UNIVERSITÀ DEL PIEMONTE ORIENTALE
University of Eastern Piedmont
Department of Health Sciences

PhD Programme in Food Health and Longevity
XXXVI Cycle

Role of intestinal microbiota, obesity and inherited predisposition in colorectal carcinogenesis

Coordinator: **Prof. Antonia Follenzi** PhD candidate: **Marta La Vecchia**
Supervisor: **Prof. Irma Dianzani** SSD: MED/04

TABLE OF CONTENTS

SUMMARY

Colorectal cancer (CRC) is one of the most frequent malignant tumors worldwide and has a multifactorial etiology, influenced by unmodifiable risk factors – such as age, sex, genetic predisposition – and modifiable risk factors – such as smoking, heavy consumption of alcohol, overweight and obesity, physical inactivity, high consumption of red and processed meat, low consumption of dietary fibers and whole grains. Increasing data indicate that changes in the gut microbiota are implicated in initiation and promotion of CRC.

To evaluate the role of various CRC risk factors, we recruited a panel of 306 patients with colon polyps at the Ospedale Maggiore della Carità (Novara) and the Azienda Ospedaliera Universitaria Città della Salute e della Scienza (Torino). We collected biological samples (swabs for microbiota and metabolome analyses, feces, peripheral blood) and clinical information (anthropometric parameters, dietary and lifestyle habits, family and personal history of cancer).

Our objective was to shed light on the role of the intestinal microbiota on colon carcinogenesis and its interaction with obesity and inherited predisposition syndromes. Thus, we stratified patients according to the grade of dysplasia of their polyps, to their body mass index and waist circumference, or to their genetic background. Both mucosa- and lumen-associated microbiota (MAM and LAM, respectively) were analyzed. However, since bacteria in close contact with enterocytes may play an important role in colon carcinogenesis, we focused our attention on MAM and mucosa-associated metabolites. For this purpose, we have devised a new technique that allows the collection of MAM and metabolites from the polyp's surface without jeopardizing tissue integrity.

Firstly, we explored the relationship among MAM, mucosa-associated metabolome and CRC initiation and progression in 78 individuals with colorectal polyps. According to the driver-passenger model, candidate driver bacteria are found on low-grade polyps and are involved in CRC initiation, while candidate passenger bacteria are found on high-grade polyps and are involved in CRC progression. Accordingly, we identified an enrichment of candidate passenger genera on high-grade dysplastic adenomas, whereas *Bacteroides fragilis*, one of the most studied driver bacteria, was enriched on low-grade dysplastic polyps. We also found that different metabolite signatures were associated to tumor stage. A microbiota-metabolome integrated analysis showed positive or negative correlations between bacteria and metabolites. These results support the involvement of mucosa-associated microbiota and metabolome in CRC initiation and progression.

Secondly, we explored the role of body weight, diet, lifestyle and intestinal microbiota cofactors in colorectal carcinogenesis, analyzing 120 patients with colon polyps, divided in normal-weight or overweight/obese. Dietary habits data, obtained using a validated EPIC questionnaire, revealed a statistically significant higher consumption of processed meat for obese patients compared to normal-

weight ones. Both MAM and LAM 16S sequencing analyses revealed different bacterial signatures in the two groups. In particular, we found an enrichment of *Finegoldia magna*, that has a high pathogenic potential due to its virulence factors, in MAM of obese patients compared to normal-weight ones. Moreover, mucosal and luminal metabolome signatures distinguished normal-weight and obese patients. Obese patients showed a higher concentration of pyroglutamic acid and a lower concentration of niacin in mucosal-associated metabolome. These data support the hypothesis that different risk factors cooperate in tumorigenesis of obese and normal-weight patients (e.g. processed meat and *F. magna* in obese patients).

