Ilona Vilkoite

ORCID 0000-0001-9663-9156

The Association of Colorectal Neoplasia with Changes in the Basic Composition of the Gut Microbiome

Summary of the Doctoral Thesis for obtaining the scientific degree “Doctor of Science (*PhD*)”

Sector Group – Medical and Health Sciences

Sector – Clinical Medicine

Sub-Sector – Internal Medicine

Riga, 2026

The Doctoral Thesis was developed at Health Center 4 branch Diagnostic Center, Academic Histology Laboratory, Latvian Biomedical Research and Study Center

Supervisors of the Doctoral Thesis:

Dr. med., Professor **Aivars Lejnieks**,
Rīga Stradiņš University, Latvia

Dr. med., Assistant Professor **Ivars Tolmanis**,
Rīga Stradiņš University, Latvia

Scientific Advisor:

Dr. biol., Researcher **Laila Silamiķele**,
Latvian Biomedical Research and Study Center

Official Reviewers:

Dr. med., Associate Professor **Andris Gardovskis**,
Rīga Stradiņš University, Latvia

PhD, Associate Professor **Elīna Sīviņa**,
University of Latvia

PhD, Professor **Corrado Pedrazzani**,
University of Verona, Italy

Chair of the Doctoral Committee:

Dr. med., Professor **Ilze Konrāde**,
Rīga Stradiņš University, Latvia

Members of the Doctoral Committee:

Dr. biol., Professor **Edvīns Miklaševičs**,
Rīga Stradiņš University, Latvia

Dr. med., Associate Professor **Viktorija Kuzema**,
Rīga Stradiņš University, Latvia

Dr. med., Associate Professor **Arvīds Irmejs**,
Rīga Stradiņš University, Latvia

PhD, Assistant Professor **Aigars Reinis**,
Rīga Stradiņš University, Latvia

Defence of the Doctoral Thesis will take place at the public session of the Promotion Council of Clinical Medicine on 10 November 2026 at 15.00 in the 1st Auditorium, 16 Dzirciema iela, Rīga Stradiņš University.

The Doctoral Thesis is available in RSU Library and on RSU website:
https://www.rsu.lv/en/dissertations



The Doctoral Thesis was developed with the support of the project
"The role of miRNAs in host-gut microbiome interactions in the context of metformin therapy in metabolic disorders" No.: 1.1.1.1/18/A/092

Table of Contents

Introduction 7
Aim of the Thesis 8
Tasks of the Thesis 9
Hypothesis of the Thesis 9
Novelty of the Thesis 9
Collaboration partners 10
Personal contribution 10
Ethical aspects 11

1 Materials and Methods 12
1.1 Study Design 12
1.2 Material acquisition and storage 13
1.2.1 Stool sample acquisition and storage 13
1.2.2 Colonoscopy examination 13
1.2.3 Morphological diagnosis of lesions detected during colonoscopy 14
1.2.4 Metagenome sequencing and determination of microbiome composition 14
1.2.5 Data analysis 15
1.2.6 Statistical analysis 15

2 Results 17
2.1 Characteristics of the study population 17
2.2 Demographic and clinical characteristics 17
2.2.1 Age 17
2.2.2 Gender 17
2.2.3 Body mass index (BMI) 18
2.2.4 Place of residence of patients 18
2.2.5 Patient education 18
2.2.6 Lifestyle factors 18
2.3 Characteristics of adenomas 20
2.4 Microbiome composition at the genus level 21
2.5 Diversity analysis at the genus level 22
2.5.1 Alpha diversity 22
2.5.2 Beta diversity at the species level 23
2.6 Differential abundance analysis between patients with and without colorectal adenomas 24
2.7 Microorganism genera with differences in relative incidence under the influence of the factors analysed in the study 25
2.7.1 Relationship between genera of microorganisms and digestive tract diseases 27
2.7.2 Association between microbial genera and body mass index (BMI) 28
2.7.3 Association between microbial genera and smoking 29
2.7.4 Significant associations of microorganisms with clinical and demographic factors (heat map visualisation of MaAsLin2 analysis results) 30
2.8 Microbial functional pathways meeting the exploratory MaAsLin2 threshold in patients with and without colorectal adenomas 30
2.8.1 PWY-7371 (1,4-dihydroxy-6-naphthoate biosynthesis II) pathway 31
2.8.2 Menaquinol-8 biosynthesis superpathway (PWY-7992) 32
2.8.3 PROPFERM-PWY: L-alanine fermentation superpathway or Stickland reaction 33
2.8.4 PWY-8188: L-alanine degradation VI (reductive Stickland reaction) 33
2.8.5 PWY-8189: L-alanine degradation V (oxidative Stickland reaction) 34

2.8.6 PWY-5494: Fermentation of pyruvate to propanoate II (acrylate pathway).....34
2.8.7 Comparative analysis of six microbial metabolic pathways 35
Discussion 37
Patient characteristics 37
Microbiome composition at the genus level.....38
Alpha diversity 38
Beta diversity.....39
Differential abundance analysis between patients with and without colorectal adenomas 40
Microbial genera with significant differences in relative abundance under the influence of the factors analysed in the study.....43
Functional differences in the microbiome between patients with and without colorectal adenomas 44
Menaquinone (vitamin K2) biosynthetic pathways (PWY-7371 and PWY-7992).....44
Stickland fermentation pathways (PROPFERM-PWY, PWY-8188, PWY-8189) 45
Acrylate pathway (PWY-5494: pyruvate fermentation to propanoate II).....45
Conclusions 47
Proposals 49
List of publications, reports and patents on the topic of the doctoral thesis 51
References 52
Acknowledgements 58

Abbreviations used in the Thesis

AD	Colorectal adenomas
BMI	Body mass index
CO	Control group
DNA	Deoxyribonucleic acid
FDR	False discovery rate
GERD	Gastroesophageal reflux disease
GTDB	Genome Taxonomy Database; a genome-based taxonomic database for prokaryotes
IL-17	Interleukin-17
K_2 vitamin	K2 – Menaquinone
Limma	Linear Models for Microarray Data
logFC	Log fold change
MaAsLin2	Multivariate Association with Linear Models 2
MAG	Metagenome-assembled genome; a microbial genome bioinformatically reconstructed from metagenomic sequencing data
NF-κB	Nuclear factor kappa B
PCA	Principal component analysis
PCoA	Principal coordinates analysis
PWY	Pathway, metacyc
SCFA	Short-chain fatty acids
Sp	Species
spp.	Multiple species
UBA	Uncultivated Bacteria and Archaea; a designation for uncultivated bacteria and archaea used for metagenome- assembled microbial genomes (MAGs) identified from genomic sequence data without prior cultivation

Introduction

Colorectal cancer (CRC) remains a major global health challenge, placing a significant burden on both healthcare systems and individuals affected by the diagnosis (Bray et al, 2024). CRC is one of the most common and lethal malignancies globally, characterised by high morbidity and mortality rates. Recent epidemiological data indicate that CRC accounts for approximately 10 % of all cancers, making it the third most commonly diagnosed cancer worldwide, with approximately 1.8 million new cases diagnosed each year (Bray et al, 2024). This prevalence highlights the need to better understand the factors that contribute to the development of CRC, as well as to identify potential preventive and therapeutic strategies. The aetiology of CRC is multifactorial, involving a complex interaction of genetic, environmental and lifestyle factors (Gagnière et al, 2016). The main modifiable risk factors include high consumption of red and processed meat, low dietary fibre intake, obesity, physical inactivity, smoking, excessive alcohol consumption, and chronic intestinal inflammation

(Koliarakis et al., 2018, Kang et al., 2017, Wang et al, 2022, Franzosa et al., 2018). In addition to these factors, new studies increasingly recognise the role of the gut microbiota in the development and progression of CRC (Mallick et al., 2021, Ahn et al., 2013, Holmes et al., 2011, Zeller et al., 2014). The human colon is home to a diverse and dynamic community of microorganisms, collectively referred to as the gut microbiota, which plays a crucial role in maintaining intestinal homeostasis and influencing various aspects of host physiology (Brennan et al., 2016). These microorganisms not only provide digestive functions, protection against pathogens, but also contribute to immunomodulation, participate in metabolic processes, which are essential for maintaining a healthy colon.

CRC usually arises from precancerous formations known as polyps. These polyps undergo a series of molecular and cellular changes, which opens up opportunities for the activation of the so-called adenoma-carcinoma pathway, when the transition from benign adenoma to malignant carcinoma occurs (Lv et al., 2023). In addition, 15–30 % of CRC develop through the so-called serrated pathway, which is an alternative mechanism of tumourigenesis that involves distinct genetic and epigenetic changes (Sze et al., 2016). Both pathways begin with precancerous polyps, which can be detected and removed during colonoscopy to prevent the development of CRC. However, if polyps are incompletely removed or ablated or for some reason not detected, they can cause so-called interval cancers – cases of colorectal cancer that occur between scheduled screening examinations. It is important to understand the factors that influence polyp formation and growth, as well as the transition from benign to malignant states. The formation of colorectal polyps and the subsequent development of colorectal cancer are influenced by a complex set of environmental influences and genetic predispositions (Lv et al.,

2023). One of the most important factors that is increasingly implicated in this process is the intestinal microbiota. It has been shown that changes in the composition and function of the colonic microbiota, referred to as dysbiosis, contribute to the pathogenesis of colorectal cancer. Dysbiotic changes in the intestinal microbiota can lead to chronic inflammation, abnormalities in immune responses and metabolic dysregulation in the colonic microenvironment, creating conditions that promote tumour formation and progression (Gagnière et al., 2016). Disturbed microbial balance can lead to an overgrowth of pathogenic bacteria and a decrease in beneficial bacteria, which together promote inflammation and increase levels of harmful metabolites that can damage the colonic epithelium and promote oncogenic mutations.

The relationship between the gut microbiota and CRC is a rapidly growing area of research. Studies have identified specific bacteria that are found in higher numbers in CRC patients, such as *Fusobacterium nucleatum*, Escherichia coli, and *Bacteroides* fragilis. These bacteria produce pro-inflammatory molecules and toxins that can cause DNA damage, disrupt cell signalling pathways, and weaken the epithelial barrier, thereby creating a microenvironment that supports tumourigenesis (Kostic et al., 2013; Lindström et al., 2017). In contrast, beneficial bacterial strains such as *Faecalibacterium prausnitzii* and *Akkermansia muciniphila*, which produce anti-inflammatory metabolites and help maintain the integrity of the intestinal barrier, are often present in lower abundance in patients with CRC, suggesting their protective role against tumour formation (Guo et al., 2020).

Understanding the complex relationship between the colonic microbiota and the pathogenesis of CRC offers hope for the development of new preventive and therapeutic strategies. Monitoring changes in specific bacterial populations or microbial metabolites could serve as biomarkers for the early detection of CRC or its precursors. In addition, strategies to restore microbial balance, such as dietary modifications, probiotics, prebiotics, and even faecal microbiota transplantation, are being investigated as a potential method to prevent CRC in high-risk patients.

Aim of the Thesis

To evaluate the differences in the composition and functional profile of the gut microbiota between patients with and without colorectal adenomas using a metagenomic sequencing approach, with the aim of identifying structural, functional and clinically relevant microbiome indicators associated with the early development of colorectal neoplasia.

Tasks of the Thesis

To achieve the goal of the doctoral thesis, the following tasks have been set:

1 To characterise the study participants and identify clinically relevant differences in demographic, anthropometric and lifestyle factors between individuals with and without colorectal adenomas.
2 To analyse the composition of the gut microbiota at the genus and species level using metagenomic sequencing data, and to compare the relative abundance of the most common taxa between the two groups.
3 To assess the alpha diversity of the microbiota at the genus level by comparing species richness, evenness and dominance between adenoma and control groups.
4 To compare the degree of structural diversity (beta diversity) of microbial communities between patients with and without colorectal adenomas, identifying specific taxa that contribute most to intergroup differentiation.
5 To perform differential abundance analysis at the species level using statistical models to identify microorganisms whose abundance differs significantly between adenoma and control groups.
6 To evaluate the association of the relative abundance of microbial genera with individual clinical and anthropometric factors, including gastrointestinal diseases, body mass index, gender, and smoking habits.
7 To identify microbiome functional pathways whose relative representation differs between patients with and without colorectal adenomas and to interpret the potential biological significance of these differences.
8 Integrate structural and functional microbiome data to conceptualise the potential role of microbiota alterations in the early stages of colorectal adenomatosis development and hypothesise potential gut microbiome biomarkers.

Hypothesis of the Thesis

Individuals with established colon adenomas have different gut microbiota composition and functional profile, including key microbial metabolic pathways, than those without adenomas.

Novelty of the Thesis

The work investigates the potential role of the gut microbiota and its specific metabolic pathways in the development of colorectal adenomas. Using shotgun metagenomic sequencing, the study provides new insight into the composition of the gut microbiota and its metabolic functions associated with the development of colorectal adenomas. This information improves

the understanding of the early stages of CRC development by identifying specific pathways, such as menaquinone biosynthesis. The study identifies microbiome compositional differences and exploratory functional signals associated with the presence of colorectal adenomas. These findings improve the understanding of microbiome alterations at an early stage of colorectal neoplasia and provide a basis for further biomarker research. However, validation in independent cohorts and complementary functional or metabolomic studies is required before these findings can be translated into diagnostic, preventive, or therapeutic applications.

The study highlights the potential of a personalised medicine approach based on an individual's gut microbiota profile, when targeted and specific interventions can be considered. Identification of disrupted metabolic pathways allows for consideration of therapeutic approaches that could alter the composition of the gut microbiota and inhibit the process of adenoma formation.

Collaboration partners

1 Endoscopic examinations and polypectomy procedures were performed at the Diagnostic Center of SIA "Veselības Centrs 4" branch.
2 Morphological analysis of the removed polyps was performed at SIA "Akadēmiskā Histoloģijas laboratorija".
3 Faecal microbiota testing was performed at SIA "Latvijas Biomedicīnas pētījumu un studiju centrs".

Personal contribution

The author of the work has performed all endoscopic examinations, polypectomy procedures, and sending the removed polyps to the Academic Histology Laboratory. Has communicated with patients, organised the correct collection and delivery of faecal material to the Biomedical Center.

Has informed patients about the study and obtained their consent to participate in the study. Has ensured the attraction of material resources and logistics necessary for conducting the examinations.

Has performed scientific and clinical planning, including patient selection and selection of study indicators.

Has compiled, analysed, and processed the obtained data, including using statistical data analysis methods. The results were used in the preparation of scientific publications. The author of the work has fully developed and written this doctoral thesis.

During the development of the doctoral thesis, the author participated in the preparation of several scientific publications related to the assessment of the quality of colonoscopy.

One of the publications analysed the accuracy of the NICE classification in diagnosing polyps depending on their localisation in the colon (Vilkoite et al., 2022), where the author performed data analysis, participated in the classification of polyps and the interpretation of the results. The author also led and participated in the development, implementation and publication of a study on the application of artificial intelligence in colonoscopy (Vilkoite et al., 2023). This study made a significant contribution to improving the quality of colorectal cancer diagnostics in Latvia, supplementing the author's scientific expertise in the assessment of the quality of colonoscopy, the impact of the artificial intelligence system on the adenoma detection rate (ADR) and polyp detection rate (PDR) in colonoscopies. These studies form the basis for a multidimensional understanding of the aspects of early diagnosis of colorectal cancer. The third scientific publication, whose author is also the author of the doctoral thesis, provided a clinical and epidemiological insight into the differences in the incidence of colorectal cancer and polyp characteristics in different age groups.

Ethical aspects

The study was conducted in accordance with the ethical aspects of the Declaration of Helsinki, and the study received the opinion of the Central Medical Ethics Committee – No. 01-29.1.2/1751.

1 Materials and Methods

1.1 Study Design

A cross-sectional case-control study was conducted at the Department of Gastrointestinal Endoscopy, Diagnostic Center, Health Center 4, Academic Histology Laboratory, and Latvian Biomedical Research and Study Center from April 1, 2021 to September 22, 2024. The study was approved by the Latvian Central Medical Ethics Committee (No. 01-29.1.2/1751). All participants signed informed consent, and the procedures were performed in accordance with the principles of the Declaration of Helsinki.

