and an accuracy of 88% with the scores and hyper-tuned parameters of the model depicted in Supplementary Tables 3&4.To further reduce the number of microbes, the Gini-importance scores of the microbes were noted, and a subset of the top 15 microbes provided the best ROC-AUC score of 97%.

These microbes at the genus level were subjected to a Spearman correlation cluster map. From the clusters, we identified 6 microbes, mainly *CAG-103*, *Ruminococcus*, *Olleya*, *Cutibacterium*, *Allisonella*, and *Centipeda* (Supplementary Fig. 2a).

For the metabolome data, LASSO was the top performer, with an AUC of 0.98 (0.91–1.00) and an accuracy of 92%. Further feature selection identified a subset of 30 metabolites that had an AUC score of 98%. From the cluster map, 8 metabolites were identified, mainly dihydrouracil, taurine, γ-butyrobetaine, pimelate, glycocholate, methionine sulfoxide, phenethylamine, and citramalate (Supplementary Fig. 2b) (Fig. 1c).

We choose to validate the top 15 microbes and top 30 metabolites across all three models. We validated the GC microbiome model using data from Jaeyun Sung et al., where Random Forest performed best, with an AUC of 0.88 (0.85–0.99). However, for metabolite validation using the UK Biobank, the AUC scores were lower than expected, with all three models showing an AUC of 0.50 (0.50–0.50) while all other performance metrics remained similar. The reduced AUC scores, despite high sensitivity and recall, are likely due to differences in sample type. The UK Biobank metabolites were derived from plasma samples, whereas the main model was trained using metabolite data extracted from fecal samples. This could have potentially limited the model's predictive power (Supplementary Table 5) (Fig. 1d).

Microbial diversity & abundance analysis for gastric cancer

Alpha diversity analysis using the selected six microbes revealed significant differences between GC patients

(See figure on next page.)

Fig. 1 Microbiome-metabolome machine learning for cross-disease predictions in GC. **a** Fecal microbiome and metabolome data from GC patients (orange) and healthy individuals(green) obtained from Erawijantari et al. **b** Data preprocessing workflow highlighting the key microbes, metabolites, and samples selected for machine learning, alongside a principal coordinates analysis (PCoA) plot used for outlier removal. **c** The receiver operator curve – area under the curve (ROC-AUC) for microbiome and metabolome data across models: XGBoost (blue), Random Forest (green), and LASSO (red). Bar graph showing the best-performing model (microbiome-Random Forest, metabolome-LASSO) based on the highest AUC-ROC score, highlighting the optimal number of features. The selection includes 6 microbial and 8 metabolite features identified through Spearman cluster map analysis. **d** Validation performance metrics of the optimal features depicted by bar plots for microbiome and metabolome analysis were evaluated using the microbiome dataset from Jaeyun Sung et al. and the metabolome dataset from the UKBB. **e** Alpha diversity for microbes was visualised with violin plots comparing healthy and GC patients using the Shannon and Gini-Simpson indices. FDR-corrected p-values ($p < 0.05$) showed significant differences within both groups. Beta diversity was evaluated using non-metric multidimensional scaling (NMDS) based on Jaccard distances, with the stress value confirming statistical significance between healthy and diseased patients. **f** Circular bar plots illustrate the performance scores of the three models trained using combined microbiome and metabolome data from GC patients. Key biomarkers from the GC dataset were identified in the IBD and CRC datasets. GC-trained models were applied to predict IBD and CRC outcomes respectively

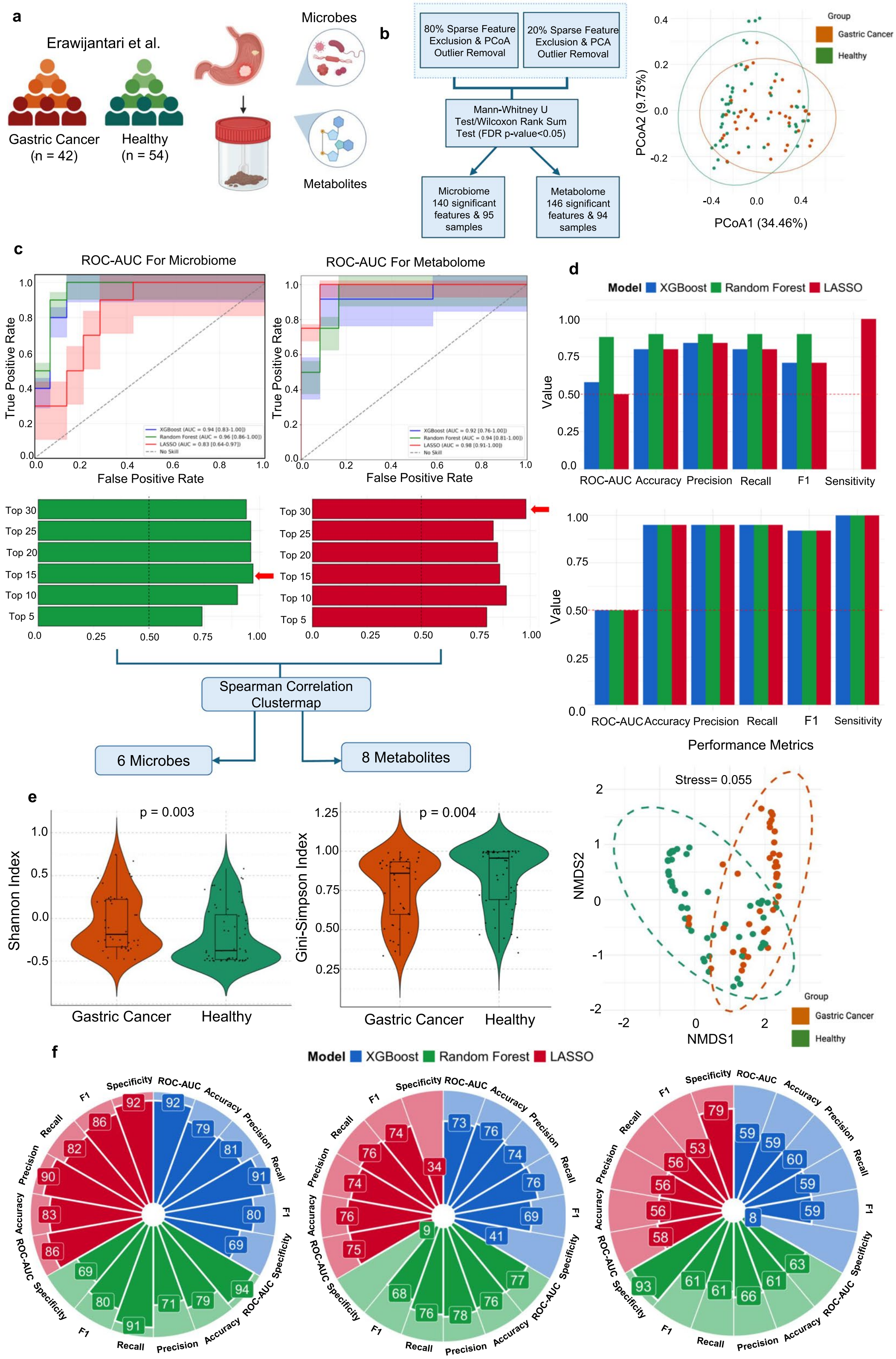

Fig. 1 (See legend on previous page.)

post-gastrectomy and healthy individuals. Shannon index (p =0.003) and Gini-Simpson index (p =0.004) consistently indicated distinct microbial diversity within the groups. Similarly, beta diversity analysis, using NMDS with Jaccard distances (stress =0.055), effectively demonstrated microbial differences between the groups (Fig. 1e).

Using gastric cancer biomarkers to predict inflammatory bowel disease and colorectal cancer

To explore cross-disease applicability, we extended our analysis from GC to include IBD and CRC. We applied the GC model, trained on the selected biomarkers, to predict IBD and classify CRC patients as either diseased or non-diseased (non-IBD and non-CRC). Among the GC models, Random Forest performed the best, achieving a ROC-AUC score of 0.94 (0.83–1.00). Surprisingly, the predictions on both IBD and CRC revealed the Random Forest model as the top-performing model, with an ROC-AUC score of 0.77 (0.71–0.83) and 0.63 (0.57–0.69), respectively (Supplementary Tables 6&7) (Fig. 1f).

Data preprocessing for colorectal cancer

For preprocessing the CRC dataset, we removed 70% of the sparse features from the microbiome dataset and 80% of the sparse features from the metabolome data as an initial step. After applying min–max scaling to the remaining features, we generated PCoA and PCA plots to identify and remove outliers. This resulted in the exclusion of 10 samples from the microbiome dataset and 12 samples from the metabolome dataset. Since the Wilcoxon test failed to identify significant microbes and metabolites, we applied the Kruskal–Wallis test to filter out insignificant features. Ultimately, 208 microbes across 267 samples and 105 metabolites across 265 samples were used as inputs for the machine learning models (Supplementary Fig. 3).

Model performance across multiple models for colorectal cancer and validation

We trained hyperparameter-tuned machine learning models on the CRC microbiome and metabolome datasets. Feature selection resulted in 82 microbes and 72 metabolites that were consistently identified

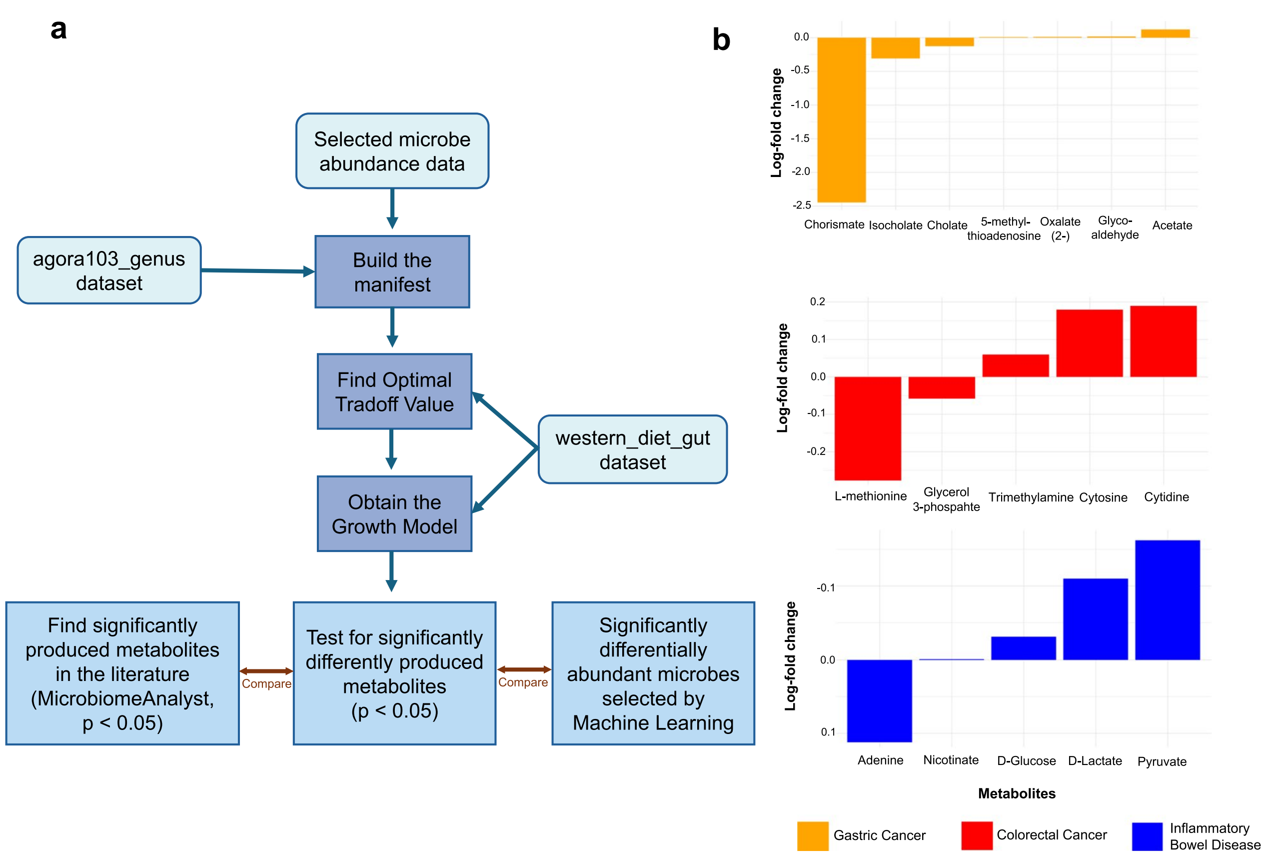

Fig. 2 Microbial community model (MICOM) results overview. **a** A summary of the process used to obtain the results. **b** The significantly differentially produced metabolites (p < 0.05) for each disease and their log-fold change abundance, where a positive change represents an increase in cases vs controls (Diseased vs Healthy)

across the three models. Among the models built for CRC microbes, Random Forest performed the best, achieving an AUC-ROC score of 0.89 (0.80–0.96) and an accuracy of 84%. However, when we focused on the top 15 microbes, the AUC score dropped significantly to 54%, with other subsets performing even worse. Despite this decrease, we decided to proceed with this subset to strike a balance between model performance and the identification of actionable biomarkers. From the heatmap, we identified 13 microbes such as *HGM04593, Parafilimonas, RUG11977, UMGS755, CACXMZ01, Desulfitobacterium, Onthousia, M0103, Fusobacterium, Enterococcus, Selenomonas, Psychrobacillus,* and *Thermus* (Supplementary Fig. 4a).