Thirdly, we explored the genetic predisposition and gut microbiota relationship. We first identified 29 germline pathogenic variants (PVs) in cancer predisposing genes in 26 of 242 (10.74%) individuals with colon polyps and classified patients as mutated or sporadic. Then, we performed MAM shotgun sequencing analysis to compare 154 patients with or without inherited predisposition to cancer (i.e. mutated *vs* sporadic). In particular, we found that *Fusobacterium nucleatum,* one of the most studied CRC-associated bacteria, was enriched in mutated patients compared to sporadic ones, supporting the hypothesis that colorectal carcinogenesis in patients with different germline genetic background is favored by different bacterial risk factors. We speculate that bacteria, such as *F. nucleatum,* cooperate with the genetic defect to promote cancer.

Overall, our data show that different mechanisms and risk factors are involved in colorectal carcinogenesis in specific groups of patients. Further studies will be necessary to test the short-term or long-term effects of specific bacteria or molecules.

RIASSUNTO

Il cancro del colon-retto (CRC) è uno dei più frequenti tumori maligni in tutto il mondo e ha eziologia multifattoriale, influenzata da fattori di rischio non modificabili (come età, sesso, predisposizione genetica) e fattori di rischio modificabili (come fumo, alto consumo di alcol, sovrappeso e obesità, inattività fisica, alto consumo di carne rossa e processata, basso consumo di fibre e cereali integrali). Sempre più dati indicano che alterazioni nel microbiota intestinale sono implicati nell'iniziazione e nella promozione del CRC.

Per valutare il ruolo di vari fattori di rischio per il CRC, abbiamo reclutato 306 pazienti con polipi intestinali presso l'Ospedale Maggiore della Carità (Novara) e l'Azienda Ospedaliera Universitaria Città della Salute e della Scienza (Torino). Abbiamo raccolto i loro campioni biologici (tamponi per le analisi sul microbiota e sul metaboloma, feci, sangue periferico) e informazioni (parametri antropometrici, abitudini riguardanti la dieta e lo stile di vita, storia familiare e personale di tumori).

Il nostro obiettivo è stato far luce sul ruolo del microbiota intestinale nella cancerogenesi del colon e sulla sua interazione con obesità e sindromi ereditarie ad alta incidenza di cancro. Abbiamo quindi stratificato i pazienti in base al grado di displasia dei loro polipi, al loro indice di massa corporea e circonferenza vita, o al loro background genetico. Sono stati analizzati sia il microbiota associato alla mucosa sia il microbiota luminale (rispettivamente MAM e LAM). Tuttavia, poiché i batteri a stretto contatto con gli enterociti potrebbero avere un ruolo importante nella cancerogenesi del colon, ci siamo focalizzati sul MAM e sui metaboliti associati alla mucosa. A questo scopo, abbiamo sviluppato una nuova tecnica che permette la raccolta di MAM e metaboliti dalla superficie del polipo senza compromettere l'integrità del tessuto.

In primo luogo, abbiamo esplorato la relazione tra MAM, metaboliti associati alla mucosa e iniziazione e progressione del CRC in 78 individui con polipi colorettali. Secondo il modello driver-passenger, sui polipi a basso grado di displasia si trovano batteri candidati come driver e probabilmente coinvolti nell'iniziazione del CRC, mentre sui polipi ad alto grado si trovano batteri candidati come passenger che potrebbero essere coinvolti nella progressione del CRC. In accordo con questa teoria, abbiamo identificato un arricchimento di possibili batteri passenger sugli adenomi ad alto grado di displasia, mentre *Bacteroides fragilis*, uno dei più studiati batteri driver, era arricchito sui polipi a basso grado. Abbiamo visto che anche i profili dei metaboliti sono significativamente diversi in rapporto allo stadio del tumore, e abbiamo svolto un'analisi integrata microbiota-metaboloma che ha mostrato correlazioni positive o negative tra i batteri e i metaboliti. Questi risultati supportano il coinvolgimento del microbiota e del metaboloma associati alla mucosa nell'iniziazione e nella progressione del CRC.