Colonoscopies were performed by a single experienced endoscopist (ADR 36 % in the screening population), ensuring methodological uniformity.

Inclusion criteria: age ≥ 18 years, colonoscopy for medical indications, and consent to participate.

Exclusion criteria:

- history of colonoscopy,
- inflammatory bowel disease,
- hereditary polyposis syndrome,
- known diagnosis of colorectal cancer (CRC),
- major bowel surgery (e. g., bowel resection, bariatric surgery), except appendectomy,
- any contraindications to polypectomy,
- incomplete bowel preparation (BBPS 0–1 in any of the three segments),
- standard contraindications to colonoscopy (e. g., acute diverticulitis or suspected perforation),
- incomplete colonoscopy (due to technical reasons or poor preparation),
- pregnancy or breastfeeding,
- any acute illness at the time of inclusion,
- less than 3 years since the oncological diagnosis or the end of specific therapy,
- chronic kidney disease, autoimmune disease, HIV, hepatitis B or C virus infection,
- chronic alcoholism,
- antibacterial, probiotic, immunosuppressive, use of glucocorticosteroids or proton pump inhibitors in the last two months,
- diarrhoea in the last two weeks.

In addition, a comparative analysis of colonoscopy data (2022.09.–2023.02.) was performed, comparing the rates of adenoma detection in patients < 50 and ≥ 50 years of age in relation to the increase in EOCRC. All polyps were classified according to WHO criteria (Odze RD et al., 2019).

1.2 Material acquisition and storage

1.2.1 Stool sample acquisition and storage

Of the 1363 patients recommended for colonoscopy, 146 agreed to participate in the study and received a stool collection kit (Zymo Research). Samples were collected the day before bowel preparation. Upon arrival for the examination, they were stored at −20 °C and transferred to −80 °C within 24 hours. A total of 136 samples were collected and available for metagenomic processing (56 from patients with adenomas and 80 from controls). After sequencing quality control and removal of host-derived reads, 123 samples were retained for downstream microbiome analyses.

1.2.2 Colonoscopy examination

Colonoscopies were performed with an Olympus EVIS EXERA III (CF-HQ190L/l) videocolonoscope. All examinations were performed by a single experienced endoscopist (9 years of experience, ~1300 colonoscopies per year), ensuring methodological uniformity. The colonic mucosa was initially assessed in white light mode, and polyps were additionally analysed with narrowband imaging (NBI) and classified according to NICE criteria (Hewett et al., 2012); the diagnostic accuracy of this approach has been previously validated (Vilkoite et al., 2022).

Examinations were performed under the supervision of an anaesthesiologist using intravenous propofol sedation. The quality of bowel preparation was assessed using the Boston Bowel Preparation Scale (BBPS); patients with incomplete preparation (0-1 in any segment) were excluded. The minimum instrument evacuation time was ≥ 7 minutes.

All polyps were documented according to the Paris and NICE classifications (Hewett et al., 2012), specifying their location and size. Biopsy material and ablated polyps were sent for histopathological examination; at least two biopsy specimens were taken before polypectomy.

The examination included retroversion into the rectum, photodocumentation of the cecum dome and vermiform appendix opening, and, if possible, retroversion into the dome. The right bowel was examined twice using NBI; during the second evacuation, chromoendoscopy with methylene blue was performed to improve the detection of flat polyps.

Polyps < 1 cm were removed by cold snare technique, >1 cm by diathermocoagulation. If endoscopic ablation was not possible, biopsies were taken and the patient was referred for surgical treatment.

1.2.3 Morphological diagnosis of lesions detected during colonoscopy

All polyps and biopsy specimens ablated during colonoscopy were sent to the Academic Histology Laboratory (Riga, Latvia) for morphological evaluation. The specimens were analysed by expert pathologists according to the WHO classification criteria (Odze RD et al., 2020).

Lesions were classified as serrated polyps, low-grade dysplasia (LGD), high-grade dysplasia (HGD), superficial submucosal invasive carcinoma (SM-s; < 1000 μm) or deep submucosal invasive carcinoma (SM-d; ≥ 1000 μm). The methodology followed a previously validated approach (Odze RD et al., 2020).

No traditional serrated adenomas (TAS), sessile serrated lesions with dysplasia (SSL-D) or unclassified serrated adenomas were morphologically detected.

Low-risk polyps were defined as < 10 mm tubular adenomas with LGD, while high-risk polyps were defined as adenomas with villous structure, HGD, diameter >10 mm, or ≥ 3 adenomas regardless of size (Levin et al., 2008), considering their increased risk of CRC.

1.2.4 Metagenome sequencing and determination of microbiome composition

Microbial DNA was isolated from faecal samples of study participants using the FastDNA Spin Kit for Soil (MP Biomedicals, LLC, USA) according to the manufacturer's protocol. DNA concentration and quality were assessed with the Qubit dsDNA HS Assay Kit. The researchers performing the sequencing did not have access to personally identifiable information. DNA samples were stored at −20 °C.

Metagenome libraries were prepared from 300 ng of microbial DNA using the MGIEasy Universal DNA Library Prep Kit (MGI Tech Co., Ltd.). Preparation included DNA fragmentation, adapter ligation, PCR amplification, purification and quality control (Agilent Bioanalyzer 2100, Qubit® 2.0). Sequencing was performed on a DNBSEQ-G400RS platform (PE150), providing approximately 20 million reads per sample. Sequencing depth could vary depending on microbial diversity and host DNA contamination.

A shotgun metagenomics approach was used to determine the composition of the microbiome, which sequences all genetic material (bacteria, viruses, archaea, fungi) present in a sample, providing a broad characterisation of the microbial community. This method allows the identification of microorganisms to the species and sometimes strain level (Lloyd-Price et al., 2016), which is essential when analysing the potential role of specific strains in health and disease pathogenesis.

The shotgun approach also provides insight into the functional potential of the microbiota by identifying genes and metabolic pathways involved in carbohydrate digestion, vitamin synthesis, short-chain fatty acid (SCFA) production, and other processes

(Franzosa et al., 2014). It is also possible to identify antibiotic resistance genes, which reflect the resistance potential in the microbiota (Sommer et al., 2009). Functional analysis allows for the evaluation of the potential impact of the microbiota on intestinal barrier function, immunomodulation, and metabolic processes.

Limitations of the method include higher costs compared to 16S rRNA sequencing, the complexity of bioinformatic analyses, and possible host DNA contamination in faecal samples, which can affect data quality (Lindgreen et al., 2016). However, in some cases, the presence of host DNA can provide additional information about host-microbiome interactions.

Shotgun sequencing is based on random DNA fragmentation and sequencing of individual fragments, after which the genome content is reconstructed. This approach allows for the determination of relative microbial abundance and provides a detailed picture of the microbial community structure.

Statistical analysis compared the relative abundance of microorganisms between patient groups. Linear mixed effects models (LME) were used to evaluate associations, calculating coefficients (log(coefficient)), p-values, and q-values (FDR correction). The models included fixed effects (e. g., gastrointestinal diseases) and took into account potential covariates, including BMI and gender. Multiple hypothesis testing adjustment was performed to reduce the risk of false positive results.

1.2.5 Data analysis

Read quality was assessed with FastQC (Andrews, 2010), while adapters and low-quality fragments were trimmed with fastp (Chen et al., 2018). Paired reads ≥ 100 bp were retained for further analysis. Reads of human origin were removed with Bowtie2 (Langmead & Salzberg, 2012) using the GRCh38 reference genome.

Taxonomic classification was performed with Kraken2 (Wood et al., 2019) and Bracken (Lu et al., 2017) using the UHGG v2 database (Almeida et al., 2021), down to the species level. Functional analysis was performed with HUMAnN 3 (Beghini et al., 2021), linking genes to MetaCyc metabolic pathways and normalising relative abundances. Samples with < 100 000 quality reads were excluded.

1.2.6 Statistical analysis

Statistical analysis was performed in R Studio (R 4.4.1; R Core Team, 2023). Alpha diversity was calculated with the vegan package (Oksanen et al., 2022) by analysing Shannon, Simpson, and inverse Simpson indices, and the results were expressed as mean ± SD. Normality was tested with the Lilliefors test (Lilliefors, 1967). Groups were compared using the Welch

t-test or the Wilcoxon rank sum test depending on the data distribution; statistical significance was set at $p < 0.05$.

Beta diversity was assessed with the Aitchison distance (Aitchison, 1982) after CLR transformation (scikit-bio 0.5.5; scikit-bio development team, 2019), excluding taxa with relative abundance < 0.01 % in at least 90 % of the samples. Data were visualised with principal component analysis (PCA) using scikit-learn (Pedregosa et al., 2011).

Differential occurrence analysis was performed with limma (Ritchie et al., 2015), normalising data with the voom function. Samples with < 100 000 reads were excluded. Additional differential taxa and functional scores were analysed with MaAsLin2 (Mallick et al., 2021), including adenoma status, gender, BMI, smoking, and comorbidities as fixed effects. FDR correction was applied for multiple testing. For exploratory MaAsLin2 analyses, $q < 0.25$ was used as the predefined exploratory threshold, while $q < 0.05$ was considered the conventional threshold for statistical significance,

2 Results

2.1 Characteristics of the study population

A total of 146 patients who underwent colonoscopy were included in the study. Of these, 10 patients were excluded (3 patients with ulcerative colitis, 3 with colorectal cancer and 4 due to insufficient bowel preparation). 136 patients were included in the final analysis. Based on the results of colonoscopy, patients were divided into two groups: adenoma-positive group (AD): n = 56 (41.2 %) and control group (CO): n = 80 (58.8 %) (Table 2.1).

Table 2.1

Distribution of patients according to inclusion, exclusion criteria and colonoscopy examination results

Patients	N	%
Colonoscopy performed	146	100
Excluded	10	6.85
Included	136	93.15
Adenoma positive group	56	41.18
Adenoma negative group	80	58.82

2.2 Demographic and clinical characteristics

2.2.1 Age

Patients in the adenoma group were statistically significantly older than those in the control group (53.03 ± 5.4 vs. 45.61 ± 3.0 years; $p < 0.001$). This difference confirms the importance of age as one of the main risk factors for colorectal adenomas (Table 2.2).

Table 2.2

Mean age of patients in the adenoma and control groups

Group	Mean age (years) ± SD
Adenoma group	53.03 ± 5.4
Control group	45.61 ± 3.0
p-value	< 0.001

2.2.2 Gender

In the adenoma group, the proportion of men was higher than in the control group (48.2 % vs. 32.5 %), and this difference was statistically significant ($p = 0.027$). This suggests a possible effect of gender on the development of adenomas (Table 2.3).

Table 2.3

Distribution of patients by gender

Group	Men N (%)	Women N (%)	Total (N)	χ^2	p-value
AD	27/56 (48.21)	29/56 (51.79)	56		
CO	26/80 (32.50)	54/80 (67.50)	80	4.86	0.027
Total	53 (39)	83 (61)	136		

2.2.3 Body mass index (BMI)

In the adenoma group, BMI was higher than in the control group (27.5 ± 4.2 vs. 25.7 ± 5.2 kg/m²; p = 0.039), which indicates the possible role of metabolic factors in the development of adenomas.

2.2.4 Place of residence of patients

In the adenoma group, 57.14 % of patients lived in the capital, while in the control group this proportion was only 25.00 % (Table 2.4). The result of the chi-square test ($\chi^2 = 16.29$; $p < 0.001$) indicates a statistically significant relationship between place of residence and the presence of adenomas, which may indicate a possible influence of urban lifestyle or environmental factors on the development of adenomas.

Table 2.4

Place of residence of patients

Group	Capital, N (%)	Outside the capital, N (%)
AD	32/56 (57.14)	24/56 (42.86)
CO	20/80 (25.00)	60/80 (75.00)

2.2.5 Patient education

The majority of the adenoma group had a bachelor's degree or equivalent (83.90 %), while the proportion in the control group was significantly lower – 46.25 %. The difference was statistically significant (p < 0.001) (Table 2.5).

Table 2.5

Patient education

Group	Higher, N (%)	Secondary and lower, N (%)
AD	47/56 (83.9)	9/56 (16.1)
CO	37/80 (46.25)	43/80 (53.75)

2.2.6 Lifestyle factors

Physical activity

The proportion of physically active patients in the adenoma group was significantly higher (73.2 % vs. 17.5 %; p < 0.001). This result is unexpected and may indicate the influence of possible selection bias or other confounding factors (Table 2.6).

Table 2.6

Prevalence of physical activity in adenoma and control groups

Group	Physically active, N (%)	Physically inactive N (%)
AD	41/56 (73.2)	15/56 (26.8)
CO	14/80 (17.50)	66/80 (82.50)

Smoking

No statistically significant differences in smoking habits were found (p = 0.28) (Table 2.7).

Table 2.7

Smoking

Group	Smoker (N, %)	Non-smoker ((N, %)
AD	5/56 (8.93)	51/56 (91.07)
CO	3/80 (3.75)	77/80 (96.25)

Alcohol use

High alcohol consumption was significantly more common in the adenoma group (85.7 % vs. 37.5 %; p < 0.001), indicating a possible role of alcohol in the development of adenomas (Table 2.8).

Table 2.8

Alcohol use

Alcohol Frequency	Total	AD	CO
Never	10	2	8
Exceptionally	10	2	8
1–2 × per week	38	4	34
≥ 3 × per week	36	22	14
Every day	42	26	16

The results of the study (Table 2.9) show statistically significant differences between the adenoma and control groups in several demographic and lifestyle indicators. Patients in the adenoma group were older, more often male, with a higher BMI, more often lived in the capital, had higher education and were more physically active (all p < 0.05). Also, high alcohol consumption was more often found in this group (p < 0.001). In contrast, no statistically significant differences were observed in smoking habits (p = 0.280).

Table 2.9

Between-group comparative analysis by demographic and lifestyle indicators

Variable	AD	CO	Statistical test	p-value
Age (years, mean ±SD)	53.03±5.4	45.61±3.0	Welch's t-test	< 0.001
Males (%)	48.21	32.50	Chi-square	0.027
Living in the capital (%)	57.14	25	Chi-square	< 0.001
Higher education (%)	83.9	46.25	Fisher's exact	< 0.001
Physically active (%)	73.2	17.50	Chi-square	< 0.001
Smoker (%)	8.93	3.75	Fisher's exact	0.280
High alcohol consumption (%)	85.7	37.5	Chi-square	< 0.001

To complement the intergroup comparison, Table 2.10 provides detailed demographic and lifestyle characteristics of the patients, reflecting both absolute numbers and the distribution into adenoma and control groups.

Table 2.10

Demographic characteristics of adenoma and control group patients

	Total (136)	AD (56)	CO (80)	p-value
Age (years, mean ±SD)		53.03 ± 5.4	45.6±13.0	P< 0.001
Gender, N%				
Men	53	27	26	
Women	83	29	54	
BMI kg/m2 (mean ± SD)		27.5±4.2	25.7±5.2	0.039
Residence				
Capital	52	32	20	< 0.001
Cities and districts	84	24	60	
Education				< 0.001
Highest	84	47	37	< 0.001
Average and lower	52	9	43	
Lifestyle				
Physical active	55	41	14	< 0.001
Physically inactive	81	15	66	
Smoking				
Smoker	8	5	3	0.273
Non-smoker	128	51	77	
Alcohol consumption				
High consumption	78	48	30	< 0.001
Low consumption	58	8	50	

2.3 Characteristics of adenomas

A total of 77 adenomas were detected in 56 patients. Of these, 85.7 % were low-risk adenomas, 14.3 % were high-risk adenomas. The majority of adenomas were 5–9 mm (71.4 %), which corresponds to the distribution typical of screening populations. Table 2.11. shows the distribution of adenomas by size in the analysed cohort.

Table 2.11

Size and number of adenomas

Size of adenomas (mm)	Number of adenomas (N = 77)
< 5 mm	11
5–9 mm	55
10 mm and larger	11

BMI of patients in the adenoma group was 27.51 ± 4.18, and in the control group – 25.73 ± 5.22 (p = 0.039).