For the metabolome data, LASSO performed the best, with an AUC score of 0.70 (0.57–0.82). From the top 15 metabolites, which yielded the highest AUC score of 61%, we selected 10 metabolites such as isoleucine, N6-methyl-2-deoxyadenosine, N1,N8-diacetylspermidine, guanine, γ-guanidinobutyrate, dTMP, nicotinamide, decanoate, dodecanedioate, and 2-hydroxyoctanoate were selected for final analysis from the heatmap(Supplementary Fig. 4b)(Supplementary Tables 8&9).

To validate the CRC models, we used the microbiome and metabolome from Kim et al. XGBoost and Random Forest achieved an AUC of 0.51 (0.47–0.54) for the microbiome, while LASSO for metabolites achieved an AUC of 0.50 (0.39–0.61). The lower performance could likely stem from differences in particular biomarker

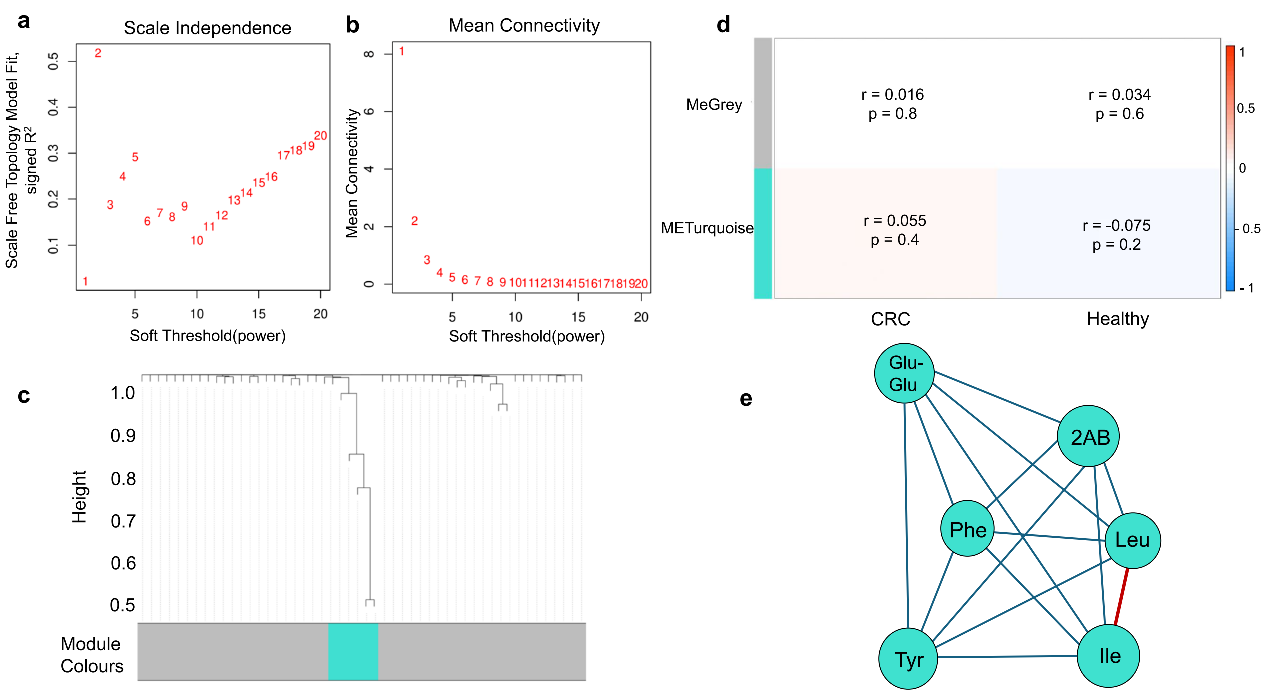

Fig. 3 Weighted Gene Co-expression Network Analysis (WGCNA) for CRC. **a** This plot shows the scale-free topology model fit (R^2) versus soft-thresholding power (β). The highest R^2 is 0.1847 at β = 9, indicating a weak but improving fit to the scale-free topology as β increases. **b** This plot displays how the mean connectivity decreases with increasing β. At β = 9, the mean connectivity is low, reflecting network sparsification while retaining some structural connections. **c** Shows a hierarchical clustering dendrogram of metabolites, where branches represent clusters of similar elements based on their co-expression. The height (Y-axis) indicates the dissimilarity between clusters, with smaller heights representing higher similarity. The horizontal bar below the dendrogram represents module assignments. The turquoise color indicates elements grouped into a co-expression module, while grey represents elements that were not assigned to any module due to low correlation or lack of clustering. **d** This heatmap represents the correlation between module eigengenes and traits (Case and Control, where Case = CRC and Control = Healthy). Each cell contains the correlation coefficient and its p-value with color intensity indicating the correlation's strength and direction (red for positive, blue for negative). The grey module, containing unassigned elements, shows a very weak positive correlation with CRC (r = 0.016, p = 0.8) and Healthy (r = 0.034, p = 0.6), both of which are statistically insignificant. The turquoise module, containing co-expressed elements, shows a weak positive correlation with CRC (r = 0.055, p = 0.4) and a weak negative correlation with Healthy (r = – 0.075, p = 0.2), neither of which are significant. This suggests no strong relationship between module expression and CRC. **e** The turquoise nodes in the network visualisation represent metabolites within the turquoise module, characterized by strong co-expression connections. The edges connecting turquoise nodes reflect the strength of co-expression: red edges represent higher strongly co-expression interactions, and blue edges indicate lower co-expression interactions. Metabolites like"Leu,"and"Ile"are central to this cluster, potentially functioning as hub metabolites coordinating module activity

profiles or technical variability between datasets (Supplementary Table 10).

Microbial diversity and abundance analysis For CRC

Alpha diversity indices such as the Shannon index ($p = 0.0617$) and Gini Simpson ($p = 0.524$) indicated no significant differences in microbial composition within the healthy and within the diseased groups. Similarly, beta diversity using the NMDS analysis had a stress value of 0.7, further confirming no significant diversity between healthy and diseased samples.

Using colorectal cancer biomarkers to predict gastric cancer and inflammatory bowel disease

We used the model trained on the selected combined biomarkers for CRC to predict and distinguish GC from non-GC patients and IBD from non-IBD patients. Within the CRC models, Random Forest achieved the highest ROC-AUC score of 0.75 (0.62–0.86). For GC predictions, Random Forest led with a top ROC-AUC of 0.86 (0.77–0.93), whereas for IBD predictions, the LASSO model produced the top AUC score of 0.65 (0.57–0.72) (Supplementary Tables 11, 12).

Data preprocessing for inflammatory bowel disease

For the IBD data, we removed 40% of sparse features from the microbiome dataset and 60% from the metabolome dataset, followed by min–max scaling of the remaining features. For outlier removal, we used PCoA on the microbiome data, identifying and removing nine outliers, and PCA on the metabolome data, removing seven outliers. After applying the Mann–Whitney U test, we identified 1089 significant microbiome features across 211 samples and 259 metabolome features across 213 samples, which were used for further analysis (Supplementary Fig. 5).

Model performance across multiple models for inflammatory bowel disease and validation

We individually trained the IBD microbiome and metabolome data using hyperparameter-tuned XGBoost, Random Forest, and LASSO models. Feature selection through RFECV and LASSO consistently identified 83 significant microbiome features and 73 metabolome features across all three models. For the microbiome data, Random Forest demonstrated the best performance, with an AUC of 0.90 (0.81–0.97) and an accuracy of 83%. The subset of the top 15 microbes had a top score of 89%, and the Spearman cluster map identified 9 microbes specifically, *Actinomarina*, *RGIG4708*, *Butyribacter, Limivivens, Faecalibaculum, UBA11774, UMGS1601, Bariatricus*, and *SIG607* (Supplementary Fig. 6a).

For the metabolome data, Random Forest was the top performer, with an AUC of 0.95 (0.89–0.99) and an accuracy of 89%. The subset of 20 metabolites had the top score of 88% and identified 10 metabolites such as urobilin, glycerate, cholestenone, acetyl-arginine, 4-hydroxy 3-methyl acetophenone, methylguanine, pseudouridine, inosine, 1,3,7-trimethyl urate, and carnosol based on the cluster map and Gini-importance scores (Supplementary Fig. 6b, Supplementary Tables 13, 14).

For validation of the IBD models, we used the Integrative Human Microbiome Project (iHMP) datasets, and for the microbes and metabolites, we have Random Forest as the best model with an AUC of 0.60 (0.50–0.64) and AUC of 0.76 (0.70–0.81), respectively. These microbes and metabolites can be investigated as potential biomarkers in IBD alone (Supplementary Table 15).

Microbial diversity and abundance analysis for inflammatory bowel disease

For IBD, the Shannon diversity index ($p = 0.617$), and Gini Simpson ($p = 0.525$) values for alpha diversity indicate that there are no microbial differences within the healthy and IBD groups. However, the beta diversity analysis, visualised through NMDS (stress value $= 0.05$), indicated distinct compositional differences between the groups.

Using inflammatory bowel disease biomarkers to predict gastric cancer and colorectal cancer

We used the model trained on the selected combined biomarkers for IBD to predict and distinguish GC from non-GC patients and CRC from non-CRC patients. Among the main IBD models, Random Forest had the highest AUC score of 0.93 (0.86–0.98). For the GC predictions, we noticed that Random Forest had the best AUC score of 0.66 (0.54–0.76), and for predictions on CRC, LASSO had a top score of 0.57 (0.51–0.63) (Supplementary Tables 16, 17).

Interactions between microbes and metabolites

An overview of the process used to obtain these results can be found in Fig. 2a. Of the 59, 82, and 83 initial genera chosen by the machine learning process, which were input into the MICOM manifest, 14, 14, and 5 were recognized by the AGORA database for GC, CRC, and IBD datasets, respectively. The optimal trade-off values for the created models were 0.8, 0.7, and 0.9 (Supplementary Table 18). Initial analysis of differentially produced metabolites between control and case groups identified 6, 7, and 16 metabolites for GC, CRC, and IBD, respectively, all of which are reported in detail within Supplementary Table 19. Of those we deemed significant, 65% of metabolites were of increased abundance in cases vs controls.

Chorismate ($p = 0.002$), isocholate ($p = 0.0002$), and cholate ($p = 0.02$) were decreased in abundance, while 5-methylthioadenosine ($p = 0.02$), oxalate ($p = 0.01$), glycolaldehyde ($p = 0.01$) and acetate ($p = 0.01$) were increased in abundance, for GC cases versus controls. L-methionine ($p = 0.004$) and glycerol 3-phosphate ($p = 0.008$) were in reduced abundance, along with trimethylamine ($p = 0.02$), cytosine ($p = 0.006$) and cytidine ($p = 0.03$) in increased abundance, for CRC cases vs controls (Fig. 2b).