In secondo luogo, abbiamo esplorato il ruolo di peso corporeo, dieta, stile di vita e microbiota intestinale come cofattori della cancerogenesi del colon, analizzando 120 pazienti con polipi del colon divisi in normopeso e sovrappeso/obesi. I dati riguardanti le abitudini alimentari, ottenuti usando un questionario validato EPIC, hanno rivelato un consumo maggiore in modo statisticamente significativo di carne processata nei pazienti obesi rispetto ai normopeso. Entrambe le analisi tramite sequenziamento 16S di MAM e LAM hanno mostrato differenti firme batteriche nei due gruppi. In particolare, abbiamo identificato un arricchimento di *Finegoldia magna*, che ha un alto potenziale patogenico dovuto ai suoi fattori di virulenza, nel MAM dei pazienti obesi comparati ai normopeso. Inoltre, le firme del metaboloma mucosale e luminale discriminano tra i pazienti normopeso e obesi. I pazienti obesi mostrano una maggior concentrazione di acido piroglutammico e una minor concentrazione di niacina nel loro metaboloma associato al tumore. Questi dati supportano l'ipotesi che fattori di rischio differenti cooperino nella tumorigenesi di pazienti obesi e normopeso (ad esempio carne processata e *F. magna* nei pazienti obesi).

In terzo luogo, abbiamo esplorato la relazione tra predisposizione genetica e microbiota intestinale. Abbiamo quindi identificato 29 varianti germinali patogeniche (PV) in geni che predispongono al cancro in 26 su 242 individui con polipi del colon-retto (10.74%) e classificato i pazienti come mutati o sporadici. Abbiamo poi svolto il sequenziamento shotgun sul MAM per confrontare 154 pazienti con o senza predisposizione ereditaria (cioè, mutati *vs* sporadici). In particolare, abbiamo trovato che *Fusobacterium nucleatum,* uno dei più studiati batteri associati al CRC, è arricchito nei pazienti mutati comparati agli sporadici, supportando l'ipotesi che la cancerogenesi del colon-retto nei pazienti con differente background genetico sia favorita da differenti fattori di rischio. L'ipotesi di lavoro è che batteri come *F. nucleatum* cooperino con il difetto genetico per promuovere il cancro.

In conclusione, i nostri dati mostrano che differenti meccanismi e fattori di rischio sono coinvolti nella cancerogenesi colorettale in specifici gruppi di pazienti. Ulteriori studi saranno necessari per valutare gli effetti a breve o lungo termine di specifici batteri o molecole.

Chapter 1

Introduction

Colorectal cancer

Colorectal cancer (CRC) is one of the most frequent malignant tumors worldwide, that ranked third in incidence and second in mortality among all cancers in 2020[1]. CRC development is influenced by unmodifiable risk factors – such as age, sex, genetic predisposition[2]– and modifiable risk factors – such as smoking, heavy consumption of alcohol, overweight and obesity, physical inactivity, high consumption of red and processed meat, low consumption of dietary fiber and whole grains[3]– which confer to CRC a complex multifactorial etiology. Increasing data indicate that changes in the gut microbiota, i.e. the collection of microorganisms colonizing the gastrointestinal tract, are implicated in initiation and promotion of CRC[3].

1.1 Epidemiology

According to GLOBOCAN updated data (https://gco.iarc.fr/), CRC is the third cancer in incidence (Fig. 1.1a) and the second in mortality (Fig. 1.1b), with more than 1.9 million new cases and 935 000 deaths in 2020, that are predicted to reach 3.2 million and 1.6 million respectively by 2040[4].

a) b)