2.4 Microbiome composition at the genus level

Of the 136 collected samples, 123 passed metagenomic quality control and were retained for downstream analysis. In these samples, the median number of reads remaining after quality filtering and removal of host-derived reads was 22,658,664 (IQR 10,326,052).

The relative abundance of the most common genera in patients with colorectal adenomas (AD) and patients without adenomas (CO) is shown in Figure 2.1. The most common genera in both groups were *Faecalibacterium* (blue), *Bacteroides* (light blue), *Blautia*_A (orange), *Alistipes* (yellow), *Prevotella* (red) and *Roseburia* (light purple). In addition, the archaeal *Methanomassiliicoccus*_A (greenish yellow) was found to be present among the 20 most common genera.

The relative proportions of genera indicate interindividual variability, which is evident in the diversity of column heights and colour distribution. Detailed information on the dominant genera in each sample and their belonging to the AD or CO group is presented in Table 2.12.

Figure 2.1 **Composition of the faecal microbiome in each of the studied samples at the genus level – 20 most common genera, N = 123**

Table 2.12

Differences in bacterial composition between adenoma (AD) and control (CO) groups, highlighting the most common genera

Genus	Adenoma group (AD)	Control group (CO)
Faecalibacterium	Reduced (~10–20%) in most samples	Dominant (~40–50%)
Bacteroides	Dominant (~55–65%) in most samples	Moderate (~20–30%)
Phocaeicola	More common (~15–25%)	Less common (~8–12%)
*Methanomassiliicoccus*_A	Slightly more common compared to the control group (~3–5%)	Less common compared to the adenoma group (~2–3%)

Table 2.12 continued

Genus	Adenoma group (AD)	Control group (CO)
*Blautia*_A	Moderately common (~5–10%)	More common (~10–15%)
Alistipes	~3–7% of samples	~8–12% of samples
Gemmiger	Low prevalence (~3–5%)	Prevalence (~7–11%)
Agathobacter	Prevalence 5–11%	Prevalence (~7–12%)
Prevotella	Prevalence 15–20%	Prevalence 20–25%
CAG-83	Prevalence 8–12%	Prevalence 12–17%

The adenoma group is characterised by the following changes: ↓ *Faecalibacterium* (~10 % vs. ~40 %), ↑ *Bacteroides* (~59 % vs. ~25 %), ↑ *Phocaeicola*. In contrast, in the control group, the following are more common: *Faecalibacterium*, *Anaerostipes* and other SCFA-producing microorganisms. These changes indicate a dysbiosis of the microbiota in the adenoma group.

2.5 Diversity analysis at the genus level

2.5.1 Alpha diversity

No statistically significant differences were found between groups in any of the alpha diversity indicators (Shannon, Simpson, evenness; $p > 0.05$). This suggests that the overall diversity of the microbiota remains similar (Figure 2.2).

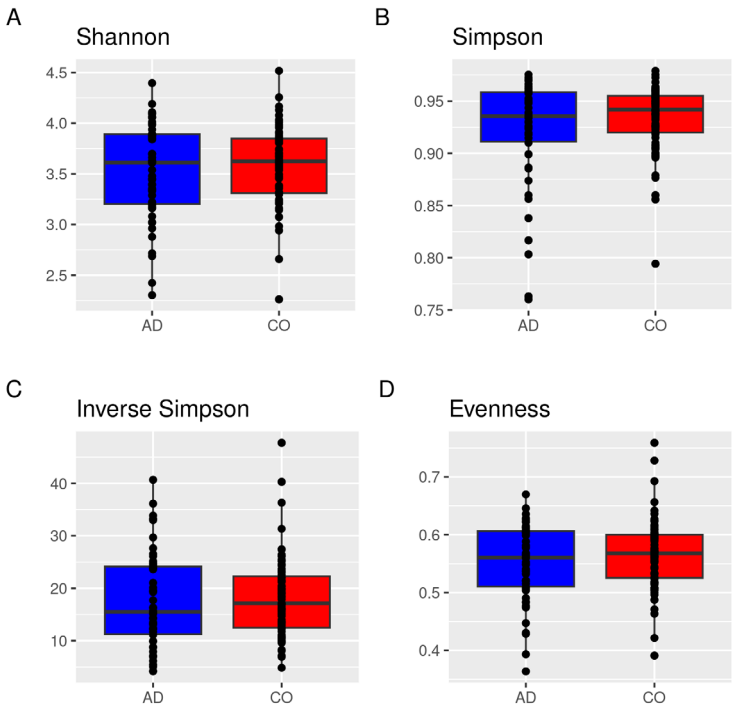


Figure 2.2 **Alpha diversity**

Table 2.13

Alpha diversity at the genus level

Measurement	Median (AD)	Median (CO)	p-value	Interpretation
Shannon index	3.6	3.8	0.31	No significant difference in richness + evenness
Simpson index	0.92	0.93	0.29	No significant difference in richness + evenness
Inverse Simpson index	15.4	14.8	0.25	No significant difference in richness
Evenness	0.58	0.6	0.34	No significant difference in evenness

2.5.2 Beta diversity at the species level

Beta-diversity analysis showed a statistically significant between-group difference ($p = 0.0002$; statistical test to be specified). AD samples also showed greater visual dispersion in ordination space (Figure 2.3).

Although there is no clear boundary between the two groups, the clustering of AD and CO samples along the PCA axes suggests subtle differences in the structure of the microbial community. Such clustering trends indicate structural differences in the microbial community, which are likely related to the presence of adenomas.

Figure 2.3 **Beta diversity**

To visualise the differences in microbiota composition, Figure 2.4 shows the distribution of relative abundance of microorganisms in adenoma and control groups.

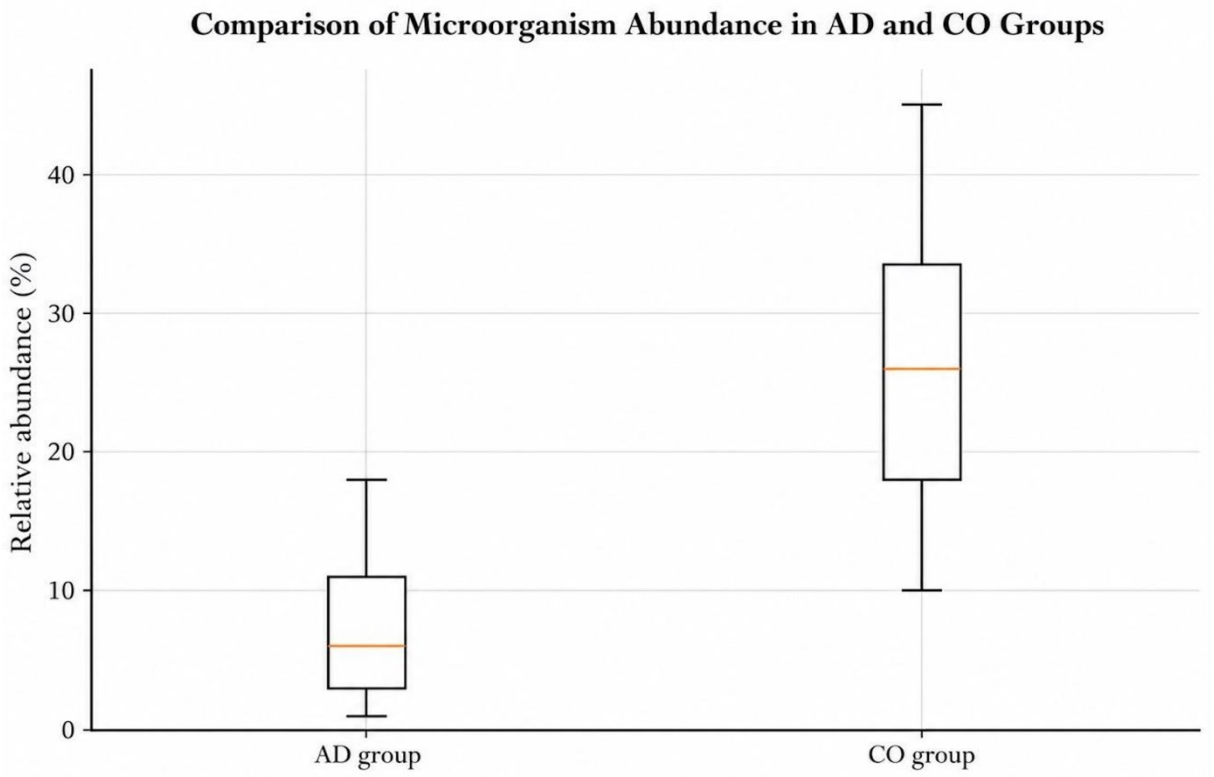


Figure 2.4 **Comparison of Microorganism Abundance in AD and CO Groups**

To quantitatively characterise the differences observed in Figure 2.4, Table 2.14 provides a comparison of the relative abundance of the main microbial genera between the groups.

Table 2.14

Statistical comparison of groups (mean values AD and CO)

Genus	AD group (%)	CO group (%)	CO–AD differences
Faecalibacterium	10.0	40.2	+30.2 (more CO gr)
Bacteroides	59.2	25.0	−34.2 (more AD gr)
Prevotella	18.0	23.0	+5.0 (more CO gr)
Phocaeicola	19.0	10.0–9.0	−9.0 (more AD gr)
Collinsella	6.0	12.8	+6.8 (more CO gr)
Anaerostipes	5.6	18.0	+12.4 (more CO gr)

2.6 Differential abundance analysis between patients with and without colorectal adenomas

After FDR correction, one statistically significant taxon was identified: *UBA7597* sp003448195 was significantly increased in the adenoma group (FDR = 0.0023). This result indicates a possible association of specific microorganisms with the development of adenomas.

In the AD group, compared with the CO group, the microorganism *UBA7597 sp003448195* showed a statistically significantly higher relative abundance (logFC = 3.4458; AveExpr = −2.6435; t = 5.1366; p = 1.03×10^{-6}; FDR = 0.0023; B = 4.6811) (see Figure 2.5).

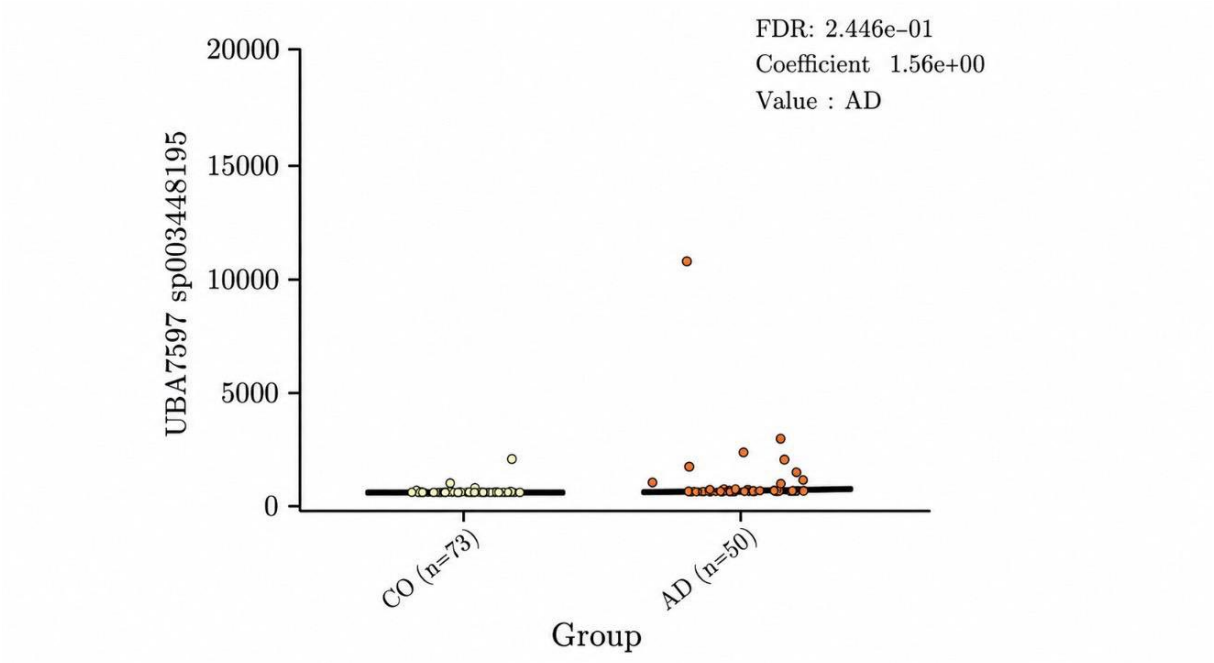


Figure 2.5 **imma differential-abundance analysis: relative abundance of UBA7597 sp003448195 in the AD and CO groups**

2.7 Microorganism genera with differences in relative incidence under the influence of the factors analysed in the study

In addition, an exploratory analysis was performed to evaluate the association of microorganisms with clinical and demographic factors (Table 2.15). FDR correction (Q < 0.25) was used in the analysis. In general, several potential associations between microorganisms and the analysed factors were observed, however, most of them did not reach statistical significance after multiple testing correction.

Ruminococcus sp.*900540005* showed an increased incidence in men (p = 7.00×10^{-5}) (Figure 2.6), but after FDR correction this association was not statistically significant (q = 0.209), therefore it should be interpreted with caution.

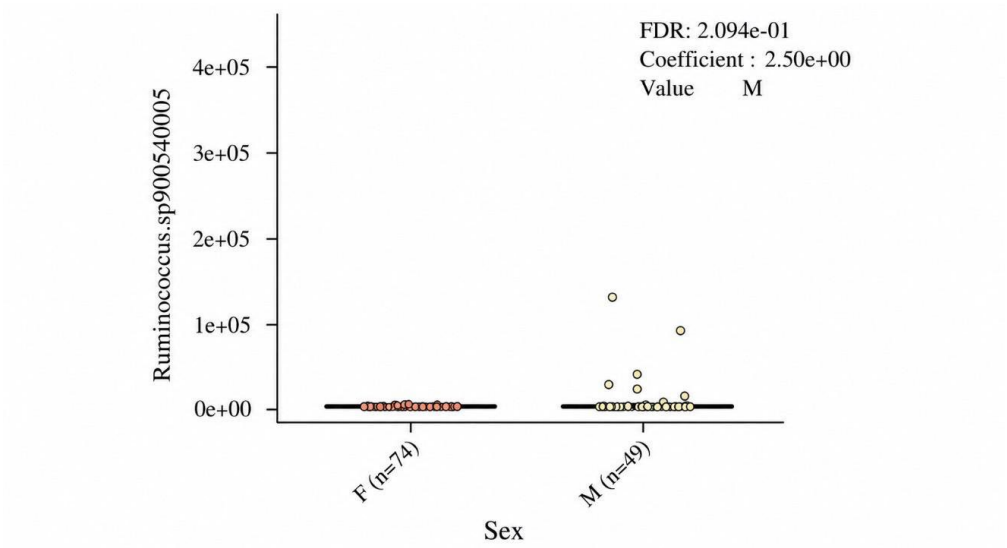


Figure 2.6 **Relative abundance of *Ruminococcus.sp900540005* by gender**

More pronounced trends were observed in relation to digestive tract diseases, however, these results also mostly failed FDR correction.