Adenine ($p = 0.03$) was diminished in IBD cases, whilst nicotinate ($p = 0.04$), D-glucose ($p = 0.04$), D-lactate ($p = 0.05$), and pyruvate ($p = 0.04$) were elevated, in IBD cases versus controls (Fig. 2b). Consequently, the overlap between the machine learning analysis and MICOM selected metabolites defines cytidine, glycine, and methionine as important for CRC definition, while no overlap occurred for either IBD or GC.

Additionally, however, MICOM identified metabolite derivatives of those identified by the feature selection process, namely glutamate, glycerate N-acetyl-histidine and nicotinic acid for IBD; glycerophosphate for CRC; and 5-methyl-2-deoxycytidine, acetyl CoA, and glycocholate for GC.

MICOM further selected metabolites as differential in one disease, which were significantly identified by the machine learning algorithm for others. For instance, cholate and cytosine were differential for IBD analysis but were selected as significant for GC and CRC, respectively, by MICOM. Similarly, alanine, glutamate, histidine, lactate, and tryptophan were differential for IBD and 1-methyladenosine differential for CRC as chosen by MICOM, while they were selected as important for CRC by the machine learning process. Finally, alanine and nicotinate were selected for IBD, glycerophosphate for CRC, and methionine for both IBD and CRC by MICOM, whereas they were discriminatory for GC in the machine learning models.

MicrobiomeAnalyst also identified similar metabolites as significantly associated with each disease: D-glucose, glycine, histamine, and L-alanine for IBD; L-phenylalanine, L-leucine, azelaic acid, cholic acid, and lactic acid for GC; and isoleucine, methionine, butyric acid, D-glucose, Serine and 3-methylhistidine for CRC.

Weighted gene co-expression network analysis

For WGCNA, although different co-expressed metabolite modules were identified for GC, CRC, and IBD when analyzed with 45, 72, and 73 metabolites, respectively, no modules were detected for the microbiome-metabolite dataset. This could be explained by the fact that, in contrast to the more stable and evolutionarily conserved gene networks, interactions between the microbiota and metabolites are highly variable and transient. Nevertheless, to determine an appropriate stable configuration for the study, an adequate level of network connectivity was chosen. Furthermore, co-expression networks were shown, emphasizing the modular structures and metabolite interactions in each dataset.

For the CRC metabolite dataset, a power of $\beta = 9$ was selected, yielding an R^2 of 0.1847 (Fig. 3a&b). Among the co-expressed metabolite modules identified, the turquoise module stood out as a unique cluster strongly associated with CRC which included metabolites like Glu-Glu, isoleucine (Ile), 2 AB, phenylalanine (Phe), leucine (Leu), and tyrosine (Tyr) (Fig. 3c).

A heatmap visualising the correlation between module eigengenes (ME) and clinical characteristics (CRC vs. Healthy, denoted as cases vs. controls revealed that the turquoise module showed a modest positive association with CRC cases (correlation $= 0.055$, $p = 0.4$) and a slight negative correlation with healthy controls. (correlation $= -0.075$, $p = 0.2$) (Fig. 3d).

The network diagram illustrated the dense interconnections among metabolites in the turquoise module, with leucine (Leu) and isoleucine (Ile) occupying central positions as hub metabolites (Fig. 3e). Metabolites from other modules, which were not part of this co-expression cluster, were represented by grey nodes.

For the GC metabolite dataset, the turquoise module was significantly associated with GC in the WGCNA analysis, developed with a soft threshold power of $\beta = 9$ (scale-free topology $R^2 = 0.1947$) (Supplementary Fig. 7a, b). The turquoise module, represented a distinct cluster of co-expressed metabolites, including N-acetylglucosamine-1-phosphate, agmatine, 5-aminolevulinate, N8-acetylspermidine, inosine, nicotinamide, S-adenosylmethionine (SAM), dihydrouracil, and uracil are the nine metabolites that made up this module (Supplementary Fig. 7c). A heatmap revealed correlation between clinical characteristics (GC vs. Healthy) and MEs. The turquoise module showed a modest positive correlation with the Healthy group (correlation $= 0.073$, $p = 0.5$) and a slight negative correlation with the GC group (correlation $= -0.13$, $p = 0.2$), suggesting its potential role in distinguishing between cases and controls (Supplementary Fig. 7 d). In the network diagram, turquoise-colored nodes represent metabolites from the turquoise module, which is strongly linked to GC, while grey nodes denote metabolites from other modules. Key metabolites such as SAM, inosine (Ino), uracil (Ura), and N8-acetylspermidine (Agm) were moderately interconnected, highlighting their central roles in the turquoise module's metabolic network.

Finally, for IBD, a soft threshold power ($\beta = 10$) was selected to construct a scale-free topology network, balancing sparsity and robustness, with a model fit (R^2) of 0.235 (Supplementary Fig. 8a, b). Hierarchical clustering based on topological overlap identified distinct co-expression modules, with the turquoise module emerging as the most biologically significant (Supplementary Fig. 8c). The turquoise module showed a positive correlation with IBD, despite a modest eigengene-disease status association (correlation = 0.056, p = 0.4) (Supplementary Fig. 8 d).

This module comprised ten metabolites moderately associated with IBD, including sebacate, deoxycholic acid, 7-ketodeoxycholate, thiamine, cholestenone, pyridoxamine, undecanedionate, lithocholic acid, 4-hydroxy-3-methylacetophenone, and urobilin.

Discussion

Early GID diagnosis is essential for both preventing disease progression and developing efficient treatment plans that can enhance patient survival. Traditionally, each GID relies on its own'gold standard' diagnostic methods, such as endoscopy, medical imaging, and biopsies. While effective, these methods are often invasive, costly, might carry the risk of radiation exposure [62–64], and may not always detect the disease at an early stage. To tackle these challenges, researchers have explored biomarkers for early and accurate detection of GC, CRC, and IBD individually using genomic, transcriptomic, microbiome, and metabolomic datasets. Given that GIDs are often interconnected, the presence of one condition can increase the risk of developing another. This study examines whether microbes and metabolites linked to one disease could serve as early indicators for diagnosing others.

Biomarkers in gastric cancer

Biomarker identification largely depends on stable feature selection to ensure reliability. To achieve this, we employed multiple feature selection methods, including RFECV and LASSO-based selection, prioritizing top-ranked features with the highest discriminative scores and ensuring that selected features were not highly correlated with one another, with the help of the Spearman correlation map. Combining the selected microbes and metabolites for GC provided optimal performance scores of AUC > 0.8 (0.63–1.00) across all three machine learning models. Since the goal of this project is to use the primary GC model to predict CC and IBD and vice versa, we applied the same selected biomarkers to IBD and CC datasets. The results revealed that GC biomarkers might also be relevant for IBD, with all models achieving AUC > 0.7 (0.66–1.00). However, while the IBD models demonstrated high accuracy, precision, specificity, and F1 scores, their sensitivity was comparatively lower, indicating a reduced ability to identify all true positive cases. This trade-off essentially reflects the model's tendency to prioritize minimizing false positives over maximizing true positives. Similar observations were made by Hodgkiss et al. [65], who also noted low sensitivity in IBD prediction models.

In contrast, although the GC biomarkers performed well for IBD, they performed poorly in predicting CRC, with most models showing an AUC just above 0.58 (0.51–0.58), except for the Random Forest model, which achieved an AUC of 0.63 (0.57–0.69).

To reinforce our findings in the literature, we observed that the microbes associated with GC belonged to three major phyla, which included *Firmicutes*, *Bacteroidota*(also known as *Bacteroidetes*), and *Actinobacteria*. Tseng et al. [66] reported that these bacterial phyla were abundant in patients who had recently undergone gastrectomy, which aligns with our observations, as the data were collected from GC patients post-gastrectomy.

In our study, we identified microbes that belong to the *Lachnospiraceae* family, which is known to participate in the production of acetic acid and butyric acid. A reduction in the abundance of *Lachnospiraceae* was associated with altered lipid metabolism, increased inflammation, and malignancy in GC [67–69]. Additionally, the *Muribaculaceae* family showed a positive correlation to amino acid and glucose metabolism pathways related to GC [70, 71].

Studies reported conflicting effects of *Ruminococcus* on GC, with some studies indicating that certain species can be beneficial in reducing the risk of CRC and stabilising the intestinal barrier [72], while certain species of *Ruminococcus* can potentially increase the risk of developing GC [73]. Additionally, *Centipeda*, a microbe found significant in GC in our study, has been strongly associated with cancer virulent *H.pylori* [74]. Similarly, *Cutibacterium*, another significant microbe in GC, has been studied extensively for its role in promoting tumor formation in renal cell carcinomas [75] and has also been found to be abundant in GC as well [76, 77].

Regarding metabolites, dihydrouracil ranked highly in LASSO feature importance. Although not directly involved in GC causation, dihydrouracil plays a key role in pyrimidine metabolism, which, when disrupted, can lead to cancerous lesions [78–80]. Shentu et al. [81] revealed that taurine exhibits dual roles in GC disease progression, promoting tumor growth in immunodeficient mice while inhibiting it in immunocompetent mice.

Moreover, Sinha et al. [82], in their meta-analysis regarding the effects of taurine on CRC, noted that most studies report an increase in the taurine levels associated with the disease. γ-Butyrobetaine, which serves as a precursor in the formation of Trimethylamine N-Oxide (TMAO) by gut flora like *Firmicutes* and Actinobacteria, which is known to be associated with the development of GC [83, 84]. Secondary bile acid glycocholate, also known as glycocholic acid, was seen to be increased in patients with GC [85] and UC [86]. In contrast, MICOM analysis identified decreased derivatives of the bile acid glycocholate, namely chorismate, isocholate, and cholate, in GC cases vs controls. This was based on the metabolite flux predicted from the differential microbes. This could highlight that the increased abundance of bile acids is not due to their differential production by our selected microbes but may be due to other alterations in downstream cellular mechanisms [87]. Methionine sulfoxide, another key metabolite from GC machine learning analysis, was produced at lower levels by the selected microbes for IBD and CC cases in MICOM analysis. The promising use of this metabolite as an adjuvant researched in all three disorders [88–90] gives insight into the potential of this pathway as a link in their pathology.

Furthermore, acetate, a derivative involved in the acetyl CoA pathway that has been linked to cancer cell growth [91], was predicted to be significantly increased in GC samples based on microbial abundance in MICOM. Additional metabolites associated with the differential taxa from machine learning analysis and an increase in GC disease by MICOM included oxalate and glycolaldehyde linked with the development of renal dysfunction in GC patients [92], and GC metastasis [93], respectively.

Biomarkers in colorectal cancer

The combined model incorporating both microbes and metabolites for the CRC dataset achieved an AUC >0.7 (0.57–0.86) across XGBoost, Random Forest, and LASSO models. When this model was applied to the GC and IBD datasets, its performance was notably better for GC, achieving an AUC >0.7 (0.58–0.93) across all three models. This suggests that the biomarkers identified for CRC may also be relevant for GC.

Microbes identified from the CRC dataset were predominantly from the phylum *Firmicutes*, followed by *Bacteroidetes*, *Fusobacteriota*, *Actinobacteriota*, and *Deinococcota*. Interestingly, although 7 of the 13 microbes belonged to the *Firmicutes* phylum, studies have reported conflicting findings regarding its abundance in CRC patients. Some studies suggest a reduction in *Firmicutes* abundance in CRC patients, particularly those species involved in butyrate production [94, 95]. In contrast, other studies indicate an increase in *Firmicutes* along with *Bacteroidetes*, *Fusobacteriota*, *Actinobacteriota*, and *Deinococcota*, which have been found to be more prevalent and abundant in CRC patients [96–99].