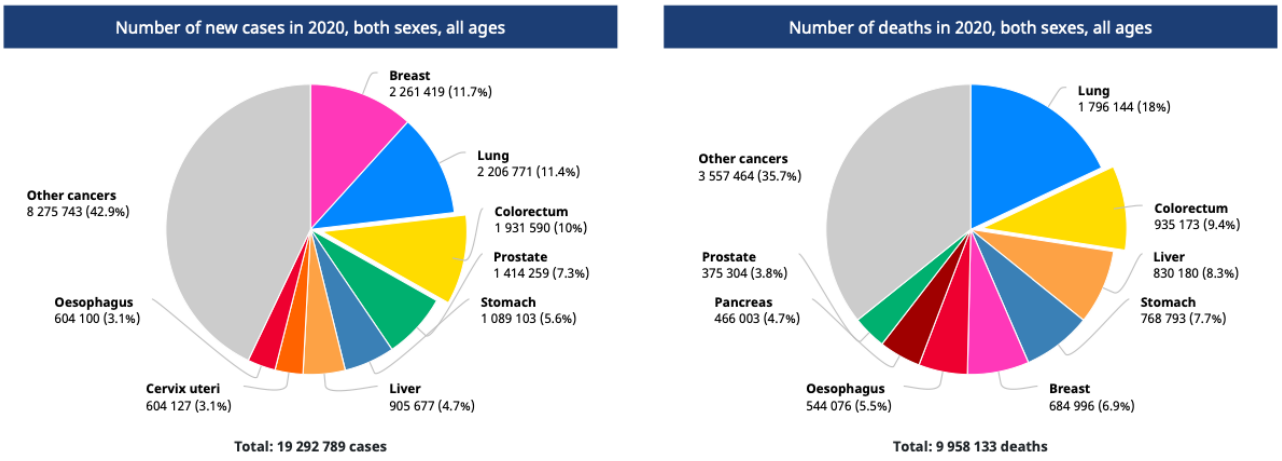


Figure 1.1. *Number of new cases of CRC (a) and deaths (b) in 2020 (GLOBOCAN 2020).*

The age standardized incidence rate at world level for 100.000 cases was equal to 23.4 in males and 16.2 in females (Fig. 1.2).

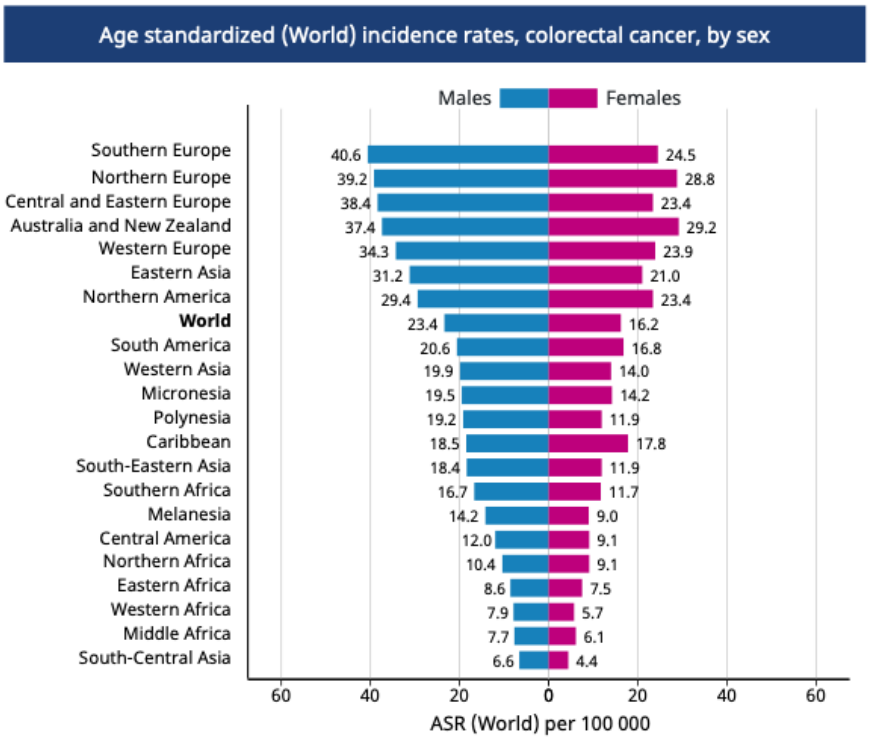


Figure 1.2. *Age standardized (World) incidence rates of CRC in 2020 (GLOBOCAN 2020).*

CRC incidence and mortality are rapidly growing worldwide, reflecting both aging of the population and changes in the prevalence and distribution of CRC risk factors associated with socioeconomic development[1].

Developed countries (Europe, Australia/New Zealand, Eastern Asia, Northern America) are at the highest risk for CRC compared to less developed ones (Africa and Southern Asia), but recent advances in early detection and treatment options, as well as lifestyle modification, help CRC mortality reduction in developed nations[5].