Table 2.15

Differentially abundant microbial genera

Taxa	Metadata	Value	Coef	Stderr	N	N.not.0	pval	Qval
Collinsella sp900541645	GI disease	Gastritis	−4.721	0.866	123	19	$2{,}97 \times 10^{-7}$	0.003
Collinsella sp900541645	GI disease	None	−4.600	0.847	123	19	$3{,}26 \times 10^{-7}$	0.003
Collinsella sp900541645	GI disease	GERD	−4.387	0.868	123	19	$1{,}68 \times 10^{-6}$	0.011
Sutterella faecalis	Smoking	Yes	6.369	1.420	123	34	$1{,}81 \times 10^{-5}$	0.091
Collinsella sp900541645	GI disease	Peptic ulcer	−4.865	1.137	123	19	$3{,}99 \times 10^{-5}$	0.161
CAG.103.sp004556965	GI disease	Peptic ulcer	−4.277	1.086	123	117	0,0001	0.209
Collinsella sp900754275	GI disease	None	−2.626	0.660	123	20	0,0001	0.209
Dialister sp900543455	GI disease	Gastritis	−7.243	1.796	123	69	0,0001	0.209
Gemmiger sp900539695	GI disease	Peptic ulcer	−6.619	1.627	123	123	$8{,}85 \times 10^{-5}$	0.209
Holdemania massiliensis	BMI	BMI	−0.208	0.053	123	117	0,0001	0.209
Lawsonibacter sp900066645	BMI	BMI	−0.144	0.036	123	121	0,0001	0.209
Parolsenella massiliensis	GI disease	Gastritis	−5.087	1.248	123	22	$8{,}63 \times 10^{-5}$	0.209
Parolsenella massiliensis	GI disease	None	−4.886	1.230	123	22	0,0001	0.209
RC9 sp900546395	BMI	BMI	−0.129	0.032	123	116	0,0001	0.209
Ruminococcus sp900540005	Sex	M	2.495	0.604	123	102	$7{,}00 \times 10^{-5}$	0,209

2.7.1 Relationship between genera of microorganisms and digestive tract diseases

Taxa associated with gastritis

A statistically significant negative association was found between *Collinsella* sp. and gastritis (q = 0.003), suggesting a possible reduced incidence or potential protective role of this taxon in gastritis (Table 2.17). Additionally, negative associations were also observed for *Dialister* sp. and *Parolsenella* sp., but these associations did not retain statistical significance after FDR correction (q = 0.209), and therefore should be interpreted with caution. The summary is presented in Table 2.16.

Table 2.16

Taxa associated with gastritis

Taxon	Coefficient	p value	q value	Interpretation
Collinsella sp.	−4.721	2.97E-07	0.003	Statistically highly significant negative association
Dialister sp.	−7.243	0.0001	0.209	Not q-significant. Could be chance
Parolsenella sp.	−5.087	8.63E-05	0.209	Not q-significant. Trend towards negative association

Only *Collinsella sp*. retained a statistically significant negative association with gastritis after FDR adjustment, while *Dialister* sp. *and Parolsenella sp*. did not retain this significance.

Taxa associated with GERD

A statistically significant negative association was found between *Collinsella* sp. and GERD (q = 0.011), indicating a reduced abundance of this taxon in patients with GERD. Despite the relatively low prevalence in the sample, this result is consistent with a similar negative association reported in gastritis, indicating a lower relative abundance of *Collinsella* sp. in patients with GERD; the biological significance of this association remains uncertain.

In the analysis, *Collinsella* sp. was significantly inversely associated with gastroesophageal reflux disease (GERD) (coefficient = −4.387; $p = 1.68 \times 10^{-6}$; q = 0.011). The negative coefficient indicates that a higher relative abundance of this taxon was associated with a lower likelihood of GERD. This finding supports the hypothesis that *Collinsella* sp. may have a potentially protective role; however, the observed association does not establish causality.

Taxa associated with peptic ulcer

Negative associations were observed between *Collinsella sp*., CAG-103 sp. and *Gemmiger sp*. and peptic ulcer (Table 2.17). Although p-values for all taxa indicated statistical significance, none of the associations remained significant after FDR correction (q > 0.05), so the results should be interpreted with caution. *Gemmiger sp*. had the largest effect size and high prevalence in all samples. This association is depicted in Table 2.17.

Table 2.17

Taxa associated with peptic ulcer

Taxon	Coefficient	p value	q value	Interpretation
Collinsella sp.	−4.865	3.99E-05	0.161	Statistically significant negative association, but not significant after adjustment; low prevalence
CAG-103 sp.	−4.277	0.0001	0.209	Statistically significant negative association, but not significant after adjustment; high prevalence
Gemmiger sp.	−6.619	8.85E-05	0.209	Statistically significant negative association, but not significant after adjustment; present in all samples

Although all three taxa showed negative associations with peptic ulcer, none of them retained statistical significance after FDR adjustment, so the results should be interpreted with caution. *Gemmiger* sp. had the largest effect size and wide prevalence.

2.7.2 Association between microbial genera and body mass index (BMI)

A negative association was observed between *Holdemania, Lawsonibacter* and *RC9* and BMI, suggesting that these taxa may be less common in individuals with higher BMI. Although p-values for all taxa indicated statistically significant associations, none of them retained significance after FDR correction (q = 0.209), so the results should be interpreted with caution (Table 2.18). These taxa were widely distributed in the analysed samples, indicating a possible trend, however, its significance needs to be confirmed in further studies. These results are visualised in Figure 2.7, where the scatter plot shows the changes in relative abundance depending on BMI.

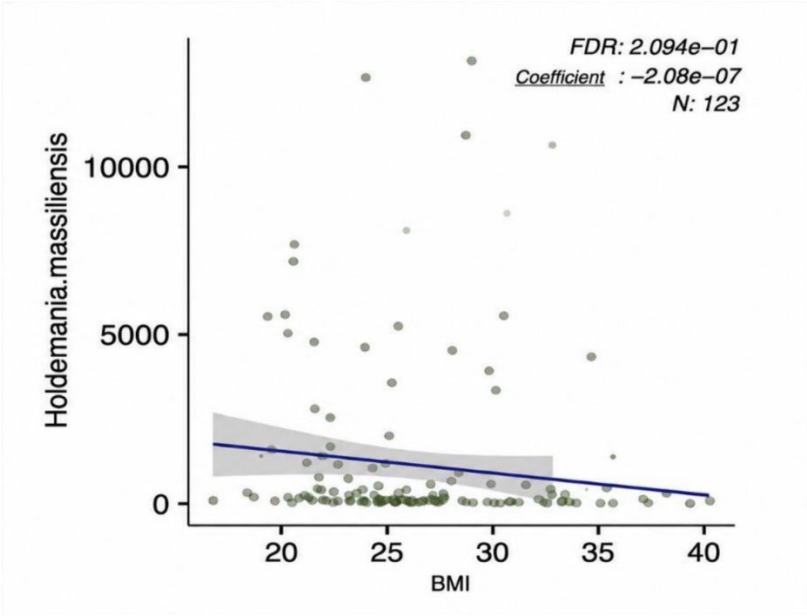


Figure 2.7 **Relative changes in abundance of *Holdemania massiliensis* relative to BMI, showing nominal p-value significance without FDR-adjusted q-value validation**

Table 2.18

Relationship between abundance of specific bacterial taxa and body mass index (BMI): regression coefficients and statistical significance

Taxon	Coefficient	FDR (q value)	Interpretation
Holdemania	−0.208	0.209	Moderately negative, not statistically significant
Lawsonibacter-	−0.145	0.209	Slightly negative, not statistically significant
RC9	−0.130	0.209	Slightly negative, not statistically significant

All three taxa show a consistent pattern in which they are more common in individuals with lower BMI, but the associations do not withstand FDR correction and should therefore be interpreted with caution.

2.7.3 Association between microbial genera and smoking

A positive association was observed between *Sutterella faecalis* and smoking, indicating an increased incidence of this taxon in smokers (Figure 2.8, Table 2.19). Although the p-value indicated statistical significance, this association did not remain significant after FDR correction (q = 0.091), so the results should be interpreted with caution (see Figure 2.8).

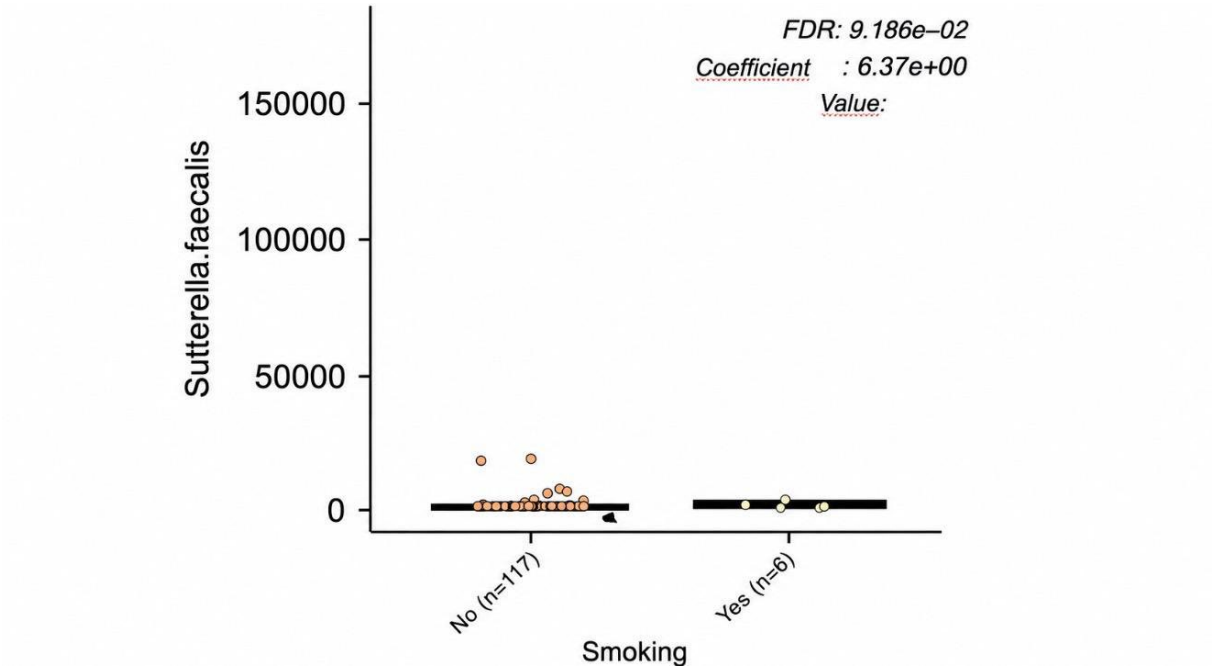


Figure 2.8 **Association between smoking and *Sutterella faecalis* prevalence (FDR > 0.05)**

Table 2.19

Association between smoking and *Sutterella faecalis* prevalence

Group	*Sutterella faecalis* prevalence	N	Interpretation
Smokers	Significantly higher	6	Strong positive association
Non-smokers	Lower	117	–
FDR	0.091	–	No statistical significance (q< 0.05)

2.7.4 Significant associations of microorganisms with clinical and demographic factors (heat map visualisation of MaAsLin2 analysis results)

The heat map (Figure 2.9) depicts the associations of microbial taxa with clinical and demographic factors based on MaAsLin2 analysis. The colour intensity reflects the direction and strength of the association: red indicates positive associations, and blue indicates negative associations, taking into account the q value. Only associations with $q < 0.25$ are included in the figure, which corresponds to the exploratory nature of the analysis and the sample size of the study.

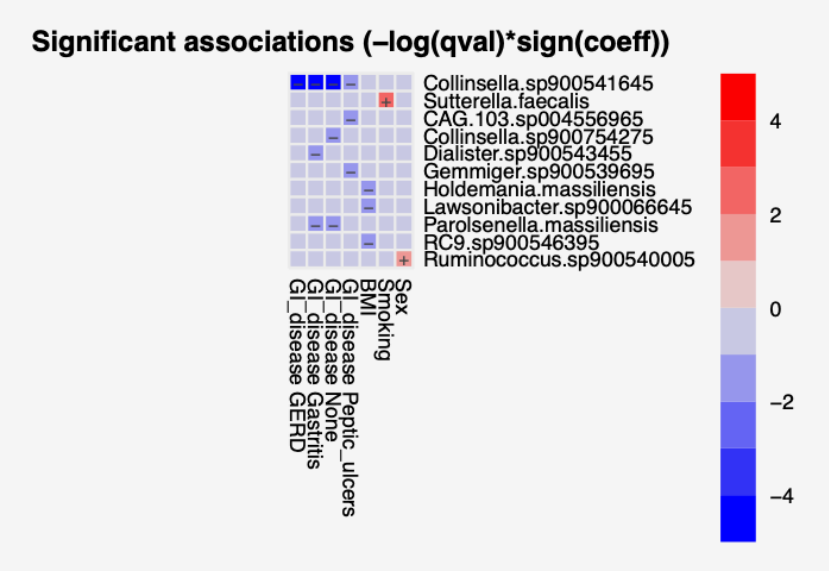


Figure 2.9 **Exploratory associations of bacterial taxa with clinical and demographic factors meeting the predefined q < 0.25 threshold** (MaAsLin2 analysis)

2.8 Microbial functional pathways meeting the exploratory MaAsLin2 threshold in patients with and without colorectal adenomas

To evaluate the functional differences of the microbiome depending on the presence of colorectal adenomas, 4511 microbial functions were analysed using MaAsLin2. Functional profiling was performed on metagenomic samples that passed quality control (n = 123). Of these, 243 showed deviations, but six functional pathways reached the exploratory threshold of $q < 0.25$; none reached FDR < 0.05. The results are summarised in Figure 2.10, which depicts the differences between the adenoma and control groups. The coefficient values reflect the direction and magnitude of the differences, while the FDR values – their statistical significance.

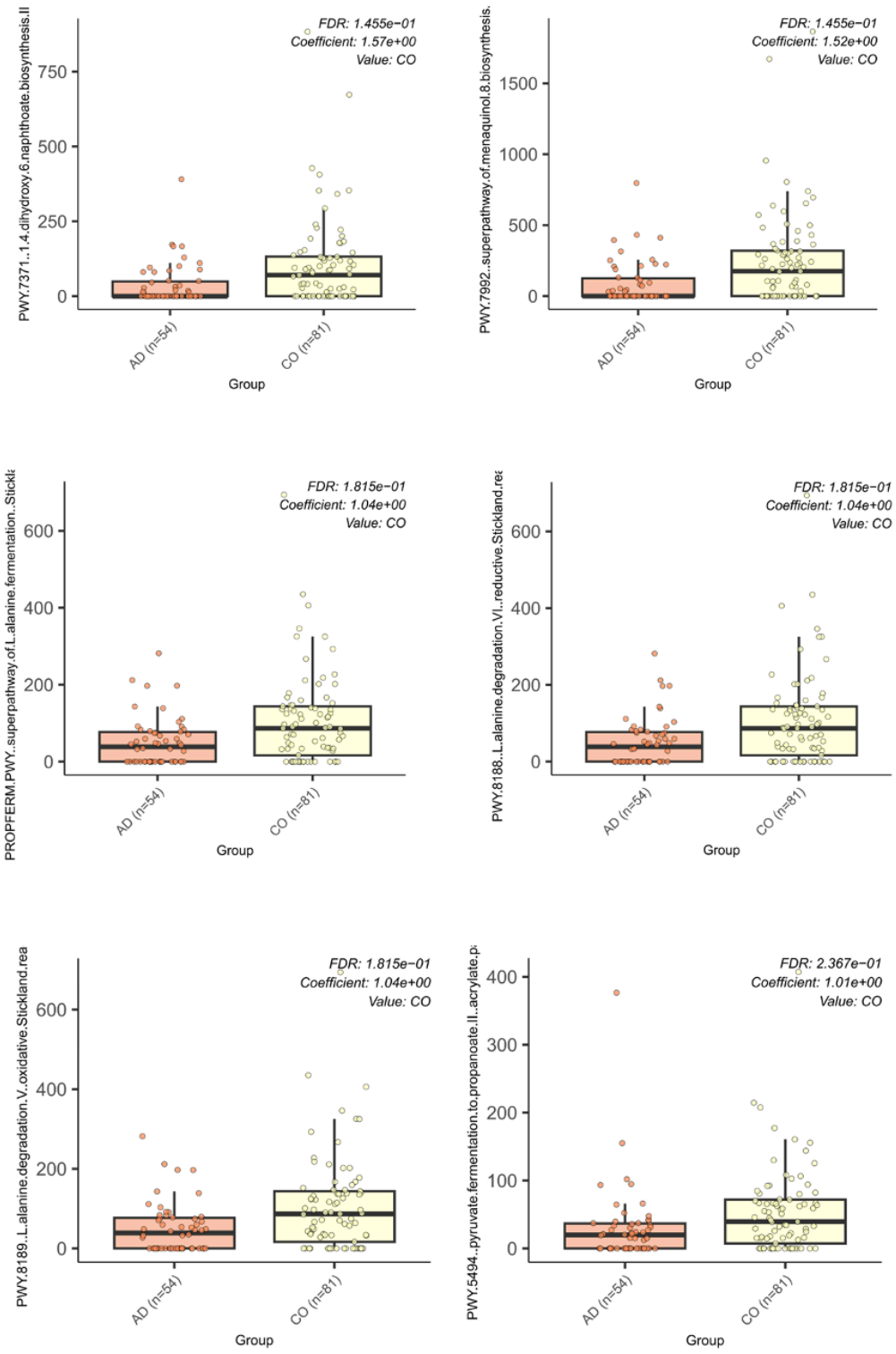


Figure 2.10 **Microbial functional pathways meeting the exploratory q < 0.25 threshold in the MaAsLin2 analysis**

2.8.1 PWY-7371 (1,4-dihydroxy-6-naphthoate biosynthesis II) pathway

The PWY-7371 pathway, which is involved in the biosynthesis of menaquinone (vitamin K_2), was with lower relative representation in patients with colorectal adenomas compared to the control group (Figure 2.10). Although the difference did not reach statistical significance after FDR correction (FDR = 0.1455), the effect size (coefficient = +1.57)

indicated a higher occurrence of the pathway in the control group. This trend may indicate a decrease in the functions of microorganisms involved in K_2 biosynthesis in adenomas, potentially reflecting functional changes in the microbiome in the development of early neoplasia (Table 2.20).