At the genus level, *Fusobacterium* emerged as a key bacterium frequently associated with periodontal diseases [100]. Recognized for its pro-inflammatory properties, *Fusobacterium* was found to be more abundant in advanced stages of both CRC and GC [101, 102]. Additionally, *Fusobacterium* is linked to the production of hydrogen sulfide, which plays a role in the synthesis of sulfur-containing amino acids, a process implicated in the initiation of CRC [103]. This highlights its potential role in driving disease progression.

Furthermore, species from the genus *Enterococcus*, known for their production of reactive oxygen species (ROS), were observed in both colonic and gastric epithelial tissues, implicating their role in epithelial damage and tumorigenesis in both CRC and GC [104, 105]. Genera like *Selenomonas* [106, 107] and *Thermus* [97, 108] were also seen to be abundant in patients with both CRC and GC, reinforcing their potential as biomarkers.

In our study, we identified isoleucine, a branched-chain amino acid (BCAA), which exhibits a complex and dual role in tumor progression. Some studies on CRC suggest that isoleucine promotes tumor growth by participating in biosynthetic pathways as an intermediate in the TCA cycle that supplies energy and contributes to oncogenic mutations [109]. Additionally, Ren et al. [110] demonstrated that *Clostridium symbiosum* produces BCAAs such as isoleucine, which enhance cholesterol synthesis, a process implicated in CRC progression. In contrast, other studies on CRC [111] and GC [112] reported a protective role for isoleucine, suggesting it may inhibit tumor formation and emphasizing its dual role in cancer biology, warranting further investigation.

In the WGCNA analysis conducted, no strong interconnections between metabolites were observed in CRC. However, network analysis identified leucine and isoleucine as hub metabolites, suggesting their involvement in a tightly interconnected metabolic network. A recent study in mouse models found that the breakdown of leucine and isoleucine played a crucial role in the development of CRC, with elevated levels of these metabolites found in CRC tumor tissues compared to normal tissues. This indicates that impaired breakdown of BCAAs supports cancer cell proliferation by providing essential nutrients for tumor growth. In CRC, the normal degradation of the BCAAs is disrupted due to the downregulation of proteins involved in their breakdown. As a result, these amino acids accumulate, promoting cancer cell metabolism and growth [113]. A study by

Wang et al. [114], revealed that B cells enriched in CRC tissues with transforming growth factor-β1(TGF-β1) dominant regulatory phenotypes are driven by leucine nutritional preferences and accelerate CRC growth. Leucine promotes tumor evasion by inducing leucine-tRNA-synthetase-2 expressing B cell (LARS B) which inhibits mitochondrial NAD + regeneration and oxidative metabolism, leading to increased TGF-β1 production.

The metabolite nicotinamide also showed interesting results. While essential for normal cellular metabolism, its abundance in CRC may confer a survival advantage to cancer cells, as suggested by Jabbari et al. [115]. Similarly, targeting nicotinamide metabolism in GC patients may improve prognosis [116]. Other metabolites demonstrated disease relevance as well. Decanoate (capric acid), a medium-chain fatty acid, showed anticarcinogenic properties in CRC [117]. Elevated levels of N1, N8-diacetylspermidine were observed in CRC patients [118], while metabolites such as guanine [119, 120] and γ-guanidinobutyrate [121, 122] were identified in both CRC and GC as potential biomarkers. Cytidine, which is strongly linked to gut inflammation in CRC [123], was proposed to be increased in CRC cases according to both MICOM and machine learning analysis, whereas glycine, which has the potential to decrease CRC tumor volume and vascularization [124], was decreased in abundance. Additionally, metabolites associated with the selected microbes in CRC analysis, such as trimethylamine and cytosine, were predicted to be of higher abundance in CRC samples. An increase of trimethylamine was also observed in CRC patients by Guo et al. [125] and linked to dysbiosis by Chan et al. [126], while the increased activity of cytosine activated pathways has been associated with CRC progression [127]. Furthermore, metabolites identified as differential for CRC in early machine learning analysis such as glutamate and alanine were selected as differential in IBD patients in MICOM analysis. Glutamate is strongly related to the maintenance of the mucosal lining, with its disruption contributing to IBD and gastrointestinal cancer [128], while amino acids such as alanine and leucine are key to mucosal healing after destruction. Their low abundance can lead to issues in both CRC and IBD [129, 130].

Biomarkers in inflammatory bowel disease

The combined model scores for IBD, which incorporated both microbes and metabolites, showed satisfactory results with AUC values > 0.84 (0.71–0.98). However, when applied to both GC and CRC, the predictions were suboptimal. While GC showed slightly better performance, with AUC scores > 0.60 (0.53–0.76) for Random Forest and LASSO, other performance metrics did not perform as well.

At the phylum level, most of the microbes identified in IBD belonged to the *Firmicutes* and *Actinobacteriota*. These phyla, typically involved in the breakdown of short-chain fatty acids (SCFAs) into butyrate and other beneficial products, were reported to be reduced in the gut microbiomes of IBD patients in a study by Tsai et al. [131] Conversely, Santoru et al. [132] observed an increase in these phyla in IBD patients, highlighting the need for further investigation by looking more into the genera and the family levels. As many of the microbes were unclassified at the genus level, we traced them to the family level classifications. Genera such as *Bariatricus*, *Butyribacter*, *Limivivens*, and *UBA11774*, all members of the *Lachnospiraceae* family, were consistently found to be decreased in abundance in IBD patients compared to healthy controls in multiple studies [133, 134]. Interestingly, certain species within *Lachnospiraceae* have been linked to tumor progression in GC [67]. At the genus level, we examined *Faecalibaculum*, which was identified in IBD was recognized for its probiotic properties, SCFA production, and production of indole-3-lactic acid, which helps reduce colonic inflammation and repair the gut epithelial barrier [135, 136] However, Chen et al. [137] confirmed the presence of *Faecalibaculum* in cases of small bowel disease and severe gastritis. In terms of metabolites in IBD, urobilin, and acetyl-arginine have been identified as possible biomarkers in previous studies [138, 139]. Elevated levels of glyceric acid (glycerate) have been observed in the fecal samples of IBD patients, likely due to the breakdown of triacylglycerols released from the colon mucosa, which contributes to metabolic disruptions such as acidosis [132]. Notably, glycerate has also been implicated in GC, where it participates in the glycolysis pathway, providing energy to support tumor growth. This aligns with the Warburg effect, a metabolic hallmark of cancer, where cancer cells preferentially rely on glycolysis for energy production, even under oxygen-rich conditions, to fuel their rapid proliferation and survival [140, 141]. Furthermore, D-glucose and glutamate were significantly increased in IBD samples of MICOM analysis, increasing the promotion of Th17 cell differentiation and the activation of transforming growth factor β [142], which further promotes cell proliferation and inflammation and weakens the auto-immune response [143]. Heightened levels of lactate in IBD patients were also found by Song et al. [144], in addition to our MICOM analysis, while contrastingly, many other studies have determined lactate is beneficial for the intestinal barrier and a potential therapeutic target for IBD [145, 146]. This difference may be due to the fact that the differential microbes used to produce the MICOM model are highly lactate producing in normal circumstances, but may not have been in the guts of our patient cohort. Furthermore, nicotinate and pyruvate, also overabundant in IBD samples of the MICOM analysis, have been under investigation as therapeutic targets, with inhibitors of their metabolic pathways showing promising results [147, 148].

Metabolites with protective roles in the context of GIDs were also identified in this study. Inosine and carnosol, for instance, have been identified for their anti-inflammatory properties and for maintaining the intestinal barrier in both IBD and GC [149–152]. Similar patterns were observed in MICOM analysis. Moreover, adenine, which was found to be underabundant in IBD, activates receptors responsible for anti-inflammatory macrophages [153], potentially by inhibiting the tumour necrosis factor-α (TNF-α) induced interleukin-8 secretion pathway [154].

Machine learning for microbiome and metabolomic disease classification

In our study, Random Forest consistently emerged as the top-performing model for both individual microbiome, and combined models for GC, CRC, and IBD indicating high discriminatory power between the healthy and diseased samples. While high performance scores alone do not always correlate with better results, Random Forest is widely used in microbiome studies due to its ability to model complex nonlinear relationships between features and outcomes, as well as its robustness to noise [155]. For example, Gao et al. [156] demonstrated that the Random Forest based pipeline achieved the best classification performance for CRC prediction compared to other models, with performance metrics improving as the number of decision trees increased in the meta-dataset. Similarly, Zheng et al. [157], identified Random Forest as a reliable diagnostic model for distinguishing between healthy controls, CD, and UC patients based on gut microbiome data with AUC = 0.81 (0.80–0.82). Appiah et al. [158] also showed that Random Forest could accurately identify potential microbial signatures that distinguish healthy controls from GC samples.

On the other hand, LASSO emerged as the top model for metabolomic data in GC and CRC due to its ability to handle high dimensional data and shrink less important variable coefficients to zero, effectively selecting the most relevant metabolites for disease prediction.

Supporting this, Chen et al. [159] developed a 10-metabolite GC diagnostic model using LASSO, achieving a sensitivity of 0.9. Similarly, Sun et al. [160] successfully used LASSO regression to distinguish between healthy individuals and CRC patients with AUC = 0.96 based on plasma and fecal metabolites. Overlapping metabolites identified by the machine learning analysis and MICOM predicted models suggest potential links between differential microbes and the metabolites in disease. Furthermore, these interaction models provide the potential to design optimized fecal microbial transplant (FMT) treatments that not only address dysbiosis but also incorporate metabolite supplementation or degradation to enhance disease therapy [161].

Shared pathological mechanisms and biomarker insights in GC, CRC, and IBD

GC, CRC, and IBD exhibit distinct yet overlapping clinical and pathological characteristics that are often driven by shared clinical risk factors and underlying disease mechanisms. For instance, GC often presents with non-specific symptoms that include acid reflux, dysphagia, abdominal pain, bloating, indigestion, weight loss, and melaena [162]. The most common subtype, gastric adenocarcinoma, is known to frequently metastasize to the lymph nodes and liver. Several risk factors increase the likelihood of developing GC, including *H. pylori* infection, gastroesophageal reflux disease (GORD), family history, and poor diet [163, 164]. Furthermore, *H. pylori* is also linked to mucosa-associated lymphatic tissue (MALT) lymphoma in the stomach, primarily due to chronic inflammation and bacterial virulence factors [165].

CRC, predominantly in the form of colonic adenocarcinoma, presents with hallmark cancer symptoms such as weight loss, fatigue, and anemia [166]. However, it also presents more specific signs like rectal bleeding (haematochezia), tenesmus, and changes in bowel habits.

Certain inherited genetic conditions, such as Lynch syndrome (Hereditary Nonpolyposis Colon Cancer) and familial adenomatous polyposis (FAP), significantly increase the risk of developing CRC [16]. Unlike some cancers, CRC tends to metastasize more widely, often affecting the lungs, liver, and peritoneum [167]. Chronic diarrhea mixed with blood in stools is a hallmark of IBD and the following inflammation, especially in UC, predisposes patients to CRC through inflammation-induced DNA damage and immune dysregulation [7, 8, 32]. IBD and GC share immune dysregulation features, including imbalances in T-helper cells(*Th1*, *Th17*), which may also contribute to CRC progression [168–170]. CRC and GC both harbor mutations in key oncogenes like PIK3 CA and exhibit dysregulation in signaling pathways like Wnt/β-catenin contributing to cancer progression and resistance to therapy [171–174].

Moreover, chronic inflammation in IBD creates a pro-tumorigenic environment by increasing oxidative stress, which leads to DNA damage, p53 mutations, and microsatellite instability, persistent inflammatory cytokine signaling, ultimately facilitating the progression from dysplasia to CRC [175]. Understanding these shared mechanisms could aid in developing therapeutic approaches that target dysbiosis, promote beneficial microbes and metabolites, and modulate the harmful ones to help predict GIDs more effectively.