1.2 Pathobiology

Colorectal carcinogenesis occurs through four stages: initiation (genetic alterations that predispose the affected cells to subsequent neoplastic transformation), promotion (abnormal growth of the initiated cells), progression (further genetic and epigenetic alterations that confer a selective growth advantage to cells) and metastasis (spread of cancer cells to other organs, generally liver and lungs)[6]. Most of colorectal cancer (85-90%) originates through a multistep process, known as "adenoma-carcinoma sequence", that affects epithelial cells leading to their transformation first into adenomas and then into adenocarcinomas. Other colorectal carcinogenic pathways are the serrated pathway (10–15% of CRC), originating from serrated polyps, and the inflammatory pathway (less than 2%), suggested to be driven by chronic inflammation[6] (Fig. 1.3).

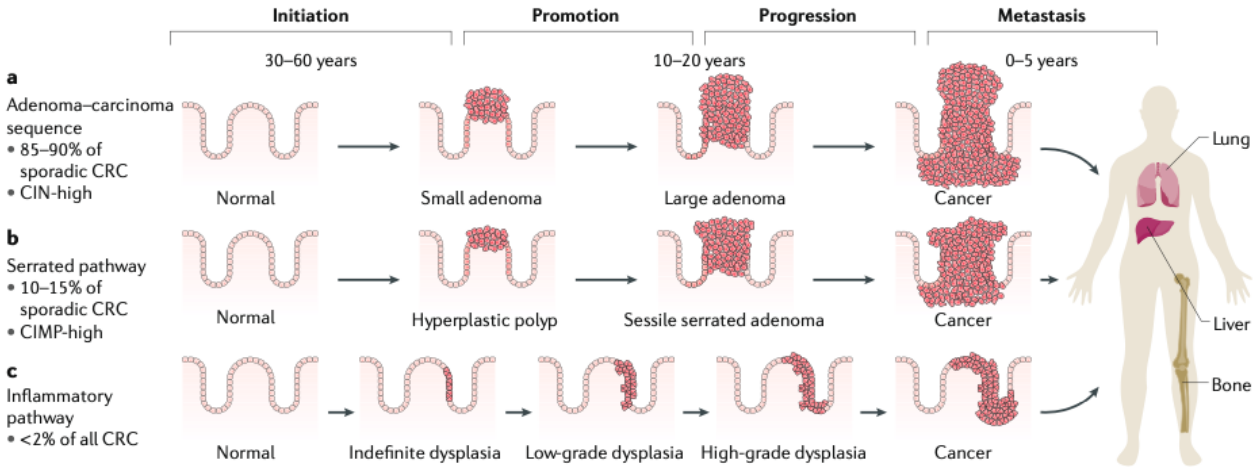


Figure 1.3. *Pathways of colorectal carcinogenesis (Keum et al., 2019). CIN: chromosomal instability; CIMP: CpG island methylator phenotype.*

Adenomatous polyps (adenomas) and serrated polyps are two major precursors of CRCs[7].

Adenomatous polyps are histologically divided into tubular, villous, and tubulovillous[8] and their risk to evolve in adenocarcinomas increases with the polyp size, with a higher risk for adenoma ≥ 1 cm [9].

Serrated polyps represent a group of heterogeneous lesions including hyperplastic polyp, traditional serrated adenoma, sessile serrated adenoma and mixed polyp. Only a subset of hyperplastic polyps can progress into serrated adenomas and CRC[6,10].

Based on their anatomical site, CRCs are categorized into proximal to the splenic flexure or right-sided (arising in cecum, ascending colon, hepatic flexure, or transverse colon), distal or left-sided (arising in descending or sigmoid colon) and rectal (arising within 15 cm from the rectum)[11,12]. The risk to develop proximal or distal tumors depends on age, sex, and ethnicity[13]. Proximal CRC is more frequent in females, older people, Africans and Afro-Americans, while distal colorectal cancer is more frequent in males, younger people, and Caucasians[6].

1.2.1 Genomic instability

The progressive accumulation of genetic and epigenetic aberrations (genomic instability) leading to CRC is caused by three different mechanisms: chromosomal instability (CIN), microsatellite instability (MSI), and CpG island methylator phenotype (CIMP)[6].