Table 2.20

Relative incidence and characteristics of the PWY-7371 (1,4-dihydroxy-6-naphthoate biosynthesis II) pathway between adenoma and control group

Indicator	Value
Pathway	PWY-7371 (1,4-dihydroxy-6-naphthoate biosynthesis II) pathway
Groups	AD, CO
Average abundance	More in the Co group
Prevalence	Most prevalent in the CO group, closely clustered in the AD group
Coefficient	+1.58 (higher in the CO group)
FDR	0.1455 (not statistically significant) (FDR > 0.05)
Interpretation	Reduced relative representation in the AD group
Potential significance	Dysbiosis and reduced functional potential

Table 2.20 summarises the relative abundance of the PWY-7371 pathway, statistical parameters, and their interpretation in adenoma and control groups.

2.8.2 Menaquinol-8 biosynthesis superpathway (PWY-7992)

The PWY-7992 pathway, which is involved in the biosynthesis of menaquinol-8 (vitamin K_2), was with lower relative representation in patients with colorectal adenomas compared to controls (Figure 2.10). Although the difference did not reach statistical significance after FDR correction (FDR = 0.1455), the effect size (coefficient = +1.52) indicated a higher abundance of this pathway in the control group. These results suggest a possible reduction in K_2 biosynthesis in adenomas and changes in the functional potential of the microbiome in early neoplasia (Table 2.21).

Table 2.21

Relative abundance and characterization of the menaquinol-8 biosynthesis superpathway (PWY-7992) in the adenoma and control groups

Indicator	Value
Pathway	Menaquinol-8 biosynthesis superpathway (PWY-7992)
Groups	AD, CO
Mean abundance	More in Co group
Prevalence	Most common in CO group CO > AD
Coefficient	+1.52 (higher in CO group)
FDR	0.1455 (not statistically significant, (FDR > 0.05))
Interpretation	Reduced menaquinone biosynthesis potential in AD group

Table 2.21 summarises the PWY-7992 pathway data, showing its reduced occurrence in the adenoma group. Although the difference is not statistically significant (FDR = 0.1455), the effect size indicates a possible reduction in the functional potential of the microbiome.

2.8.3 PROPFERM-PWY: L-alanine fermentation superpathway or Stickland reaction

The PROPFERM-PWY pathway, which is associated with L-alanine fermentation (Stickland reaction), was reduced in patients with colorectal adenomas compared to the control group (Figure 2.10). Although the difference did not reach statistical significance after FDR correction (FDR = 0.1815), the effect size (coefficient = +1.04) indicated a higher functional potential of this pathway in the control group. These results may indicate a decrease in anaerobic bacterial function and amino acid fermentation in adenomas, reflecting functional changes in the microbiome in early stages of neoplasia (Table 2.22).

Table 2.22

PROPFERM-PWY: L-alanine fermentation superpathway or Stickland reaction relative abundance and characteristics between adenoma and control group

Indicator	Value
Pathway	PROPFERM-PWY: L-alanine fermentation superpathway, also known as the Stickland reaction
Groups	AD, CO
Mean abundance	More in CO group Greater relative representation observed in CO group CO>AD
Prevalence	Most prevalent in CO group
Coefficient	+1.04 (higher in CO group), Higher functional potential in control group
FDR	0.1815 (not statistically significant) (FDR > 0.05)
Interpretation	Microbial amino acid lower functional potential for fermentation in AD group

2.8.4 PWY-8188: L-alanine degradation VI (reductive Stickland reaction)

The PWY-8188 pathway involved in L-alanine degradation (reductive Stickland reaction) was with lower relative representation in patients with colorectal adenomas compared to controls (Figure 2.10). Although the difference did not reach statistical significance (FDR = 0.1815), the effect size (coefficient = +1.04) indicated a higher functional potential of this pathway in the controls. These results are consistent with the trends observed in other fermentation and menaquinone pathways and may indicate a decrease in the function of anaerobic microorganisms in adenomas (Table 2.23).

Table 2.23

PWY-8188: L-alanine degradation VI (reductive Stickland reaction). Relative abundance and characteristics between adenoma and control group

Indicator	Value
Pathway	PWY-8188: L-alanine degradation VI (reductive Stickland reaction)
Groups	AD, CO
Mean abundance	More in CO group Greater relative representation observed in CO group CO > AD
Prevalence	More prevalent in CO group
Coefficient	+1.04 (higher in CO group), Higher functional potential in control group
FDR	0.1815 (not statistically significant) (FDR > 0.05)
Interpretation	Reduced microbial functions in AD group

2.8.5 PWY-8189: L-alanine degradation V (oxidative Stickland reaction)

The PWY-8189 pathway, which represents the oxidative branch of the Stickland reaction, was with lower relative representation in patients with colorectal adenomas compared to controls (Figure 2.10). Although statistical significance was not reached after FDR correction (FDR = 0.1815), the effect size (coefficient = +1.04) suggests a higher functional potential of this pathway in the controls. This trend is consistent with other Stickland fermentation pathways and may indicate a decrease in the metabolic functions of anaerobic microorganisms in adenomas (Table 2.24).

Table 2.24

PWY-8189: L-alanine degradation V (oxidative Stickland reaction) Relative abundance and characteristics between adenoma and controls

Indicator	Value
Pathway	PWY-8189: L-alanine degradation V (oxidative Stickland reaction).
Groups	AD, CO
Mean abundance	More CO in CO group Greater relative representation is observed in CO group CO > AD
Prevalence	Most common in CO group
Coefficient	+1.04 (higher in CO group), higher functional potential in control group
FDR	0.1815 (not statistically significant) (FDR > 0.05)
Interpretation	Loss of anaerobic microbial function in AD group, consistent with other Stickland and menaquinone functions

2.8.6 PWY-5494: Fermentation of pyruvate to propanoate II (acrylate pathway).

The PWY-5494 pathway, which is involved in the formation of propionate (short-chain fatty acid) in the acrylate pathway, was with lower relative representation in patients with colorectal adenomas compared to controls (Figure 2.10). Although statistical significance was not reached (FDR = 0.2367), the effect size (coefficient = +1.01) indicates a higher functional potential of this pathway in the control group. These results may indicate a decrease in

the function of SCFA-producing microorganisms in adenomas and a possible weakening of intestinal defence mechanisms (Table 2.25).

Table 2.25

PWY-5494: Pyruvate fermentation to propanoate II (acrylate pathway). Relative abundance and characteristics between adenoma and control group

Indicator	Value
Pathway	PWY-5494: Pyruvate fermentation to propanoate II (acrylate pathway).
Groups	AD, CO
Mean abundance	More CO in CO group Greater relative representation is observed in CO group CO > AD
Prevalence	More prevalent in CO group
Coefficient	+1.01 (higher in CO group), higher functional potential in control group
FDR	0.2367 (not statistically significant) (FDR > 0.05)
Interpretation	Reduced functional potential for propionate biosynthesis in AD group

2.8.7 Comparative analysis of six microbial metabolic pathways

Analysis of six microbial metabolic pathways showed a consistent decrease in these functions in patients with colorectal adenomas compared to controls. Although none of the associations reached statistical significance after FDR correction, the effect sizes (coefficients +1.01 to +1.57) indicate biologically significant trends. The control group had higher functional potential in the menaquinone (vitamin K_2) biosynthetic pathways (PWY-7371, PWY-7992), the Stickland fermentation pathways (PROPFERM-PWY, PWY-8188, PWY-8189), and the propionate formation pathway (PWY-5494). This suggests a possible decrease in the capacity for anaerobic metabolism, vitamin biosynthesis, and short-chain fatty acid (SCFA) formation in the adenoma group. The consistent reduction of these interconnected pathways may reflect early functional changes in the microbiome and dysbiosis in the development of colorectal neoplasia (Table 2.26).

Table 2.26

Comparative analysis of six microbial metabolic pathways between patients with colorectal adenomas and controls

Pathway	Pathway ID	Metabolic function	Predominant group	Coefficient	FDR	Role	Interpretation
1,4-dihydroxy-6-naphthoate biosynthesis Il	PWY-7371	Menaquinone (vitamin K2) biosynthesis	CO	+1.57	0.1455	Electron transport, intestinal microbial respiration	Reduced potential for vitamin K2 biosynthesis AD group
Super-pathway of menaquinol-8 biosynthesis	PWY-7992	Mena-quinone (vitamin K2) biosynthesis	CO	+1.52	0.1455	Vitamin K2 production, microbial metabolism	Supports the trend of impaired electron transport metabolism AD
Super-pathway of L-alanine fermentation (Stickland)	PROPFERM-PWY	Amino acid fermentation	Co	+1.04	0.1815	Redox balance, energy production by Clostridia	Suggests loss of anaerobic enzyme core: AD
L-alanine degradation VI (reductive Stickland)	PWY-8188	Amino acid fermentation	CO	+1.04	0.1815	Stickland reaction reductive pathway	Functional loss of amino acid-based energy production AD
L-alanine degradation V (oxidative Stickland)	PVT-8189	Amino acid fermentation	CO	+1.04	0.1815	Oxidative part of Stickland reaction	Confirms anaerobic metabolism decline in AD
Pyruvate fermentation to propanoate II (acrylate pathway)	PWY-5494	SCFA (pro-pionate) production	CO	+1.01	0.2367	Anti-inflammatory SCFA generation	Reduced SCFA production potential in AC microbiota

Discussion

The whole-genome shotgun metagenomic sequencing used in this study provided high taxonomic resolution and allowed for the analysis of not only the composition of the microbiota, but also its functional potential. Unlike 16S rRNA sequencing, this approach allows for the identification of microorganisms to the species level and the characterisation of metabolic pathways that may be involved in intestinal homeostasis and neoplastic progression. However, it should be noted that metagenomics reflects functional potential, not real activity, and the results may be affected by database limitations and technical factors. The analysis used CLR transformation, Aitchison distance, and multiple hypothesis correction to ensure the reliability of the results. This doctoral thesis analyses the composition and functions of the intestinal microbiota in individuals with and without colorectal adenomas to identify potential microbiome biomarkers for early diagnosis. Alpha and beta diversity, as well as taxonomic and functional profiles, were assessed, taking into account relevant covariates (gender, age, BMI, smoking, etc.). Although no significant differences were found in alpha diversity, specific changes in taxa and microbiome functions were identified, which may indicate the involvement of the microbiota in adenoma development and its potential as a non-invasive biomarker for early diagnosis of colorectal cancer.

Patient characteristics

The study included 136 patients after colonoscopy, of whom 56 (41.18 %) had at least one colorectal adenoma, while 80 (58.82 %) had no adenomas. This prevalence is consistent with the rates observed in screening populations (Lieberman et al., 2007) and is suitable for the study of early adenomas. Patients in the adenoma group were statistically significantly older (53.03 ± 5.4 vs 45.61 ± 3.0 years; $p < 0.001$), which is consistent with the known role of age in the development of adenomas (Leslie et al., 2002). Adenomas were also more frequently detected in men ($p = 0.027$), which is consistent with epidemiological data (Zhou et al., 2018; Kim et al., 2019). The adenoma group was more likely to live in a capital city ($p < 0.001$), which may be related to urban factors and changes in microbiota (Sonnenburg & Sonnenburg, 2019; Vangay et al., 2018). This group was also more likely to have higher education ($p < 0.001$), which may reflect indirect lifestyle factors such as diet or sedentary work (O'Keefe et al., 2015). Contrary to the literature (Wolin et al., 2009), physical activity was more likely to be observed in the adenoma group ($p < 0.001$), which may be explained by confounding factors or behavioural changes after diagnosis. No significant differences were found in smoking ($p = 0.28$), possibly due to low prevalence (Botteri et al., 2008), but BMI was higher in the adenoma group ($p = 0.039$), consistent with the known risk (Omata et al., 2013).

High alcohol consumption was significantly more common in the adenoma group ($p < 0.001$), consistent with the literature on its role in the development of colorectal adenomas (Bagnardi et al., 2015). The majority of patients had low-risk adenomas (85.7 %), mostly tubular and medium-sized (5–9 mm), consistent with screening data (Click et al., 2018). The results confirm that the development of adenomas is associated with the interaction of several factors. In addition to traditional risk factors, there is also significant potential for the integration of new approaches, including the use of artificial intelligence in polyp detection (Vilkoite et al., 2023) and the use of microbiome biomarkers to improve screening accuracy (Vilkoite et al., 2022).

Microbiome composition at the genus level

Shotgun metagenomic analysis of 123 patients (AD and CO groups) revealed a typical Western-type microbiome dominated by the genera *Faecalibacterium*, *Bacteroides*, *Blautia*_A, *Alistipes*, *Prevotella* and *Roseburia* (Lozupone et al., 2012; Charbonneau et al., 2016). Descriptive differences in genus-level microbiome composition were observed between the groups This finding aligns with previous studies on its association with inflammation and precancerous changes (Cao et al., 2014; Martin et al., 2023). Conversely, *Bacteroides* showed a higher relative abundance in the AD group, a shift potentially associated with the production of procarcinogenic metabolites and the activation of inflammatory pathways (Chen et al., 2014; Cheng et al., 2020).

Phocaeicola was also more frequently found in the AD group, which may indicate its possible role in inflammatory and metabolic processes (Hu et al., 2023), while the methanogenic archaea *Methanomassiliicoccus*_A was slightly more frequent in the AD group (Cisek et al., 2024). At the same time, the butyrate-producing genera *Blautia*_A and *Gemmiger* were reduced, pointing to a weakening of protective mechanisms (Louis et al., 2014). Contrary to some literature data, *Prevotella* and *Alistipes* did not differ significantly between groups (Fu et al., 2024; Lo et al., 2022).

Overall, the observed changes – decreased butyrate producers and increased potentially proinflammatory taxa – indicate dysbiosis, which may contribute to the development of adenomas. Similar trends were also described by Deng et al. (2025), where *Bacteroides* was enriched, but *Faecalibacterium* was reduced in the polyp group.

Alpha diversity

Alpha diversity describes the richness and evenness of microbial species within a sample and is considered an indicator of microbiome stability and health (Lozupone et al., 2012). In this study, it was analysed using Shannon, Simpson, and other indices, comparing patients with colorectal adenomas (AD) and controls (CO).

The results showed no statistically significant differences between the groups in any of the indicators, suggesting that the presence of adenomas is not associated with changes in the overall diversity of the microbiome. These data are consistent with other studies in which alpha diversity is often unchanged in the early stages of colorectal carcinogenesis (Zeller et al., 2014; Deng et al., 2025), although a slight decrease has been described in individual cases (Ai et al., 2019).