Limitations

One of the key limitations of our study was the failure to account for confounding factors such as age, gender, BMI, and diet in our predictive models. These variables can significantly influence model performance and introduce bias if not properly controlled [176]. Secondly, there was variability in the methods used for data collection, preprocessing, and analysis across different datasets and countries. This inconsistency can significantly impact the performance of the models [177], especially when integrating multiple datasets from diverse disease types. For instance, during validation, we observed suboptimal performance in some of our models, which we attribute to these methodological differences. To address this, establishing standardized protocols for data collection and analysis could greatly enhance the robustness and accuracy of machine learning models in future studies. Another limitation arose from the way we categorized diseases.

In this study, CD and UC were grouped under IBD, and different stages of CRC were analyzed collectively without distinguishing the adenoma specific data. This may have resulted in the loss of insights specific to individual disease subtypes or stages, potentially affecting the performance of our models. Additionally, while we used an independent dataset to validate the predictive power of our machine learning models, the validation process was limited to a small subset of the identified microbes and metabolites. This constraint highlights the need for larger, more diverse validation datasets to ensure the generalisability of our findings. Finally, our use of the MICOM model, which simulates microbial community interactions, was limited by the scope of the AGORA database. Since the database only includes a fraction of known microbes, only 6–24% of the differentially abundant microbes in our study could be incorporated into the disease models. This limitation underscores the importance of further research into the functional roles of unclassified microbes and their metabolic pathways. By expanding our understanding of these microbes, we can fill critical gaps in model creation and improve their predictive accuracy.

Future work

The use of a longitudinal study could enhance this study, as it would allow us to follow microbial and metabolomic changes in the gut at different stages, and we could gain a much clearer picture of how the disease starts and develops. Additionally, adding more GI disorder types can be incorporated in the future to develop a complete diagnostic model for different disorders with high specificity and sensitivity. This deeper understanding could help improve early detection and lead to more effective treatment strategies down the line.

Conclusion

Findings from our study suggest that differential microbes and metabolites associated with GC could also serve as potential biomarkers for predicting IBD. Interestingly, when we examined the microbes and metabolites linked to CRC, we found that they had a stronger predictive performance for GC than for IBD. These observations point to the possibility of overlapping disease pathways and biological mechanisms, supporting the idea that microbes and metabolites from one GID can be used to predict another. To further validate these findings, we cross referenced our identified microbes and metabolites with existing literature which reinforced the notion that certain biomarkers are shared across different GIDs. This opens up the possibility of developing diagnostic tools that could enhance our understanding and treatment of GIDs.

abbreviations

GIDs	Gastrointestinal disorders
GC	Gastric cancer
CRC	Colon cancer
IBD	Inflammatory bowel disease
XGBoost	EXtreme gradient boosting
LASSO	Least absolute shrinkage and selection operator
MICOM	Microbial community
WGCNA	Weighted Gene Co-expression Network Analysis
ROC-AUC	Receiver Operator Curve-Area Under the Curve
PCA	Principal components analysis
PcoA	Principal coordinates analysis
CD	Crohn's disease
UD	Ulcerative colitis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06552-w.

Additional file 1.

Additional file 2.

Acknowledgements
Not applicable.

Author contributions
DP: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing-original draft, Writing-review, and editing. RH: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing-original draft, Writing-review, and editing. SKR: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing-original draft, Writing-review, and editing. AS: Investigation, Validation, Writing-original draft, Writing-review, and editing. AA: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Writing-original draft, Writing-review, and editing. All authors read and approved the final manuscript.

Funding
MRC Heath Data Research UK and HDRUK midlands regional community project [QQ2], initiatives funded by UK Research and Innovation, Department of Health and Social Care (England) and the devolved administrations, and leading medical research charities. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research, the Medical Research Council, or the Department of Health.

Data availability
This paper analyzes existing, publicly available data. Information on access and associated publications is provided in Table 1. The Python and R packages used in this study can be found in Supplementary Table 1, and the underlying code generated for this study has been deposited in Figshare and can be accessed via this link: https://doi.org/https://doi.org/10.6084/m9.figshare.28464557.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that this research was conducted without any competing personal or financial relationships that could be considered potential influences on the work presented in this paper.

Author details
[1]Cancer and Genomic Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham Dubai, Dubai, UAE. [2]Cancer and Genomic Sciences, School of Medical Sciences, College of Medicine and Health, University of Birmingham, Birmingham, UK. [3]Centre for Health Data Research, University of Birmingham, Birmingham, UK. [4]Institute of Translational Medicine, University Hospitals Birmingham NHS, Foundation Trust, Birmingham, UK.

Received: 16 March 2025 Accepted: 4 May 2025
Published online: 16 May 2025

References

1. Ogobuiro I, Gonzales J, Shumway KR, et al. Physiology, gastrointestinal. Treasure Island: StatPearls; 2023.
2. Morgado-Diaz JA. Gastrointestinal cancers 2022. Brisbane: Exon Publications; 2022. https://doi.org/10.36255/EXON-PUBLICATIONS-GASTROINTESTINAL-CANCERS.
3. Ranjbar R, Ghasemian M, Maniati M, et al. Gastrointestinal disorder biomarkers. Clin Chim Acta. 2022;530:13–26. https://doi.org/10.1016/J.CCA.2022.02.013.
4. Wroblewski LE, Peek RM, Wilson KT. *Helicobacter pylori* and gastric cancer: factors that modulate disease risk. Clin Microbiol Rev. 2010;23:713. https://doi.org/10.1128/CMR.00011-10.
5. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229–63. https://doi.org/10.3322/CAAC.21834.
6. Morgan E, Arnold M, Camargo MC, et al. The current and future incidence and mortality of gastric cancer in 185 countries, 2020–40: a population-based modelling study. EclinicalMedicine. 2022. https://doi.org/10.1016/J.ECLINM.2022.101404.
7. Dahal RH, Kim S, Kim YK, et al. Insight into gut dysbiosis of patients with inflammatory bowel disease and ischemic colitis. Front Microbiol. 2023;14:1174832. https://doi.org/10.3389/FMICB.2023.1174832/BIBTEX.
8. Guan Q. A comprehensive review and update on the pathogenesis of inflammatory bowel disease. J Immunol Res. 2019;2019:7247238. https://doi.org/10.1155/2019/7247238.
9. Alavinejad P, Hashemi SJ, Behl N, et al. Inflammatory bowel disease evolution in the past two decades: a chronological multinational study. EClinicalMedicine. 2024. https://doi.org/10.1016/j.eclinm.2024.102542.
10. Wang R, Li Z, Liu S, et al. Global, regional and national burden of inflammatory bowel disease in 204 countries and territories from 1990 to 2019: a systematic analysis based on the Global burden of disease study 2019. BMJ Open. 2023;13: e065186. https://doi.org/10.1136/BMJOPEN-2022-065186.
11. Bertuccio P, Rosato V, Andreano A, et al. Dietary patterns and gastric cancer risk: a systematic review and meta-analysis. Ann Oncol. 2013;24:1450–8. https://doi.org/10.1093/ANNONC/MDT108.
12. Owczarek D, Rodacki T, Domagała-Rodacka R, et al. Diet and nutritional factors in inflammatory bowel diseases. World J Gastroenterol. 2016;22:895. https://doi.org/10.3748/WJG.V22.I3.895.
13. Zargar T, Kumar D, Sahni B, et al. Dietary risk factors for colorectal cancer: a hospital-based case-control study. Cancer Res Statist Treat. 2021;4:479–85. https://doi.org/10.4103/CRST.CRST_116_21.
14. Slavin TP, Weitzel JN, Neuhausen SL, et al. Genetics of gastric cancer: what do we know about the genetic risks. Transl Gastroenterol Hepatol. 2019;4:55. https://doi.org/10.21037/TGH.2019.07.02.
15. El Hadad J, Schreiner P, Vavricka SR, et al. The genetics of inflammatory bowel disease. Mol Diagn Ther. 2024;28:27–35. https://doi.org/10.1007/S40291-023-00678-7/METRICS.
16. Munteanu I, Mastalier B. Genetics of colorectal cancer. J Med Life. 2014;7:507. https://doi.org/10.1016/b978-0-12-091075-5.50016-0.
17. Durko L, Malecka-Panas E. Lifestyle modifications and colorectal cancer. Curr Colorectal Cancer Rep. 2014;10:45. https://doi.org/10.1007/S11888-013-0203-4.
18. Rozich JJ, Holmer A, Singh S. Effect of lifestyle factors on outcomes in patients with inflammatory bowel diseases. Am J Gastroenterol. 2020;115:832. https://doi.org/10.14309/AJG.0000000000000608.
19. Ko KP. Risk factors of gastric cancer and lifestyle modification for prevention. J Gastric Cancer. 2023;24:99. https://doi.org/10.5230/JGC.2024.24.E10.
20. Guinane CM, Cotter PD. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. Therap Adv Gastroenterol. 2013;6:295. https://doi.org/10.1177/1756283X13482996.
21. Zeng R, Gou H, Lau HCH, et al. Stomach microbiota in gastric cancer development and clinical implications. Gut. 2024;73:2062–73. https://doi.org/10.1136/GUTJNL-2024-332815.
22. Villéger R, Lopès A, Veziant J, et al. Microbial markers in colorectal cancer detection and/or prognosis. World J Gastroenterol. 2018;24:2327. https://doi.org/10.3748/WJG.V24.I22.2327.
23. Ma J, Wang K, Wang J, et al. Microbial disruptions in inflammatory bowel disease: a comparative analysis. Int J Gen Med. 2024;17:1355–67. https://doi.org/10.2147/IJGM.S448359.
24. Kaji S, Irino T, Kusuhara M, et al. Metabolomic profiling of gastric cancer tissues identified potential biomarkers for predicting peritoneal