The results obtained indicate that more significant changes in the context of adenomas may affect not the overall diversity of the microbiome, but its taxonomic composition and functional potential, especially in metabolic pathways, such as the synthesis of butyrate and other metabolites.

Beta diversity

Beta diversity characterises differences in the microbiome structure between individuals and is considered a more sensitive indicator of microbiota changes than alpha diversity, especially in the early stages of colorectal carcinogenesis (Zhang et al., 2022; Hale et al., 2016). In this study, using PCA analysis, statistically significant differences were found between patients with colorectal adenomas (AD) and the control group (CO) ($p = 0.0002$). The samples of the AD group showed a greater dispersion, indicating higher interindividual microbiome variability, while the CO group was characterised by a more stable and homogeneous microbiome profile. This suggests that in the case of adenomas, the microbiome structure becomes less stable and more prone to dysbiosis. At the taxonomic level, inflammation-related microorganisms, such as *Prevotella*, *Collinsella* and *Holdemanella*, which have previously been associated with intestinal barrier disruption and proinflammatory processes, were more prevalent in the AD group (Larsen et al., 2017). In contrast, butyrate producers, such as *Faecalibacterium prausnitzii* and *Anaerostipes*, which play an important role in maintaining intestinal homeostasis and reducing inflammation (Louis et al., 2017), dominated in the CO group (Louis et al., 2017). The observed changes – a reduced proportion of butyrate producers and an increased presence of potentially proinflammatory taxa – indicate a structural shift of the microbiome towards dysbiosis, which may contribute to a preneoplastic environment in the colon. These results are consistent with Deng et al. (2025), who also found statistically significant differences in beta diversity and adverse changes in the microbiota in patients with colorectal polyps. Overall, the data obtained confirm that structural changes in the microbiome may be an important early indicator in the development of colorectal adenomas and a potential biomarker for a personalised screening approach.

Differential abundance analysis between patients with and without colorectal adenomas

Differential abundance analysis is essential in microbiome studies because it allows the identification of taxa whose relative abundance differs between patients with colorectal adenomas (AD) and controls (CO) (Gloor et al., 2017; Weiss et al., 2017). This approach helps to identify potentially pathogenic or protective microorganisms and provides insight into disease mechanisms. The analysis used shotgun metagenomics data, which ensures high accuracy and reduces the impact of technical artifacts. Linear models (MaAsLin2, limma) were used to compare taxa, taking into account confounding factors (e. g. age, BMI), and multiple testing correction with FDR was performed. Log fold change (logFC) was used to interpret the results: a positive logFC indicates a higher abundance of the taxon in a particular group, a negative one indicates a reduced abundance. Results with FDR < 0.25 were considered statistically significant. A total of 123 patient samples were analysed.

MaAsLin2 analysis did not reveal statistically significant associations between the occurrence of specific microorganisms and colorectal adenomas. Such a result may indicate high interindividual variability of the microbiome or insufficient statistical power (Mallick et al., 2021). In contrast, limma analysis, which is particularly suitable for high-dimensional data, revealed one statistically significant species – *UBA7597* sp003448195, which was significantly more common in the AD group (logFC = 3.45, FDR = 0.0023). The identity of this bacterium has not been extensively described, but its increased occurrence may indicate a hitherto insufficiently studied role of this taxon in the pathogenesis of colorectal adenomas. These data echo examples described in the literature, where atypical or non-specific intestinal microorganisms are identified as potential disease indicators (Pasolli et al., 2019). The adjusted p-value of 0.0023 indicates that this finding is statistically significant after correction for multiple testing (using the false discovery rate – FDR).

Given that the adjusted p-value is much lower than 0.25 (the significance threshold used by MaAsLin2), it can be concluded that the increase in the abundance of this microorganism is not due to chance changes. The logFC value of 3.4458 indicates a relatively large change in the AD and CO groups, as it is measured on a logarithmic scale – it can be concluded that *UBA7597* sp003448195 is more common in the AD group. The p-value of 1.03E-06 further confirms that there is a large difference between the groups, supporting the claim of a significantly different abundance.

UBA7597 sp003448195 is more common in patients with colorectal adenomas. The increase is statistically significant, and the adjusted p-value supports the reliability of this result.

Such a discrepancy between MaAsLin2 and limma results is not uncommon in microbiome studies – different statistical models (e. g., variance assumptions, normalisation methods, and inclusion of confounding factors) can seriously affect the final results (Nearing et al., 2022). Therefore, it is recommended to use more than one analysis tool to ensure the stability of the results.

The MaAsLin2 analysis did not reveal significant differences between groups, which could be explained by several reasons. Since MaAsLin2 focuses on the associations of individual taxa with specific variables, it may not capture the overall diversity of microbiota structure between groups. One possible methodological difficulty may be the small number of samples, which limits statistical power and reduces the ability of the analysis to detect significant differences. Also, if the microbial communities in the two groups are very similar, detecting differences may be more difficult.

Several environmental and individual factors, such as diet, age, medication use or lifestyle factors, can affect the composition of the microbiota and mask real differences between groups. Technical aspects, such as uneven sequencing depth – if some samples have more reads than others, this can affect the results. Another important aspect is statistical adjustments, such as false discovery rate (FDR) control, which are applied to reduce the number of false positives, but can sometimes be too stringent, especially in the case of small sample sizes.

It should also be taken into account that the human microbiome is highly dynamic and is influenced by various independent factors, which create so-called biological noise and can suppress differences between groups. Rare microorganisms that could potentially be important are sometimes excluded from the analyses or their statistical significance is not reached. Furthermore, if too many irrelevant covariates are included in the analysis, or if significant covariates are not included, this may reduce the ability to detect significant associations. It is also possible that changes in the overall microbiota structure, such as alpha or beta diversity, are more pronounced than changes in specific taxa. Since MaAsLin2 focuses specifically on individual microorganisms, this method may not be sufficient to detect overall variation in the microbiota profile between groups.

The significantly higher abundance of *UBA7597* sp003448195 within the AD group constitutes a novel finding, although its biological role and functional implications remain poorly understood. Given its statistical significance and consistent presence in only one group, this taxon may serve as a biomarker or indicator of microbiome dysbiosis during colorectal adenoma development. Overall, this analysis identifies a single differentially abundant microorganism with potential diagnostic or even pathogenetic relevance. These results underscore the necessity of validating *UBA7597* sp003448195 in larger, independent cohorts.

The relative occurrence of *UBA7597* sp003448195 is significantly increased in patients with colorectal adenomas (log2 fold change = +3.45; FDR = 0.0023). UBA7597 sp003448195 is a taxon defined within the GTDB classification and represented by a metagenome- assembled genome (MAG), corresponding to a previously uncultivated microorganism. Therefore, its official taxonomic name, as well as its genetic and functional precise affiliation, is currently tentative, and it is not precisely linked to classical genera or species according to current classifications.

Contrary to the traditional bacteriological nomenclature, MAG discoveries are increasingly used in microbiome research, as they allow the identification and analysis of microorganisms that are difficult or impossible to cultivate outside the host. Such newly identified MAG entities have revealed both previously unknown functions of the microbiota and potential links to clinical conditions (Pasolli et al., 2019; Almeida et al., 2021).

The relative increase of *UBA7597* sp003448195 in the adenoma group may reflect specific shifts in microbiome composition associated with neoplastic processes in the colon. It is possible that this MAG represents a group of anaerobic bacteria with unique metabolic capacities that may be associated with microenvironmental conditions that support the development or progression of adenomatous tumours. The uncertainty surrounding the precise taxonomic positioning of this taxon indicates the need for further taxonomic validation, including broader phylogenetic analysis and functional annotation of the genome to determine its metabolic properties and potential role in host pathophysiology. Similarly, recently identified and metagenomic-based taxa have proven to be valuable biomarkers in other clinical conditions, including inflammatory bowel disease studies (Franzosa et al., 2019; Duvallet et al., 2017), where MAG identification provides the opportunity to uncover functionally relevant, previously unknown microbial groups. *UBA7597 sp003448195* should currently be regarded as a potential microbiome biomarker candidate. Its association requires validation in independent cohorts and targeted quantitative assays before any diagnostic or prognostic application can be considered

From a translational perspective, a biomarker does not need to be identified solely as a fully cultured microorganism; on the contrary, metagenomic characterisation has successfully identified several microbiome markers that were subsequently associated with clinical phenotype (Thomas et al., 2019; Zeller et al., 2014). Therefore, *UBA7597* sp003448195 is a promising candidate for further validation, especially if it can be confirmed in independent cohorts and possibly by quantitative diagnostic methods (e. g., qPCR-based measurement).

Although the specific biological function of this taxon has not yet been characterised, it is reasonable to expect that it may be related to enzymatic metabolic pathways, redox

reactions, or microenvironmental signalling, which are associated with early neoplastic processes. Since the study also observed reduced Stickland fermentation and short-chain fatty acid biosynthesis potential in the adenoma group, the relative abundance of *UBA7597* sp003448195 may be part of a broader metabolic shift that distinguishes adenoma from control profiles. Future studies would be valuable to precisely place *UBA7597* sp003448195 using genome-wide phylogenetic analysis, investigate its potential metabolic pathway properties using metatranscriptomics or metaproteomics approaches, validate the association with adenoma status in independent populations, and develop quantitative diagnostic tools that include this MAG as a potential biomarker.

Using LEfSe and metagenomeSeq, Deng et al. (2025) identified 48 statistically significant, differentially abundant operational taxonomic units (OTUs), the majority of which (34) were enriched in patients with polyps. These taxa included, for example, *Clostridium perfringens*, *Escherichia coli* and *Ruminococcus gnavus*. These findings are broadly consistent with taxonomic alterations reported by Deng et al. (2025), while the direction and biological relevance of these associations require further validation.

Microbial genera with significant differences in relative abundance under the influence of the factors analysed in the study

Analysis of changes in microbiota composition depending on clinical and demographic factors was performed using linear mixed effects models (MaAsLin2), which allows for the consideration of confounding variables and provides a more accurate interpretation of the results (Mallick et al., 2021; Gloor et al., 2017).

The most pronounced association was found for *Collinsella* sp., which was found to be statistically significantly negatively associated with gastritis and GERD. Given the association of this genus with lipid metabolism and intestinal mucosal function, its depletion may reflect changes in the microbiome in the context of inflammatory and upper gastrointestinal diseases (Gomez et al., 2018). Similar, but non-significant trends following FDR correction were also observed in the case of *Dialister* and *Parolsenella*, indicating a possible, but not yet proven, role of these taxa (Duvallet et al., 2017).

Regarding BMI, negative associations were found with Holdemania, Lawsonibacter and RC9, which may indicate a decrease in these taxa at higher body mass index. Such trends are consistent with previous studies describing changes in the microbiota in obesity, especially in relation to short-chain fatty acid metabolism (Dao et al., 2016).

Sex-specific differences were observed in Ruminococcus sp., which is more common in men, possibly due to hormonal and immunological factors (Wu et al., 2024). *Sutterella faecalis*, on the other hand, was more common in smokers, which is consistent with data on

the effects of smoking on the gut microbiota, especially in relation to mucosal permeability and immune response (Allais et al., 2016).

Although some of the identified associations did not reach statistical significance after correction for multiple testing, the observed trends are biologically plausible and consistent with the literature. They may serve as a basis for further studies and potential microbiome biomarkers for personalised medicine. In the study by Deng et al. (2025), patients with colorectal polyps demonstrated an increased relative abundance of *Escherichia coli.* Certain *E. coli* strains have been associated with processes involved in colorectal carcinogenesis; therefore, this finding may indicate microbiota compositional changes associated with the presence of polyps and supports further investigation of *E. coli* as a potential microbiome biomarker.

Functional differences in the microbiome between patients with and without colorectal adenomas

Analysis of the functional profile of the microbiome revealed a relative reduction in six microbial metabolic pathways in patients with colorectal adenomas (AD group) compared to the control group (CO). Although none of the differences reached statistical significance after FDR correction (all FDR > 0.05), all six pathways had strong positive effect ratios (from +1.01 to +1.57), indicating a higher occurrence of these functions in the control group. Such results suggest biologically relevant changes in gut microbiome functions that may be associated with early neoplastic changes.

Menaquinone (vitamin K2) biosynthetic pathways (PWY-7371 and PWY-7992)

Both pathways – PWY-7371 (1,4-dihydroxy-6-naphthoate biosynthesis II) and PWY-7992 (menaquinol-8 biosynthetic superpathway) were more abundant in the CO group (coefficients +1.57 and +1.52, respectively; FDR = 0.1455 for both). Menaquinone is an essential component of bacterial respiration, acting as an electron transporter and involved in both bacterial metabolism and the regulation of host inflammation and oxidative homeostasis (Brooijmans et al., 2009). The lower relative abundance of these pathways in the AD group may indicate a reduced functional potential for menaquinone biosynthesis within the microbial groups in which these pathways are represented. As metagenomic analysis characterizes functional potential, these data do not allow direct conclusions to be drawn regarding actual vitamin K2 synthesis activity or vitamin K2 concentrations.

Menaquinone is an important electron carrier in bacterial respiratory chains. In the present study, the lower relative representation of PWY-7371 and PWY-7992 in the adenoma group suggests a reduced functional potential for menaquinone biosynthesis

within microbial groups carrying these pathways. However, shotgun metagenomics measures gene and pathway representation rather than pathway activity or metabolite concentrations. Therefore, these findings do not demonstrate reduced vitamin K_2 production or vitamin K_2 deficiency. They should be interpreted as exploratory, hypothesis-generating functional signals requiring validation by targeted metabolomic, metatranscriptomic, or other functional approaches

Stickland fermentation pathways (PROPFERM-PWY, PWY-8188, PWY-8189)

The three Stickland-related pathways showed lower relative representation in the AD group (coefficient +1.04; FDR = 0.1815 for each pathway), indicating an exploratory trend toward lower functional potential for anaerobic amino-acid metabolism. Stickland fermentation is specific to bacteria of the order Clostridiales and provides an important mechanism for anaerobic energy production (Buckel et al., 1974). The observed reduction in these pathways suggests a possible loss of metabolic potential of anaerobic bacteria, which may be related to dysbiosis and inflammation-related changes in the microenvironment (Liu Y et al., 2022).

Acrylate pathway (PWY-5494: pyruvate fermentation to propanoate II)

This pathway is one of the ways in which microorganisms synthesise propionate, a short-chain fatty acid (SCFA) with known anti-inflammatory and colonocyte-protective effects (Louis et al., 2017). Lower functional potential for propionate biosynthesis in the AD group (coefficient = +1.01, FDR = 0.2367) complements previous evidence of a loss of functional microorganisms associated with intestinal barrier integrity and inflammation control (Canani et al., 2011).

Although none of the six pathways reached the conventional FDR < 0.05 threshold, all showed a consistent direction toward lower relative representation in the adenoma group. This represents a biologically interesting, hypothesis-generating pattern that requires validation in independent cohorts. The consistency of the direction across biologically related pathways supports further investigation of this exploratory signal (Thomas et al., 2019; Gagnière et al., 2016).

Several studies have indicated that microbiome functions can serve as more sensitive biomarkers of digestive tract diseases than taxonomic components alone (Franzosa et al., 2019). Our results support this statement, emphasizing the fact that in patients with colon adenomas not only changes in the composition of microorganisms are observed, but also significant changes in their functional manifestations are detected. Of particular importance are the observed disturbances in metabolic processes related to microbial energy production, ensuring the integrity of the intestinal epithelium and the regulation of low-level inflammation.

Analysis of microbiome functions allowed the identification of several metabolic activities of microorganisms that are reduced in cases of colorectal adenomas. The obtained results indicate the need to further develop functional microbiome profiling as a potential biomarker for early screening diagnostics. Larger sample analyses are needed to confirm these functions as part of a personalised approach for early detection and prevention of colorectal cancer.

Deng et al. (2025) emphasise the presence of structural imbalance in the composition of the microbiota – with a decrease in beneficial bacteria (e. g. Bifidobacterium, *Faecalibacterium*) and an increase in pathogenic bacteria (e. g. Escherichia coli, Fusobacteria). In our study, a similar trend was reflected in the functional annotation with HUMAnN3, where pathways related to lipopolysaccharide biosynthesis and oxidative stress dominated in the adenoma group. These results indicate functional alterations of the microbiome as an important component of the pathogenesis of CRC.

Despite the differences in the sequencing methods used (16S rRNA/ shotgun metagenomics), both approaches consistently indicate the importance of the gut microbiota in the pathogenesis of colorectal formations. Future studies with combined 16S rRNA and metagenomic data analysis could provide higher taxonomic precision and allow for the synthesis of both compositional and functional aspects in assessing the potential of the microbiome as a biomarker (Deng et al., 2025).

From a clinical perspective, the functional profile of the microbiome may be a more promising source of biomarkers than the relative abundance of individual taxa. Functional pathways reflect the metabolic capacity of the microbiota regardless of the specific taxon that provides this function. This is especially important given the interindividual variability of the microbiome. Therefore, in future screening approaches, functional pathway analysis could complement colonoscopy or faecal testing, creating combined risk models for early detection of adenomas.

Conclusions

1 Patients with colorectal adenomas were significantly older, had a higher BMI, were more likely to live in the capital, and more frequently had higher education. The proportion of men was numerically higher in the adenoma group, although this difference did not reach statistical significance. Unexpectedly, the proportion of physically active individuals was higher in the adenoma group (73.2% vs. 17.5%; $p < 0.001$); this finding should be interpreted cautiously considering potential confounding, selection characteristics, and differences in the assessment of physical activity.

2 Descriptive genus-level profiles showed higher relative abundance of *Bacteroides* and *Phocaeicola* and lower abundance of *Faecalibacterium* and *Anaerostipes* in the adenoma group. Beta diversity differed between groups, whereas alpha diversity did not show significant differences, indicating differences primarily in microbiome composition rather than overall within-sample diversity

3 limma differential-abundance analysis identified *UBA7597 sp003448195* as the only taxon that remained statistically significant after FDR correction. Its relative abundance was higher in the adenoma group; however, because its biological function remains poorly characterised and the finding has not been independently validated, it should be regarded as a potential biomarker candidate

4 The structure of the microbiota correlates with some clinical variables. In particular, *Collinsella* spp. showed a significant association with digestive diseases (gastritis and GERD), but in general, most statistically significant associations lost significance after FDR correction, indicating the need for larger samples or better defined covariates.

5 In the adenoma group, a trend toward lower functional potential was observed across several microbial metabolic pathways, including menaquinone biosynthesis, Stickland alanine fermentation and degradation, and propionate formation. None of these pathways reached the conventional FDR < 0.05 threshold; therefore, the findings should be interpreted as exploratory, hypothesis-generating functional trends.

6 The trend toward a reduced functional potential of menaquinone biosynthesis pathways in the adenoma group may indicate alterations in energy metabolism within the microbial groups in which these pathways are represented. Menaquinone (vitamin K2) is an essential component of the bacterial electron transport chain; therefore, the lower relative abundance of these specific biosynthetic pathways may reflect changes in the functional potential of certain microorganisms, which could potentially be associated with intestinal mucosal homeostasis.

7 Decreased anaerobic fermentation pathways indicate impaired microbiota functionality. Stickland fermentation, especially that carried out by members of the Clostridioides genus, was impaired in the AD group. These pathways are important for energy balance and amino acid metabolism, and their reduction may be part of the early metabolic disorders of colorectal neoplasia.

8 The lower relative representation of the PWY-5494 pathway in the adenoma group may indicate reduced functional potential for propionate biosynthesis. This finding requires validation in independent cohorts and complementary functional analyses.

Proposals

1 Conduct further studies to assess how the composition and functionality of the gut microbiota change during adenoma development, from low-risk adenomas to adenocarcinoma development. Such an approach would allow us to identify changes in the microbiome over time and their potential role in carcinogenesis. Particular attention should be paid to the appearance or disappearance of specific bacterial taxa or functional pathways at certain stages of adenoma development. Such studies would help determine whether changes in the microbiome could serve as an early, non-invasive biomarker with diagnostic or prognostic significance. This would open up opportunities to use microbiome profiling for risk assessment and the development of personalised screening strategies.

2 In future studies, pay more attention to the study of the metabolic functions of the gut microbiota. With particular attention to the synthesis of metabolites such as vitamin K2, butyrate, ammonia and hydrogen sulphide (H_2S) and their relationship with the development of colon adenomas. These metabolites may significantly affect the integrity of the intestinal mucosa and epithelial proliferation, thereby promoting or inhibiting the development and growth of adenomas. For example, butyrate has anti-inflammatory and antitumour properties, while ammonia and H_2S may have toxic effects on the epithelium. Functional microbiome analysis may help identify risk factors even before the occurrence of histological changes. Such information would open the way to develop preventive or therapeutic intervention strategies by influencing the metabolic profile of the microbiome. Future studies should complement metagenomic profiling with targeted metabolomic measurements to quantify these metabolites directly.

3 Conduct further studies with targeted diets or specific probiotic preparations to promote the restoration of a healthy intestinal microbiota in patients with colorectal adenomas or at high risk of developing them. Diets rich in fibre, prebiotics and anti-inflammatory agents, as well as probiotics with proven effects on microbiota diversity and anti-inflammatory functions, should be evaluated for their ability to modify microbiome composition and functional potential and, in longitudinal studies, for any association with adenoma recurrence or progression. Such studies would be important to assess both changes in microbiome composition and function, as well as the impact on biomarkers associated with colorectal adenomas. The data obtained would help identify the most effective dietary and microbiome modulation strategies, especially in high-risk populations. Such an approach could become an important part of colorectal cancer prevention.

4 Organise pilot projects within national colorectal cancer screening programs that would include the coupling of gut microbiome data with traditional examinations, such as faecal

occult blood tests and colonoscopy. The aim of such pilot projects would be to assess whether microbiome-specific composition or function indicators can improve the sensitivity and specificity of existing screening methods, especially in the detection of early lesions. The projects would allow the development of more personalised screening strategies in which high-risk individuals would be identified more accurately and earlier. The results of such pilot projects could form the basis for the inclusion of microbiome biomarkers in national screening algorithms in the future.

5 Collaborate with nutritionists and family physicians to develop evidence-based dietary recommendations that promote the development of a beneficial and diverse gut microbiota. Such recommendations would be particularly important for patients with an adverse family history, previously diagnosed adenomas, or other risk factors for colorectal cancer. Dietary interventions should be prospectively evaluated as a non-invasive strategy for modifying microbiome composition and metabolic function in individuals at increased colorectal neoplasia risk. Involving family physicians would ensure early prevention at the primary care level, while nutritionists could provide personalised recommendations based on individual risk factors and microbiome profile. Such a multidisciplinary approach could become an important step in improving the primary prevention of colorectal cancer.

6 To better understand the role of the gut microbiota in the development of colorectal adenomatous lesions and the possible reversibility after ablation of the formation, it would be advisable to perform a comparative microbiome analysis before and after adenoma ablation. Such an approach would allow us to assess whether and how the composition and functionality of the gut microbiota change after tumour removal and to determine whether signs of dysbiosis persist after ablation, which could indicate a persistent unfavourable microbiological environment. Such an approach would allow us to identify potential opportunities for microbiome restoration and consider additional interventions (e. g., probiotic or prebiotic therapies). Understanding the potential of the microbiota as a dynamic biomarker could allow us to assess the effectiveness of therapy or assess the risks of relapse.

List of publications, reports and patents on the topic of the doctoral thesis

Publications:

1. **Vilkoite, I**., Mezmale, L., Tolmanis, I., Abu Meri, H., Veide, L., Dzerve, Z., Anarkulova, L., Nevidovska, K., Lejnieks, A. 2022. Comparison of NICE classification for optical diagnosis of colorectal polyps and morphology of removed lesions depending on localisation in colon. *Proceedings of the Latvian Academy of Sciences. Section B*, 76(6), 279–285.
2. **Vilkoite, I**., Tolmanis, I., Abu Meri, H., Polaka, I., Mezmale, L., Anarkulova, L., Leja, M., Lejnieks, A. 2023. The role of an artificial intelligence method of improving the diagnosis of neoplasms by colonoscopy. *Diagnostics (Basel)*, 13(4), 701. https://doi.org/10.3390/diagnostics13040701
3. Poļaka, I., Mežmale, L., Anarkulova, L., Kononova, E., **Vilkoite, I**., Veliks, V., Ļeščinska, A.M., Stonāns, I., Pčolkins, A., Tolmanis, I., Shani, G., Haick, H., Mitrovics, J., Glöckler, J., Mizaikoff, B., Leja, M. 2023. The detection of colorectal cancer through machine learning-based breath sensor analysis. *Diagnostics (Basel)*, 13(21), 3355. https://doi.org/10.3390/diagnostics13213355
4. **Vilkoite, I**., Tolmanis, I., Abu Meri, H., Polaka, I., Mezmale, L., Lejnieks, A. 2023. Age-based comparative analysis of colorectal cancer colonoscopy screening findings. *Medicina (Kaunas)*, 59(11), 2017. https://doi.org/10.3390/medicina59112017
5. Kononova, E., Mežmale, L., Poļaka, I., Veliks, V., Anarkulova, L., **Vilkoite, I**., Tolmanis, I., Ļeščinska, A. M., Stonāns, I., Pčolkins, A., Mochalski, P., Leja, M. 2024. Breath fingerprint of colorectal cancer patients based on the gas chromatography-mass spectrometry analysis. *International Journal of Molecular Sciences*, 25(3), 1632. https://doi.org/10.3390/ijms25031632
6. **Vilkoite I**, Silamiķelis I, Kloviņš J, et al. Colorectal adenoma presence is associated with decreased menaquinone pathway functions in the gut microbiome of patients undergoing routine colonoscopy. *PLoS One*. 2026;21(3):e0344050. Published 2026 Mar 11. doi:10.1371/journal. pone.0344050

References

1. Ahn, J., Sinha, R., Pei, Z., Dominianni, C., Wu, J., Shi, J., et al. (2013). Human gut microbiome and risk for colorectal cancer. *J Natl Cancer Inst.*;105(24):1907–1911. doi:10.1093/jnci/djt300.

2. Ai, D., Pan, H., Li, X., Gao, Y., Liu, G., Xia, L. C. (2019). Identifying gut microbiota associated with colorectal cancer using a zero-inflated lognormal model. *Front Microbiol.*;10:826. doi:10.3389/fmicb.2019.00826. PMID:31068913; PMCID:PMC6491826.

3. Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society: Series B (Methodological).*;44(2):139–177.

4. Allais, L., Kerckhof, F. M., Verschuere, S., Bracke, K. R., De Smet, R., Laukens, D., et al. (2016). Chronic cigarette smoke exposure induces microbial and inflammatory shifts and mucin changes in the murine gut. *Environ Microbiol.*;18(5):1352–1363. doi:10.1111/1462-2920.12934.

5. Almeida, A., Mitchell, A. L., Boland, M., et al. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology.*;39:105–114.

6. Almeida, A., Nayfach, S., Boland, M., et al. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology.*;39:105–114. doi:10.1038/s41587-020-0603-3.

7. Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data [Internet]. Cambridge, UK: Babraham Bioinformatics, Babraham Institute. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

8. Bagnardi, V., Rota, M., Botteri, E., et al. (2015). Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *Br J Cancer.*;112(3):580–593. doi:10.1038/bjc.2014.579.

9. Beghini, F., McIver, L. J., Blanco-Míguez, A., et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife.*;10:e65088. doi:10.7554/eLife.65088.

10. Botteri, E., Iodice, S., Bagnardi, V., Raimondi, S., Lowenfels, A. B., Maisonneuve, P. (2008). Smoking and colorectal cancer: a meta-analysis. *JAMA.*;300(23):2765–2778. doi:10.1001/jama.2008.839. PMID:19088354.

11. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.*;74(3):229–263. doi:10.3322/caac.21834.

12. Brennan, C. A., Garrett, W. S. (2016). Gut microbiota, inflammation, and colorectal cancer. *Annu Rev Microbiol.*;70:395–411. doi:10.1146/annurev-micro-102215-095513.

13. Brooijmans, R., Smit, B., Santos, F., van Riel, J., de Vos, W. M., Hugenholtz, J. (2009). Heme and menaquinone induced electron transport in lactic acid bacteria. *Microb Cell Fact.*;8:28. doi:10.1186/1475-2859-8-28.

14. Buckel, W., Barker, H. A. (1974). Two pathways of glutamate fermentation by anaerobic bacteria. *J Bacteriol.*;117(3):1248–1260. doi:10.1128/jb.117.3.1248-1260.1974.

15. Canani, R. B., Costanzo, M. D., Leone, L., et al. (2011). Potential beneficial effects of butyrate in intestinal and extraintestinal diseases. *World J Gastroenterol.*;17(12):1519–1528. doi:10.3748/wjg.v17.i12.1519.

16. Cao, Y., Shen, J., Ran, Z. H. (2014). Association between Faecalibacterium prausnitzii reduction and inflammatory bowel disease: a meta-analysis and systematic review of the literature. *Gastroenterol Res Pract.*;2014:872725. doi:10.1155/2014/872725. PMID:24799893; PMCID:PMC3985188.

17. Charbonneau, M., Blanton, L., DiGiulio, D., Relman, D. A., Lebrilla, C., Mills, D. A., et al. (2016). A microbial perspective of human developmental biology. *Nature.*;535(7610):48–55. doi:10.1038/nature18845.

18. Chen, G., Ren, Q., Zhong, Z., Li, Q., Huang, Z., Zhang, C., et al. (2024). Exploring the gut microbiome's role in colorectal cancer: diagnostic and prognostic implications. *Front Immunol.*;15:1431747. doi:10.3389/fimmu.2024.1431747. PMID:39483461; PMCID: PMC11524876.

19. Chen, S., Zhou, Y., Chen, Y., Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.*;34(17):i884–i890. doi:10.1093/bioinformatics/bty560.

20. Cheng, W. T., Kantilal, H. K., Davamani, F. (2020). The mechanism of Bacteroides fragilis toxin contributes to colon cancer formation. *Malays J Med Sci.*;27(4):9–21. doi:10.21315/mjms2020.27.4.2. PMID:32863742; PMCID:PMC7444842.

21. Cisek, A. A., Szymańska, E., Aleksandrzak-Piekarczyk, T., Cukrowska, B. (2024). The role of methanogenic archaea in inflammatory bowel disease – A review. *J Pers Med.*;14(2):196. doi:10.3390/jpm14020196.

22. Click, B., Pinsky, P. F., Hickey, T., Doroudi, M., Schoen, R. E. (2018). Association of colonoscopy adenoma findings with long-term colorectal cancer incidence. *JAMA.*;319(19):2021–2031. doi:10.1001/jama.2018.5809. PMID:29800214; PMCID:PMC6583246.

23. Dao, M. C., Everard, A., Aron-Wisnewsky, J., Sokolovska, N., Prifti, E., Verger, E. O., et al. (2016). Akkermansia muciniphila and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. *Gut.*;65(3):426–436. doi:10.1136/gutjnl-2014-308778.

24. Deng, D., Zhao, L., Song, H., Wang, H., Cao, H., Cui, H., Zhou, Y., Cui, R. (2025). Microbiome analysis of gut microbiota in patients with colorectal polyps and healthy individuals. *Sci Rep.*;15:7126. doi:10.1038/s41598-025-14680-6.

25. Duvallet, C., Gibbons, S. M., Gurry, T., Irizarry, R. A., Alm, E. J. (2017). Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun.*;8:1784. doi:10.1038/s41467-017-01973-8.

26. Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., et al. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nat Methods.*;15(11):962–968. doi:10.1038/s41592-018-0176-y.

27. Franzosa, E. A., Morgan, X. C., Segata, N., Waldron, L., Reyes, J., Earl, A. M., et al. (2014). Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci USA.*;111(22):E2329–E2338. doi:10.1073/pnas.1319284111.

28. Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nature Microbiology.*;4:293–305. doi:10.1038/s41564-018-0306-4.

29. Franzosa, E. A., Sirota-Madi, A., Avila-Pacheco, J., Fornelos, N., Haiser, H. J., Reinker, S., et al. (2019). Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat Microbiol.*;4(2):293–305. doi:10.1038/s41564-018-0306-4.

30. Fu, J., Li, G., Li, X., Song, S., Cheng, L., Rui, B., et al. (2024). Gut commensal Alistipes as a potential pathogenic factor in colorectal cancer. *Discov Oncol.*;15(1):473. doi:10.1007/s12672-024-01393-3. PMID:39331213; PMCID:PMC11436608.

31. Gagnière, J., Raisch, J., Veziant, J., Barnich, N., Bonnet, R., Buc, E., et al. (2016). Gut microbiota imbalance and colorectal cancer. *World J Gastroenterol.*;22(2):501–518. doi:10.3748/wjg.v22.i2.501.

32. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., Egozcue, J. J. (2017). Microbiome datasets are compositional: and this is not optional. *Front Microbiol.*;8:2224. doi:10.3389/fmicb.2017.02224.

33. Gomez-Arango, L. F., Barrett, H. L., Wilkinson, S. A., Callaway, L. K., McIntyre, H. D., Morrison, M., et al. (2018). Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. *Gut Microbes.*;9(3):189–201. doi:10.1080/19490976.2017.1406584.

34. Guo, M., Miao, M., Wang, Y., Du, Z., Jin, Z., Jiang, J., et al. (2020). Developmental differences in the intestinal microbiota of Chinese 1-year-old infants and 4-year-old children. *Sci Rep.*;10:19470. doi:10.1038/s41598-020-76591-4.

35. Hale, V. L., Chen, J., Johnson, S., Harrington, S. C., Yab, T. C., Smyrk, T. C., et al. (2016). Shifts in the fecal microbiota associated with adenomatous polyps. *Cancer Epidemiol Biomarkers Prev.*;26(1):85–94. doi:10.1158/1055-9965.EPI-16-0337.

36. Hewett, D. G., Kaltenbach, T., Sano, Y., Tanaka, S., Saunders, B. P., Ponchon, T., Soetikno, R., Rex, D. K. (2012). Validation of a simple classification system for endoscopic diagnosis of small colorectal polyps using narrow-band imaging. *Gastroenterology.*;143(3):599–607.e1. doi:10.1053/j.gastro.2012.05.006. PMID:22609383.

37. Holmes, E., Li, J. V., Athanasiou, T., Ashrafian, H., Nicholson, J. K. (2011). Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol.*;19(7):349–359. doi:10.1016/j.tim.2011.05.006.

38. Hu, M., Caldarelli, G., Gili, T. (2023). Inflammatory bowel disease biomarkers revealed by the human gut microbiome network. *Sci Rep.*;13:19428. doi:10.1038/s41598-023-46184-y.

39. Kang, Y., Pan, W., Cai, Y. (2017). Gut microbiota and colorectal cancer: insights into pathogenesis for novel therapeutic strategies. *Z Gastroenterol.*;55(9):872–880. doi:10.1055/s-0043-116387.

40. Kim, N. H., Jung, Y. S., Lee, M. Y., Park, J. H., Park, D. I., Sohn, C. I. (2019). Risk of developing metachronous advanced colorectal neoplasia after polypectomy in patients with multiple diminutive or small adenomas. *Am J Gastroenterol.*;114(10):1657–1664. doi:10.14309/ajg.0000000000000296. PMID:31211708.

41. Koliarakis, I., Psaroulaki, A., Nikolouzakis, T. K., Apostolopoulos, K., Alegakis, A. K., Sgantzos, M., et al. (2018). Intestinal microbiota and colorectal cancer: a new aspect of research. *J BUON.*;23(5):1216–1234.

42. Kostic, A. D., Chun, E., Robertson, L., Glickman, J. N., Gallini, C. A., Michaud, M., et al. (2013). Fusobacterium nucleatum potentiates intestinal tumourigenesis and modulates the tumour-immune microenvironment. *Cell Host Microbe.*;14(2):207–215. doi:10.1016/j.chom.2013.07.007.

43. Langmead, B., Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods.*;9(4):357–359. doi:10.1038/nmeth.1923.

44. Larsen, J. M. (2017). The immune response to Prevotella bacteria in chronic inflammatory disease. *Immunology.*;151(4):363–374. doi:10.1111/imm.12760. PMID:28542929; PMCID:PMC5506432.

45. Leslie, A., Carey, F. A., Pratt, N. R., Steele, R. J. (2002). The colorectal adenoma–carcinoma sequence. *Br J Surg.*;89(7):845–860. doi:10.1046/j.1365-2168.2002.02120.x. PMID:12081733.

46. Levin, B., Lieberman, D. A., McFarland, B., Andrews, K. S., Brooks, D., Bond, J., Dash, C., Giardiello, F. M., Glick, S., Johnson, D., Johnson, C. D., Levin, T. R., Pickhardt, P. J., Rex, D. K., Smith, R. A., Thorson, A., Winawer, S. J. (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology.*;134(5):1570–1595. doi:10.1053/j.gastro.2008.02.002.

47. Lieberman, D. A., Weiss, D. G., Harford, W. V., Ahnen, D. J., Provenzale, D., Sontag, S. J., et al. (2007). Five-year colon surveillance after screening colonoscopy. *Gastroenterology.*;133(4):1077–1085. doi:10.1053/j.gastro.2007.07.006. PMID:17698067.

48. Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association.*;62(318):399–402. doi:10.2307/2283970.

49. Lindgreen, S., Adair, K. L., Gardner, P. P. (2016). An evaluation of the accuracy and speed of metagenome analysis tools. *Sci Rep.*;6:19233. doi:10.1038/srep19233.

50. Lindström, M. (2017). Commentary on Wang et al. (2017): differing patterns of short-term transitions of nondaily smokers for different indicators of socioeconomic status (SES). *Addiction.*;112(5):873–874. doi:10.1111/add.13758.

51. Liu, Y., Chen, H., Van Treuren, W., Hou, B. H., Higginbottom, S. K., Dodd, D. (2022). Clostridium sporogenes uses reductive Stickland metabolism in the gut to generate ATP and produce circulating metabolites. *Nat Microbiol.*;7(5):695–706. doi:10.1038/s41564-022-01109-9.

52. Lloyd-Price, J., Abu-Ali, G., Huttenhower, C. (2016). The healthy human microbiome. *Genome Med.*;8(1):51. doi:10.1186/s13073-016-0307-y.

53. Lo, C. H., Wu, D. C., Jao, S. W., et al. (2022). Enrichment of Prevotella intermedia in human colorectal cancer and its additive effects with Fusobacterium nucleatum on the malignant transformation of colorectal adenomas. *J Biomed Sci.*;29:88. doi:10.1186/s12929-022-00869-0.

54. Louis, P., Flint, H. J. (2017). Formation of propionate and butyrate by the human colonic microbiota. *Environ Microbiol.*;19(1):29–41. doi:10.1111/1462-2920.13589. PMID:27928878.

55. Louis, P., Hold, G. L., Flint, H. J. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. *Nat Rev Microbiol.*;12(10):661–672. doi:10.1038/nrmicro3344.

56. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K., Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature.*;489(7415):220–230. doi:10.1038/nature11550.

57. Lu, J., Breitwieser, F. P., Thielen, P., Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ Computer Science.*;3:e104. doi:10.7717/peerj-cs.104.

58. Lv, M., Zhang, J., Deng, J., Wang, Y., Li, Z., Li, L., et al. (2023). Analysis of the relationship between the gut microbiota enterotypes and colorectal adenoma. *Front Microbiol.*;14:1097892. doi:10.3389/fmicb.2023.1097892.

59. Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., et al. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol.*;17(11):e1009442. doi:10.1371/journal.pcbi.1009442.

60. Martín, R., Rios-Covian, D., Huillet, E., Auger, S., Khazaal, S., Bermúdez-Humarán, L. G., et al. (2023). Faecalibacterium: a bacterial genus with promising human health applications. *FEMS Microbiol Rev.*;47(4):fuad039. doi:10.1093/femsre/fuad039. PMID:37451743; PMCID: PMC10410495.

61. MGI Tech Co., Ltd. (n.d.). DNBSEQ-G400RS High-throughput Sequencing Platform. Available from: https://en.mgi-tech.com.

62. MP Biomedicals. (n.d.). FastDNA™ Spin Kit for Soil. Available from: https://www.mpbio.com.

63. Nearing, J. T., Douglas, G. M., Comeau, A. M., Langille, M. G. I. (2022). Microbiome differential abundance methods produce different results across 38 datasets. *Nat Commun.*;13(1):342. doi:10.1038/s41467-022-28034-z.

64. O'Keefe, S. J., Li, J. V., Lahti, L., Ou, J., Carbonero, F., Mohammed, K., et al. (2015). Fat, fibre and cancer risk in African Americans and rural Africans. *Nat Commun.*;6:6342. doi:10.1038/ncomms7342. PMID:25919227; PMCID:PMC4415091.

65. Odze, R. D., Klimstra, D., Paradis, V., Rugge, M., Schirmacher, P., Washington, K. M., Carneiro, F., Cree, I. A.; WHO Classification of Tumours Editorial Board. (2020). The 2019 WHO classification of tumours of the digestive system. *Histopathology.*;76(2):182–188. doi:10.1111/his.13975. PMID:31433515; PMCID:PMC7003895.

66. Oksanen, J., Blanchet, F. G., Friendly, M., et al. (2022). vegan: Community Ecology Package. R package version 2.6-2. Available from: https://CRAN.R-project.org/package=vegan.

67. Omata, F., Deshpande, G. A., Ohde, S., Mine, T., Fukui, T. (2013). The association between obesity and colorectal adenoma: systematic review and meta-analysis. *Scand J Gastroenterol.*;48(2):136–146. doi:10.3109/00365521.2012.737364. PMID:23130996.

68. Participants in the Paris Workshop. (2003). The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon. *Gastrointest Endosc.*;58(6 Suppl):S3–S43. doi:10.1016/S0016-5107(03)02159-X.

69. Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F., et al. (2019). Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell.*;176(3):649–662.e20. doi:10.1016/j.cell.2019.01.001.

70. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research.*;12:2825–2830.

71. R Core Team. (2023). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

72. Ritchie, M. E., Phipson, B., Wu, D., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research.*;43(7):e47. doi:10.1093/nar/gkv007.

73. Scikit-bio development team. (2019). scikit-bio: A Bioinformatics Library for Data Scientists, Students, and Developers. Available from: https://scikit-bio.org.

74. Smajdor, J., Jedlińska, K., Porada, R., et al. (2023). The impact of gut bacteria producing long chain homologs of vitamin K_2 on colorectal carcinogenesis. *Cancer Cell Int.*;23(1):268. doi:10.1186/s12935-023-03114-2.

75. Sommer, M. O. A., Dantas, G., Church, G. M. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science.*;325(5944):1128–1131. doi:10.1126/science.1176950.

76. Sonnenburg, E. D., Sonnenburg, J. L. (2019). The ancestral and industrialized gut microbiota and implications for human health. *Nat Rev Microbiol.*;17(6):383–390. doi:10.1038/s41579-019-0191-8. PMID:31089293.

77. Sze, M. A., Schloss, P. D. (2016). Looking for a signal in the noise: revisiting obesity and the microbiome. *mBio.*;7(4):e01018-16. doi:10.1128/mBio.01018-16.

78. Thermo Fisher Scientific. (n.d.). Qubit dsDNA HS Assay Kits. Available from: https://www.thermofisher.com.

79. Thomas, A. M., Manghi, P., Asnicar, F., Pasolli, E., Armanini, F., Zolfo, M., et al. (2019). Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med.*;25(4):667–678. doi:10.1038/s41591-019-0405-7.

80. Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., et al. (2018). US immigration westernizes the human gut microbiome. *Cell.*;175(4):962–972.e10. doi:10.1016/j.cell.2018.10.029. PMID:30388453; PMCID:PMC6498444.

81. Vilkoite, I., Mezmale, L., Tolmanis, I., Lejnieks, A., et al. (2022). Comparison of NICE classification for optical diagnosis of colorectal polyps and morphology of removed lesions depending on localisation in colon. *Proc Latv Acad Sci Sect B Nat Exact Appl Sci.*;76(5–6):650–656. doi:10.2478/prolas-2022-0100.

82. Wang, Y., Li, H. (2022). Gut microbiota modulation: a tool for the management of colorectal cancer. *J Transl Med.*;20(1):178. doi:10.1186/s12967-022-03378-8.

83. Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome.*;5(1):27. doi:10.1186/s40168-017-0237-y.

84. Wolin, K. Y., Yan, Y., Colditz, G. A., Lee, I. M. (2009). Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer.*;100(4):611–616. doi:10.1038/sj.bjc.6604917. PMID:19209175; PMCID:PMC2653744.

85. Wood, D. E., Lu, J., Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biology.*;20(1):257. doi:10.1186/s13059-019-1891-0.

86. Wu, Z., Huang, Y., Zhang, R., et al. (2024). Sex differences in colorectal cancer: with a focus on sex hormone–gut microbiome axis. *Cell Commun Signal.*;22:167. doi:10.1186/s12964-024-01549-2.

87. Zeller, G., Tap, J., Voigt, A. Y., Sunagawa, S., Kultima, J. R., Costea, P. I., et al. (2014). Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol.*;10(11):766. doi:10.15252/msb.20145645. PMID:25432777; PMCID:PMC4299606.

88. Zhang, H., Wu, J., Ji, D., Liu, Y., Lu, S., Lin, Z., et al. (2022). Microbiome analysis reveals universal diagnostic biomarkers for colorectal cancer across populations and technologies. *Front Microbiol.*;13:1005201. doi:10.3389/fmicb.2022.1005201. PMID:36406447; PMCID: PMC9668862.

89. Zhou, H., Shen, Z., Zhao, J., Zhou, Z., Xu, Y. (2018). Distribution characteristics and risk factors of colorectal adenomas. *Zhonghua Wei Chang Wai Ke Za Zhi.*;21(6):678–684. Chinese. PMID:29968244.

90. Zymo Research. (n.d.). DNA/RNA Shield Fecal Collection Tube. Available from: https://zymoresearch.eu/products/dna-rna-shield-fecal-collection-tube.

Acknowledgements

The greatest gratitude to my supervisors – Professor Aivars Lejnieks, who, unfortunately, is no longer with us, but whose legacy will always remain in the hearts and minds of his students and colleagues. Without his encouragement, support and professional guidance, my work would not be what it is today. Thank you to the professor for showing me the right direction and encouraging me to think more broadly, for his support, patience and encouragement to start scientific work in the hitherto unknown field of the microbiome. Thanks also to Dr. Ivars Tolmanis for his presence, faith and guidance in my growth.

Special gratitude to my scientific consultant Laila Silamiķele from the Latvian Biomedical Research and Study Center – for her dedicated time, advice, invaluable assistance during laboratory examinations and support in complex statistical challenges, as well as to my colleagues Ivars Silamiķelis, Professor Jānis Kloviņš and other specialists of the Biomedical Center for their responsiveness, consultations and valuable suggestions.

I would like to express my heartfelt gratitude to Māris Rēvalds for the opportunity to carry out this extensive work using the infrastructure of the health centre "Veselības centrs 4". I would like to thank my colleagues at the "Diagnostics Center" branch of SIA "Veselības Centrs 4" and SIA "Akadēmiskā Histology Laboratory" for their significant contribution to the examination and sample processing.

I highly appreciate the inspiration and encouragement of Professor Sandra Lejniece and Professor Ilze Konrāde not to be afraid to start working in a new, then unknown field – microbiome research. Their confidence allowed me to take bold steps and believe in success.

I am also immensely grateful to my family – my husband and children – for their understanding, patience and support during the times when I had to put work on publications and the dissertation in the first place. Thank you to all my colleagues and friends who were there for me during the process of writing this work.