recurrence. Gastric Cancer. 2020;23:874–83. https://doi.org/10.1007/S10120-020-01065-5.
25. Huang S, Guo Y, Li Z, et al. A systematic review of metabolomic profiling of gastric cancer and esophageal cancer. Cancer Biol Med. 2020;17:181. https://doi.org/10.20892/J.ISSN.2095-3941.2019.0348.
26. Zhang W, An Y, Qin X, et al. Gut microbiota-derived metabolites in colorectal cancer: the bad and the challenges. Front Oncol. 2021;11: 739648. https://doi.org/10.3389/FONC.2021.739648.
27. Zheng L, Wen XL, Duan SL. Role of metabolites derived from gut microbiota in inflammatory bowel disease. World J Clin Cases. 2022;10:2660. https://doi.org/10.12998/WJCC.V10.I9.2660.
28. Gasaly N, de Vos P, Hermoso MA. Impact of bacterial metabolites on gut barrier function and host immunity: a focus on bacterial metabolism and its relevance for intestinal inflammation. Front Immunol. 2021;12: 658354. https://doi.org/10.3389/FIMMU.2021.658354/BIBTEX.
29. Huang S, Zhou C, Wang B, et al. Recent advances in oral drug delivery materials for targeted diagnosis or treatment of gastrointestinal diseases. J Drug Deliv Sci Technol. 2023;88: 104903. https://doi.org/10.1016/J.JDDST.2023.104903.
30. Fretwell V, Kane E, MacPherson S, et al. Metastases from gastric cancer presenting as colorectal lesions: a report of two cases and systematic review. Annals. 2023. https://doi.org/10.1308/RCSANN.2023.0023.
31. Tak DH, Moon HS, Kang SH, et al. Prevalence and risk factors of gastric adenoma and gastric cancer in colorectal cancer patients. Gastroenterol Res Pract. 2016;2016:2469521. https://doi.org/10.1155/2016/2469521.
32. Sato Y, Tsujinaka S, Miura T, et al. Inflammatory bowel disease and colorectal cancer: epidemiology, etiology, surveillance, and management. Cancers (Basel). 2023;15:4154. https://doi.org/10.3390/CANCERS15164154.
33. Cammarota G, Ianiro G, Ahern A, et al. Gut microbiome, big data and machine learning to promote precision medicine for cancer. Nat Rev Gastroenterol Hepatol. 2020;17:635–48. https://doi.org/10.1038/s41575-020-0327-3.
34. Caballé NC, Castillo-Sequera JL, Gómez-Pulido JA, et al. Machine learning applied to diagnosis of human diseases: a systematic review. Appl Sci. 2020;10:5135. https://doi.org/10.3390/APP10155135.
35. Radhakrishnan SK, Nath D, Russ D, et al. Machine learning-based identification of proteomic markers in colorectal cancer using UK Biobank data. Front Oncol. 2024;14:1505675. https://doi.org/10.3389/FONC.2024.1505675/BIBTEX.
36. Bravo-Merodio L, Acharjee A, Russ D, et al. Translational biomarkers in the era of precision medicine. Adv Clin Chem. 2021;102:191–232. https://doi.org/10.1016/BS.ACC.2020.08.002.
37. Jayakrishnan TT, Sangwan N, Barot SV, et al. Multi-omics machine learning to study host-microbiome interactions in early-onset colorectal cancer. NPJ Prec Oncol. 2024;8:1–8. https://doi.org/10.1038/s41698-024-00647-1.
38. Liñares-Blanco J, Fernandez-Lozano C, Seoane JA, et al. Machine learning based microbiome signature to predict inflammatory bowel disease subtypes. Front Microbiol. 2022;13: 872671. https://doi.org/10.3389/FMICB.2022.872671/BIBTEX.
39. Freitas P, Silva F, Sousa JV, et al. Machine learning-based approaches for cancer prediction using microbiome data. Sci Rep. 2023;13:1–15. https://doi.org/10.1038/s41598-023-38670-0.
40. Liu Y, Du W, Guo Y, et al. Identification of high-risk factors for recurrence of colon cancer following complete mesocolic excision: an 8-year retrospective study. PLoS ONE. 2023. https://doi.org/10.1371/JOURNAL.PONE.0289621.
41. de Amorim LBV, Cavalcanti GDC, Cruz RMO. The choice of scaling technique matters for classification performance. Appl Soft Comput. 2023;133: 109924. https://doi.org/10.1016/J.ASOC.2022.109924.
42. Maćkiewicz A, Ratajczak W. Principal components analysis (PCA). Comput Geosci. 1993;19:303–42. https://doi.org/10.1016/0098-3004(93)90090-R.
43. Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers. Facta Univ Ser Math Inform. 2019. https://doi.org/10.22190/FUMI1903583G.
44. Gower JC. Principal coordinates analysis. Encycl Biostat. 2005. https://doi.org/10.1002/0470011815.B2A13070.
45. Ricotta C, Pavoine S. A new parametric measure of functional dissimilarity: Bridging the gap between the Bray-Curtis dissimilarity and the Euclidean distance. Ecol Modell. 2022;466: 109880. https://doi.org/10.1016/J.ECOLMODEL.2022.109880.
46. Nachar N. The Mann-Whitney U: a test for assessing whether two independent samples come from the same distribution. Tutor Quant Methods Psychol. 2008;4:13–20. https://doi.org/10.20982/TQMP.04.1.P013.
47. Ostertagová E, Ostertag O, Kováč J. Methodology and application of the Kruskal–Wallis test. Appl Mech Mater. 2014;611:115–20. https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMM.611.115.
48. Haynes W. Benjamini-Hochberg Method. Encycl Syst Biol. 2013. https://doi.org/10.1007/978-1-4419-9863-7_1215.
49. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International conference on knowledge discovery and data mining. 2016. pp. 13–17. https://doi.org/10.1145/2939672.2939785/SUPPL_FILE/KDD2016_CHEN_BOOSTING_SYSTEM_01-ACM.MP4
50. Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324/METRICS.
51. Tang J, Alelyani S, Liu H. Feature selection for classification: a review. Data Classification. 2014. https://doi.org/10.1201/B17320.
52. Bergstra J, Ca JB, Ca YB. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13:281–305.
53. Snoek J, Larochelle H, Adams RP. Practical Bayesian optimization of machine learning algorithms. Adv Neural Inf Process Syst. 2012;4:2951–9.
54. Bradley AP. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. 1997;30:1145–59. https://doi.org/10.1016/S0031-3203(96)00142-2.
55. Andermann T, Antonelli A, Barrett RL, et al. Estimating alpha, beta, and gamma diversity through deep learning. Front Plant Sci. 2022;13: 839407. https://doi.org/10.3389/FPLS.2022.839407/BIBTEX.
56. Ortiz-Burgos S. Shannon-Weaver Diversity Index. Encycl Earth Sci Ser. 2016. https://doi.org/10.1007/978-94-017-8801-4_233.
57. Somerfield PJ, Clarke KR, Warwick RM. Simpson Index. Encycl Ecol. 2008;1–5:3252–5. https://doi.org/10.1016/B978-008045405-4.00133-6.
58. Dexter E, Rollwagen-Bollens G, Bollens SM. The trouble with stress: a flexible method for the evaluation of nonmetric multidimensional scaling. Limnol Oceanogr Methods. 2018;16:434–43. https://doi.org/10.1002/LOM3.10257.
59. Diener C, Gibbons SM, Resendis-Antonio O. MICOM: metagenome-scale modeling to infer metabolic interactions in the gut microbiota. Msystems. 2020. https://doi.org/10.1128/MSYSTEMS.00606-19.
60. Diener C. Artifacts for q2-micom 2020. https://doi.org/10.5281/ZENODO.3755182.
61. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008;9:1–13. https://doi.org/10.1186/1471-2105-9-559/FIGURES/4.
62. Yücel M, Demirpolat MT, Yildirak MK. Colorectal cancer screening; colonoscopy and biopsy results in people undergoing colonoscopy due to positive fecal occult blood test. Turk J Surg. 2024;40:59–64. https://doi.org/10.47717/TURKJSURG.2024.6352.
63. Clin M, Lin EC. Radiation risk from medical imaging. Mayo Clin Proc. 2010;85:1142. https://doi.org/10.4065/MCP.2010.0260.
64. Moore LE. The advantages and disadvantages of endoscopy. Clin Tech Small Anim Pract. 2003;18:250–3. https://doi.org/10.1016/S1096-2867(03)00071-9.
65. Hodgkiss R, Acharjee A. Unravelling metabolite-microbiome interactions in inflammatory bowel disease through AI and interaction-based modelling. Biochim Biophys Acta Mol Bas Dis. 2025;1871:167618. https://doi.org/10.1016/J.BBADIS.2024.167618.
66. Tseng CH, Lin JT, Ho HJ, et al. Gastric microbiota and predicted gene functions are altered after subtotal gastrectomy in patients with gastric cancer. Sci Rep. 2016;6:1–8. https://doi.org/10.1038/srep20701.
67. Park JY, Seo H, Kang CS, et al. Dysbiotic change in gastric microbiome and its functional implication in gastric carcinogenesis. Sci Rep. 2022;12:4285. https://doi.org/10.1038/S41598-022-08288-9.
68. Vacca M, Celano G, Calabrese FM, et al. The controversial role of human gut lachnospiraceae. Microorganisms. 2020;8:573. https://doi.org/10.3390/MICROORGANISMS8040573.

69. Cui MY, Yi X, Zhu DX, et al. The role of lipid metabolism in gastric cancer. Front Oncol. 2022;12: 916661. https://doi.org/10.3389/FONC.2022.916661.
70. Yang Y, Dai D, Jin W, et al. Microbiota and metabolites alterations in proximal and distal gastric cancer patients. J Transl Med. 2022;20:1–14. https://doi.org/10.1186/S12967-022-03650-X/FIGURES/6.
71. Chung YW, Gwak HJ, Moon S, et al. Functional dynamics of bacterial species in the mouse gut microbiome revealed by metagenomic and metatranscriptomic analyses. PLoS ONE. 2020;15: e0227886. https://doi.org/10.1371/JOURNAL.PONE.0227886.
72. Crost EH, Coletto E, Bell A, et al. *Ruminococcus gnavus*: friend or foe for human health. FEMS Microbiol Rev. 2023;47:1–23. https://doi.org/10.1093/FEMSRE/FUAD014.
73. Zhang J, Dong C, Lin Y, et al. Causal relationship between gut microbiota and gastric cancer: a two-sample Mendelian randomization analysis. Mol Clin Oncol. 2024;20:38. https://doi.org/10.3892/MCO.2024.2736.
74. Rabenhorst SHB, Ferrasi AC, de Barboza MM, et al. Microbial composition of gastric lesions: differences based on *Helicobacter pylori* virulence profile. Sci Rep. 2024;14:1–15. https://doi.org/10.1038/s41598-024-80394-2.
75. Jingushi K, Kawashima A, Tanikawa S, et al. *Cutibacterium acnes*-derived extracellular vesicles promote tumor growth in renal cell carcinoma. Cancer Sci. 2024;115:2578–87. https://doi.org/10.1111/CAS.16202.
76. Ai B, Mei Y, Liang D, et al. Uncovering the special microbiota associated with occurrence and progression of gastric cancer by using RNA-sequencing. Sci Rep. 2023;13:1–11. https://doi.org/10.1038/s41598-023-32809-9.
77. Chua EG, Loke MF, Gunaletchumy SP, et al. The influence of modernization and disease on the gastric microbiome of Orang Asli, Myanmars and Modern Malaysians. Microorganisms. 2019;7:174. https://doi.org/10.3390/MICROORGANISMS7060174.
78. Bhattacharya S. An empirical review on the resistance mechanisms of epidermal growth factor receptor inhibitors and predictive molecular biomarkers in colorectal cancer. Crit Rev Oncol Hematol. 2023;183: 103916. https://doi.org/10.1016/J.CRITREVONC.2023.103916.
79. Wettergren Y, Carlsson G, Odin E, et al. Pretherapeutic uracil and dihydrouracil levels of colorectal cancer patients are associated with sex and toxic side effects during adjuvant 5-fluorouracil-based chemotherapy. Cancer. 2012;118:2935–43. https://doi.org/10.1002/CNCR.26595.
80. Wang W, Cui J, Ma H, et al. Targeting pyrimidine metabolism in the era of precision cancer medicine. Front Oncol. 2021;11: 684961. https://doi.org/10.3389/FONC.2021.684961.
81. Shentu J, Su X, Yu Y, et al. Unveiling the role of taurine and SLC6A6 in tumor immune evasion: Implications for gastric cancer therapy. Int J Biochem Cell Biol. 2024. https://doi.org/10.1016/J.BIOCEL.2024.106661.
82. Sinha A, Griffith L, Acharjee A. Systematic review and meta-analysis: taurine and its association with colorectal carcinoma. Cancer Med. 2024;13: e70424. https://doi.org/10.1002/CAM4.70424.
83. Stonāns I, Kuzmina J, Poļaka I, et al. The association of circulating L-carnitine, γ-butyrobetaine and trimethylamine N-oxide levels with gastric cancer. Diagnostics. 2023;13:1341. https://doi.org/10.3390/DIAGNOSTICS13071341.
84. Tacconi E, Palma G, De Biase D, et al. Microbiota effect on trimethylamine N-oxide production: from cancer to fitness—a practical preventing recommendation and therapies. Nutrients. 2023;15:563. https://doi.org/10.3390/NU15030563.
85. Wang S, Kuang J, Zhang H, et al. Bile acid-microbiome interaction promotes gastric carcinogenesis. Adv Sci. 2022;9:2200263. https://doi.org/10.1002/ADVS.202200263.
86. Kolho KL, Pessia A, Jaakkola T, et al. Faecal and serum metabolomics in paediatric inflammatory bowel disease. J Crohns Colitis. 2017;11:321–34. https://doi.org/10.1093/ECCO-JCC/JJW158.
87. Wu YC, Chiu CF, Hsueh CT, et al. The role of bile acids in cellular invasiveness of gastric cancer. Cancer Cell Int. 2018;18:75. https://doi.org/10.1186/S12935-018-0569-0.
88. Liu G, Yu L, Fang J, et al. Methionine restriction on oxidative stress and immune response in dss-induced colitis mice. Oncotarget. 2017;8:44511. https://doi.org/10.18632/ONCOTARGET.17812.
89. Zhou ZY, Wan XY, Cao JW. Dietary methionine intake and risk of incident colorectal cancer: a meta-analysis of 8 prospective studies involving 431,029 participants. PLoS One. 2013. https://doi.org/10.1371/JOURNAL.PONE.0083588.
90. Cao WX, Ou JM, Fei XF, et al. Methionine-dependence and combination chemotherapy on human gastric cancer cells in vitro. World J Gastroenterol. 2002;8:230. https://doi.org/10.3748/WJG.V8.I2.230.
91. Schug ZT, Peck B, Jones DT, et al. Acetyl-CoA synthetase 2 promotes acetate utilization and maintains cancer cell growth under metabolic stress. Cancer Cell. 2015;27:57–71. https://doi.org/10.1016/J.CCELL.2014.12.002.
92. Shimizu H, Ichikawa D, Tamagaki K, et al. Evaluation of postoperative nephrolithiasis and renal dysfunction in gastric cancer patients. Gastric Cancer. 2013;16:338–44. https://doi.org/10.1007/S10120-012-0192-Z.
93. Deng R, Mo F, Chang B, et al. Glucose-derived AGEs enhance human gastric cancer metastasis through RAGE/ERK/Sp1/MMP2 cascade. Oncotarget. 2017;8:104216. https://doi.org/10.18632/ONCOTARGET.22185.
94. Shastry RP, Ghate SD, Hameed A, et al. Emergence of rare and low abundant anaerobic gut Firmicutes is associated with a significant downfall of *Klebsiella* in human colon cancer. Microb Pathog. 2024;193: 106726. https://doi.org/10.1016/J.MICPATH.2024.106726.
95. Elahi Z, Shariati A, Bostanghadiri N, et al. Association of *Lactobacillus*, *Firmicutes*, *Bifidobacterium*, *Clostridium*, and *Enterococcus* with colorectal cancer in Iranian patients. Heliyon. 2023;9: e22602. https://doi.org/10.1016/J.HELIYON.2023.E22602.
96. Fang CY, Chen JS, Hsu BM, et al. Colorectal cancer stage-specific fecal bacterial community fingerprinting of the Taiwanese population and underpinning of potential taxonomic biomarkers. Microorganisms. 2021;9:1548. https://doi.org/10.3390/MICROORGANISMS9081548/S1.
97. Li M, Jin M, Zhao L, et al. Tumor-associated microbiota in colorectal cancer with vascular tumor thrombus and neural invasion and association with clinical prognosis: microbiota in colorectal cancer with vascular and neural invasion. Acta Biochim Biophys Sin (Shanghai). 2023;56:366. https://doi.org/10.3724/ABBS.2023255.
98. Zhao L, Grimes SM, Greer SU, et al. Characterization of the consensus mucosal microbiome of colorectal cancer. NAR Cancer. 2021. https://doi.org/10.1093/NARCAN/ZCAB049.
99. Lawrence GW, Begley M, Cotter PD, et al. The more we learn, the less we know: deciphering the link between human gut fusobacteria and colorectal cancer. Dig Med Res. 2020;3:21–21. https://doi.org/10.21037/DMR-2020-16.
100. Chen Y, Huang Z, Tang Z, et al. More than just a periodontal pathogen–the research progress on *Fusobacterium nucleatum*. Front Cell Infect Microbiol. 2022;12: 815318. https://doi.org/10.3389/FCIMB.2022.815318.
101. Amitay EL, Werner S, Vital M, et al. Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis. 2017;38:781–8. https://doi.org/10.1093/CARCIN/BGX053.
102. Boehm ET, Thon C, Kupcinskas J, et al. *Fusobacterium nucleatum* is associated with worse prognosis in Lauren's diffuse type gastric cancer patients. Sci Rep. 2020;10:16240. https://doi.org/10.1038/S41598-020-73448-8.
103. Wang M, Wang Z, Lessing DJ, et al. *Fusobacterium nucleatum* and its metabolite hydrogen sulfide alter gut microbiota composition and autophagy process and promote colorectal cancer progression. Microbiol Spectr. 2023;11:e02292-e2323. https://doi.org/10.1128/SPECTRUM.02292-23.
104. Strickertsson JAB, Desler C, Martin-Bertelsen T, et al. *Enterococcus faecalis* infection causes inflammation, intracellular oxphos-independent ROS production, and DNA damage in human gastric cancer cells. PLoS ONE. 2013;8: e63147. https://doi.org/10.1371/JOURNAL.PONE.0063147.
105. Williamson AJ, Jacobson R, van Praagh JB, et al. *Enterococcus faecalis* promotes a migratory and invasive phenotype in colon cancer cells. Neoplasia. 2022;27: 100787. https://doi.org/10.1016/J.NEO.2022.100787.
106. Xu J, Xiang C, Zhang C, et al. Microbial biomarkers of common tongue coatings in patients with gastric cancer. Microb Pathog. 2019;127:97–105. https://doi.org/10.1016/J.MICPATH.2018.11.051.
107. Purcell RV, Visnovska M, Biggs PJ, et al. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. Sci Rep. 2017;7:1–12. https://doi.org/10.1038/s41598-017-11237-6.
108. Wang B, Luan J, Zhao W, et al. Comprehensive multiomics analysis of the signatures of gastric mucosal bacteria and plasma metabolites

across different stomach microhabitats in the development of gastric cancer. Cell Oncol. 2024. https://doi.org/10.1007/S13402-024-00965-3/FIGURES/9.
109. Long L, Yang W, Liu L, et al. Dietary intake of branched-chain amino acids and survival after colorectal cancer diagnosis. Int J Cancer. 2021;148:2471–80. https://doi.org/10.1002/IJC.33449.
110. Ren YM, Zhuang ZY, Xie YH, et al. BCAA-producing *Clostridium symbiosum* promotes colorectal tumorigenesis through the modulation of host cholesterol metabolism. Cell Host Microbe. 2024;32:1519-1535.e7. https://doi.org/10.1016/J.CHOM.2024.07.012.
111. Budhathoki S, Iwasaki M, Yamaji T, et al. Association of plasma concentrations of branched-chain amino acids with risk of colorectal adenoma in a large Japanese population. Ann Oncol. 2017;28:818–23. https://doi.org/10.1093/ANNONC/MDW680.
112. Yu L, Bao S, Zhu F, et al. Association between branched-chain amino acid levels and gastric cancer risk: large-scale prospective cohort study. Front Nutr. 2024. https://doi.org/10.3389/FNUT.2024.1479800.
113. Lian S, Liu S, Wu A, et al. Branched-chain amino acid degradation pathway was inactivated in colorectal cancer: results from a proteomics study. J Cancer. 2024;15:3724. https://doi.org/10.7150/JCA.95454.
114. Wang Z, Lu Z, Lin S, et al. Leucine-tRNA-synthase-2-expressing B cells contribute to colorectal cancer immunoevasion. Immunity. 2022;55:1067-1081.e8. https://doi.org/10.1016/J.IMMUNI.2022.04.017/ATTACHMENT/44A660DA-17F2-492A-8A14-EE11AC4D1F68/MMC9.PDF.
115. Jabbari P, Yazdanpanah O, Benjamin DJ, et al. Supplement use and increased risks of cancer: unveiling the other side of the coin. Cancers. 2024;16:880. https://doi.org/10.3390/CANCERS16050880.
116. Shi M, Jiang Y, Wang Y, et al. Examination of antagonistic metabolic reactions and nicotinamide/methylnicotinamide as a biomarker in gastric cancer. J Clin Oncol. 2023;41:442–442. https://doi.org/10.1200/JCO.2023.41.4_SUPPL.442.
117. Narayanan A, Baskaran SA, Amalaradjou MAR, et al. Anticarcinogenic properties of medium chain fatty acids on human colorectal, skin and breast cancer cells in vitro. Int J Mol Sci. 2015;16:5014. https://doi.org/10.3390/IJMS16035014.
118. Umemori Y, Ohe Y, Kuribayashi K, et al. Evaluating the utility of N1, N12-diacetylspermine and N1, N8-diacetylspermidine in urine as tumor markers for breast and colorectal cancers. Clin Chim Acta. 2010;411:1894–9. https://doi.org/10.1016/J.CCA.2010.07.018.
119. Brown DG, Rao S, Weir TL, et al. Metabolomics and metabolic pathway networks from human colorectal cancers, adjacent mucosa, and stool. Cancer Metab. 2016. https://doi.org/10.1186/S40170-016-0151-Y.
120. Lin KY, Wang LH, Hseu YC, et al. Clinical significance of increased guanine nucleotide exchange factor Vav3 expression in human gastric cancer. Mol Cancer Res. 2012;10:750–9. https://doi.org/10.1158/1541-7786.MCR-11-0598-T.
121. Sakurai T, Katsumata K, Udo R, et al. Validation of urinary charged metabolite profiles in colorectal cancer using capillary electrophoresis-mass spectrometry. Metabolites. 2022;12:59. https://doi.org/10.3390/METABO12010059/S1.
122. Hwang Y, Jeong CS. Inhibitory effects of 4-guanidinobutyric acid against gastric lesions. Biomol Ther (Seoul). 2012;20:239–44. https://doi.org/10.4062/BIOMOLTHER.2012.20.2.239.
123. Endo Y, Marusawa H, Kou T, et al. Activation-induced cytidine deaminase links between inflammation and the development of colitis-associated colorectal cancers. Gastroenterology. 2008. https://doi.org/10.1053/J.GASTRO.2008.06.091.
124. Maneikyte J, Bausys A, Leber B, et al. Dietary glycine decreases both tumor volume and vascularization in a combined colorectal liver metastasis and chemotherapy model. Int J Biol Sci. 2019;15:1582. https://doi.org/10.7150/IJBS.35513.
125. Guo J, Pan Y, Chen J, et al. Serum metabolite signatures in normal individuals and patients with colorectal adenoma or colorectal cancer using UPLC-MS/MS method. J Proteomics. 2023. https://doi.org/10.1016/J.JPROT.2022.104741.
126. Chan CWH, Law BMH, Waye MMY, et al. Trimethylamine-N-oxide as one hypothetical link for the relationship between intestinal microbiota and cancer—where we are and where shall we go? J Cancer. 2019;10:5874. https://doi.org/10.7150/JCA.31737.
127. Issa JPJ, Vertino PM, Wu J, et al. Increased cytosine DNA-methyltransferase activity during colon cancer progression. J Natl Cancer Inst. 1993;85:1235–40. https://doi.org/10.1093/JNCI/85.15.1235.
128. Rao RK, Samak G. Role of glutamine in protection of intestinal epithelial tight junctions. J Epithel Biol Pharmacol. 2011;5:47. https://doi.org/10.2174/1875044301205010047.
129. Wu J, Li M, Zhou C, et al. Changes in amino acid concentrations and the gut microbiota composition are implicated in the mucosal healing of ulcerative colitis and can be used as noninvasive diagnostic biomarkers. Clin Proteomics. 2024. https://doi.org/10.1186/S12014-024-09513-5.
130. Wang Y, Jia Z, Wang Q, et al. Amino acids and risk of colon adenocarcinoma: a Mendelian randomization study. BMC Cancer. 2023;23:1–10. https://doi.org/10.1186/S12885-023-11514-W/FIGURES/4.
131. Tsai YC, Tai WC, Liang CM, et al. Alternations of the gut microbiota and the Firmicutes/Bacteroidetes ratio after biologic treatment in inflammatory bowel disease. J Microbiol Immunol Infect. 2024. https://doi.org/10.1016/J.JMII.2024.09.006.
132. Santoru ML, Piras C, Murgia A, et al. Cross sectional evaluation of the gut-microbiome metabolome axis in an Italian cohort of IBD patients. Sci Rep. 2017;7:1–14. https://doi.org/10.1038/s41598-017-10034-5.
133. Zhang ZJ, Qu HL, Zhao N, et al. Assessment of causal direction between gut microbiota and inflammatory bowel disease: a Mendelian randomization analysis. Front Genet. 2021;12: 631061. https://doi.org/10.3389/FGENE.2021.631061/FULL.
134. Sasaki K, Inoue J, Sasaki D, et al. Construction of a model culture system of human colonic microbiota to detect decreased lachnospiraceae abundance and butyrogenesis in the feces of ulcerative colitis patients. Biotechnol J. 2019;14:1800555. https://doi.org/10.1002/BIOT.201800555.
135. Takahashi K, Nishida A, Fujimoto T, et al. Reduced abundance of butyrate-producing bacteria species in the fecal microbial community in Crohn's disease. Digestion. 2016;93:59–65. https://doi.org/10.1159/000441768.
136. Rocha CS, Alexander KL, Herrera C, et al. Microbial remodeling of gut tryptophan metabolism and indole-3-lactate production regulate epithelial barrier repair and viral suppression in human and simian immunodeficiency virus infections. Mucosal Immunol. 2025. https://doi.org/10.1016/J.MUCIMM.2025.01.011.
137. Chen M, Fan HN, Chen XY, et al. Alterations in the saliva microbiome in patients with gastritis and small bowel inflammation. Microb Pathog. 2022;165: 105491. https://doi.org/10.1016/J.MICPATH.2022.105491.
138. Ghiboub M, Penny S, Verburgt CM, et al. Metabolome changes with diet-induced remission in pediatric Crohn's disease. Gastroenterology. 2022;163:922-936.e15. https://doi.org/10.1053/J.GASTRO.2022.05.050.
139. Vich Vila A, Hu S, Andreu-Sánchez S, et al. Faecal metabolome and its determinants in inflammatory bowel disease. Gut. 2023;72:1472. https://doi.org/10.1136/GUTJNL-2022-328048.
140. Jiang W, Zhou L, Lin S, et al. Metabolic profiles of gastric cancer cell lines with different degrees of differentiation. Int J Clin Exp Pathol. 2018;11:869.
141. Yuan LW, Yamashita H, Seto Y. Glucose metabolism in gastric cancer: the cutting-edge. World J Gastroenterol. 2016;22:2046. https://doi.org/10.3748/WJG.V22.I6.2046.
142. Rindflesch TC, Blake CL, Cairelli MJ, et al. Investigating the role of interleukin-1 beta and glutamate in inflammatory bowel disease and epilepsy using discovery browsing. J Biomed Semant. 2018;9:25. https://doi.org/10.1186/S13326-018-0192-Y.
143. Zhang D, Jin W, Wu R, et al. High glucose intake exacerbates autoimmunity through reactive-oxygen-species-mediated TGF-β cytokine activation. Immunity. 2019;51:671-681.e5. https://doi.org/10.1016/J.IMMUNI.2019.08.001.
144. Song WB, Lv YH, Zhang ZS, et al. Soluble intercellular adhesion molecule-1, D-lactate and diamine oxidase in patients with inflammatory bowel disease. World J Gastroenterol. 2009;15:3916–9. https://doi.org/10.3748/WJG.15.3916.
145. Li X, Yao Z, Qian J, et al. Lactate protects intestinal epithelial barrier function from dextran sulfate sodium-induced damage by GPR81 signaling. Nutrients. 2024. https://doi.org/10.3390/NU16050582.
146. Iraporda C, Romanin DE, Bengoa AA, et al. Local treatment with lactate prevents intestinal inflammation in the TNBS-induced colitis model. Front Immunol. 2016;7:651. https://doi.org/10.3389/FIMMU.2016.00651/FULL.

147. Chen C, Yan W, Tao M, et al. NAD+ metabolism and immune regulation: new approaches to inflammatory bowel disease therapies. Antioxidants. 2023;12:1230. https://doi.org/10.3390/ANTIOX12061230.
148. Liang QH, Li QR, Chen Z, et al. Anemoside B4, a new pyruvate carboxylase inhibitor, alleviates colitis by reprogramming macrophage function. Inflamm Res. 2024;73:345–62. https://doi.org/10.1007/S00011-023-01840-X.
149. Mabley JG, Pacher P, Liaudet L, et al. Inosine reduces inflammation and improves survival in a murine model of colitis. Am J Physiol Gastrointest Liver Physiol. 2003. https://doi.org/10.1152/AJPGI.00060.2002.
150. Guo S, Wang Y, Zhou D, et al. Association of alteration of nucleosides and nucleotides with gastric cancer microenvironment. Int J Mass Spectrom. 2018;434:37–42. https://doi.org/10.1016/J.IJMS.2018.08.012.
151. Veenstra JP, Vemu B, Tocmo R, et al. Pharmacokinetic analysis of carnosic acid and carnosol in standardized rosemary extract and the effect on the disease activity index of DSS-induced colitis. Nutrients. 2021;13:773. https://doi.org/10.3390/NU13030773.
152. El-Huneidi W, Bajbouj K, Muhammad JS, et al. Carnosic acid induces apoptosis and inhibits Akt/mTOR signaling in human gastric cancer cell lines. Pharmaceuticals. 2021;14:230. https://doi.org/10.3390/PH14030230.
153. Fukuda T, Majumder K, Zhang H, et al. Adenine has an anti-inflammatory effect through the activation of adenine receptor signaling in mouse macrophage. J Funct Foods. 2017;28:235–9. https://doi.org/10.1016/J.JFF.2016.11.013.
154. Fukuda T, Majumder K, Zhang H, et al. Adenine inhibits TNF-α signaling in intestinal epithelial cells and reduces mucosal inflammation in a dextran sodium sulfate-induced colitis mouse model. J Agric Food Chem. 2016;64:4227–34. https://doi.org/10.1021/ACS.JAFC.6B00665.
155. Zhang L, Wang Y, Chen J, et al. RFtest: a robust and flexible community-level test for microbiome data powerfully detects phylogenetically clustered signals. Front Genet. 2022;12: 749573. https://doi.org/10.3389/FGENE.2021.749573/BIBTEX.
156. Gao Y, Zhu Z, Sun F. Increasing prediction performance of colorectal cancer disease status using random forests classification based on metagenomic shotgun sequencing data. Synth Syst Biotechnol. 2022;7:574–85. https://doi.org/10.1016/J.SYNBIO.2022.01.005.
157. Zheng J, Sun Q, Zhang M, et al. Noninvasive, microbiome-based diagnosis of inflammatory bowel disease. Nat Med. 2024;30:3555–67. https://doi.org/10.1038/s41591-024-03280-4.
158. Appiah EM, Yakubu B, Salifu SP. Comprehensive microbial network analysis of gastric microbiome reveal key species affecting gastric carcinogenesis. Microbe. 2023;1: 100009. https://doi.org/10.1016/J.MICROB.2023.100009.
159. Chen Y, Wang B, Zhao Y, et al. Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. Nat Commun. 2024;15:1–13. https://doi.org/10.1038/s41467-024-46043-y.
160. Sun Y, Zhang X, Hang D, et al. Integrative plasma and fecal metabolomics identify functional metabolites in adenoma-colorectal cancer progression and as early diagnostic biomarkers. Cancer Cell. 2024;42:1386-1400.e8. https://doi.org/10.1016/J.CCELL.2024.07.005.
161. Mocanu V, Rajaruban S, Dang J, et al. Repeated fecal microbial transplantations and antibiotic pre-treatment are linked to improved clinical response and remission in inflammatory bowel disease: a systematic review and pooled proportion meta-analysis. J Clin Med. 2021;10:1–26. https://doi.org/10.3390/JCM10050959.
162. Maconi G, Manes G, Porro GB. Role of symptoms in diagnosis and outcome of gastric cancer. World J Gastroenterol: WJG. 2008;14:1149. https://doi.org/10.3748/WJG.14.1149.
163. Ghoshal UC, Chourasia D. Gastroesophageal reflux disease and *Helicobacter pylori*: what may be the relationship? J Neurogastroenterol Motil. 2010;16:243. https://doi.org/10.5056/JNM.2010.16.3.243.
164. Riihimäki M, Hemminki A, Sundquist K, et al. Metastatic spread in patients with gastric cancer. Oncotarget. 2016;7:52307. https://doi.org/10.18632/ONCOTARGET.10740.
165. Nakamura S, Hojo M. Diagnosis and treatment for gastric mucosa-associated lymphoid tissue (MALT) lymphoma. J Clin Med. 2022;12:120. https://doi.org/10.3390/JCM12010120.
166. Knowlton CA, Mackay MK, Speer TW, et al. Colon cancer. Encycl Radiat Oncol. 2024. https://doi.org/10.1007/978-3-540-85516-3_1047.
167. Stewart CL, in Surgical Oncology F, Warner S, et al. Cytoreduction for colorectal metastases: liver, lung, peritoneum, lymph nodes, bone, brain. When does it palliate, prolong survival, and potentially cure. Curr Probl Surg. 2018;55:330. https://doi.org/10.1067/J.CPSURG.2018.08.004.
168. De Simone V, Pallone F, Monteleone G, et al. Role of TH17 cytokines in the control of colorectal cancer. Oncoimmunology. 2013;2: e26617. https://doi.org/10.4161/ONCI.26617.
169. Cao H, Diao J, Liu H, et al. The pathogenicity and synergistic action of Th1 and Th17 cells in inflammatory bowel diseases. Inflamm Bowel Dis. 2023;29:818–29. https://doi.org/10.1093/IBD/IZAC199.
170. Zhou Y, Yu K. Th1, Th2, and Th17 cells and their corresponding cytokines are associated with anxiety, depression, and cognitive impairment in elderly gastric cancer patients. Front Surg. 2022;9: 996680. https://doi.org/10.3389/FSURG.2022.996680/FULL.
171. Disoma C, Zhou Y, Li S, et al. Wnt/β-catenin signaling in colorectal cancer: Is therapeutic targeting even possible? Biochimie. 2022;195:39–53. https://doi.org/10.1016/J.BIOCHI.2022.01.009.
172. Mao J, Fan S, Ma W, et al. Roles of Wnt/β-catenin signaling in the gastric cancer stem cells proliferation and salinomycin treatment. Cell Death Dis. 2014;5:e1039–e1039. https://doi.org/10.1038/cddis.2013.515.
173. Kim JW, Lee HS, Nam KH, et al. PIK3CA mutations are associated with increased tumor aggressiveness and Akt activation gastric cancer. Oncotarget. 2017;8:90948. https://doi.org/10.18632/ONCOTARGET.18770.
174. Wang H, Tang R, Jiang L, et al. The role of PIK3CA gene mutations in colorectal cancer and the selection of treatment strategies. Front Pharmacol. 2024;15:1494802. https://doi.org/10.3389/FPHAR.2024.1494802.
175. Stidham RW, Higgins PDR. Colorectal cancer in inflammatory bowel disease. Clin Colon Rectal Surg. 2018;31:168–78. https://doi.org/10.1055/S-0037-1602237.
176. Spisak T. Statistical quantification of confounding bias in machine learning models. Gigascience. 2022;11:1–15. https://doi.org/10.1093/GIGASCIENCE/GIAC082.
177. Du H, Yang Q, Ge A, et al. Explainable machine learning models for early gastric cancer diagnosis. Sci Rep. 2024;14:1–15. https://doi.org/10.1038/s41598-024-67892-z.
178. Erawijantari PP, Mizutani S, Shiroma H, et al. Influence of gastrectomy for gastric cancer treatment on faecal microbiome and metabolome profiles. Gut. 2020;69:1404–15. https://doi.org/10.1136/GUTJNL-2019-319188.
179. 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. https://doi.org/10.1038/s41591-019-0458-7.
180. Franzosa EA, Sirota-Madi A, Avila-Pacheco J, et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol. 2019;4:293–305. https://doi.org/10.1038/S41564-018-0306-4.
181. Park CH, Hong C, Lee A reum, et al. Multi-omics reveals microbiome, host gene expression, and immune landscape in gastric carcinogenesis. IScience. 2022. https://doi.org/10.1016/J.ISCI.2022.103956.
182. Kim M, Vogtmann E, Ahlquist DA, et al. Fecal metabolomic signatures in colorectal adenoma patients are associated with gut microbiota and early events of colorectal cancer pathogenesis. MBio. 2020. https://doi.org/10.1128/MBIO.03186-19.
183. Lloyd-Price J, Arze C, Ananthakrishnan AN, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature. 2019;569:655–62. https://doi.org/10.1038/s41586-019-1237-9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.