1.2.1.1 Chromosomal instability (CIN)

The chromosomal instability (CIN) pathway, characterized by the presence of structural and numerical chromosomal abnormalities and loss of heterozygosity (LOH), is the most common form of genomic instability, and is responsible for almost 80-85% of CRC cases[14].

The CIN pathway is associated with mutations in the tumor-suppressor gene *APC*, that lead to hyperactivation of Wnt/β-catenin pathway and tumor development. The APC protein is part of a multiprotein destruction complex, that also includes axis inhibitor (Axin), protein phosphatase 2A, glycogen synthase kinase-3 beta (GSK3β), and casein kinase 1 (CK1). In the absence of the WNT ligand, this complex is responsible for proteasomal degradation of the β-catenin transcription factor. Consequently, cells do not proliferate and start their maturation processes[15]. In the presence of a loss of function (LOF) mutation in *APC* or other components of the destruction complex, β-catenin accumulates, translocates into the nucleus, and constitutively activates cell proliferation[14].

Cells that have acquired one of such mutations may accumulate further mutations in other genes, including the oncogene *KRAS* (leading to a constant activation of MAP kinase and increasing cell proliferation) and the tumor-suppressor *TP53.* As a consequence, they progress to adenomas with increasing grade of dysplasia and ultimately to adenocarcinoma[14].

1.2.1.2 Microsatellite instability (MSI)

Microsatellite instability (MSI) is characterized by alterations in the length of microsatellites (i.e. short tandem repeats in DNA sequences) and is detected in about 15% of CRCs. MSI is caused by somatic inactivation of both alleles of a mismatch repair (MMR) gene – *MLH1, MSH2, MSH6* or *PMS2* – (sporadic CRCs) or by a germline mutation in one of the MMR genes (Lynch syndrome) followed by somatic inactivation of the wild-type allele[16]. Two distinct MSI tumor phenotypes are found: MSI-high (MSH-H) and MSI-low (MSI-L), depending on the proportion of microsatellite markers showing instability[16].

The presence of MSI has been associated with a better outcome in sporadic CRC compared to microsatellite stable tumors (MSS)[17].

1.2.1.3 CpG island methylator phenotype (CIMP)

CpG island methylator phenotype (CIMP) is caused by hypermethylation of CpG islands in tumor-suppressor genes promoters, leading to the silencing of those genes, which in turn increases cell proliferation and inhibits apoptosis[18,19].

CpG islands hypermethylation frequently arises from the serrated pathway of colorectal carcinogenesis[6].

1.3 Risk factors

1.3.1 Sex and age

CRC risk increases with age and is higher in males compared to females, who are however more prone to the high aggressive right-sided colon cancers[5]. The rate of CRC development and death increases after the age of 50 years.

The incidence rate for the disease in elderly people has decreased over the past decades, whereas the incidence rate for individuals under 50 years has increased, probably because of a more sedentary lifestyle and Western diet[5]. A portion of CRC cases in young patients (<50 years old) is caused by genetic predisposition[20].

1.3.2 Genetic predisposition

A positive familial history is a risk factor for CRC and familiarity is observed in about 30% of CRC cases[21,22]. The risk associated with family history is proportional to the number of affected family members, their age at diagnosis, and the degree of relationship[23].

However, only 2-8% of CRC cases are associated with germline pathogenic variants in high-risk cancer genes[24]. Hereditary CRC syndromes can be divided into non-polyposis and polyposis syndromes (Fig. 1.4). The first group includes Lynch syndrome (LS), caused by DNA mismatch repair (MMR) genes mutations, and familial colorectal cancer type X (FCCTX), describing MMR-proficient cases with a heterogeneous genetic background. The second group is distinguished in adenomatous or non-adenomatous (hamartomatous/serrated) polyp histology and comprises familial adenomatous polyposis (FAP), PTEN hamartoma tumor syndrome (PHTS), hamartomatous polyposis syndromes (Peutz-Jeghers syndrome, Juvenile polyposis), polymerase proof-reading associated polyposis (PPAP), MUTYH-associated polyposis (MAP), NTHL1-associated polyposis (NAP) and constitutional mismatch repair deficiency syndrome (CMMRD)[22].

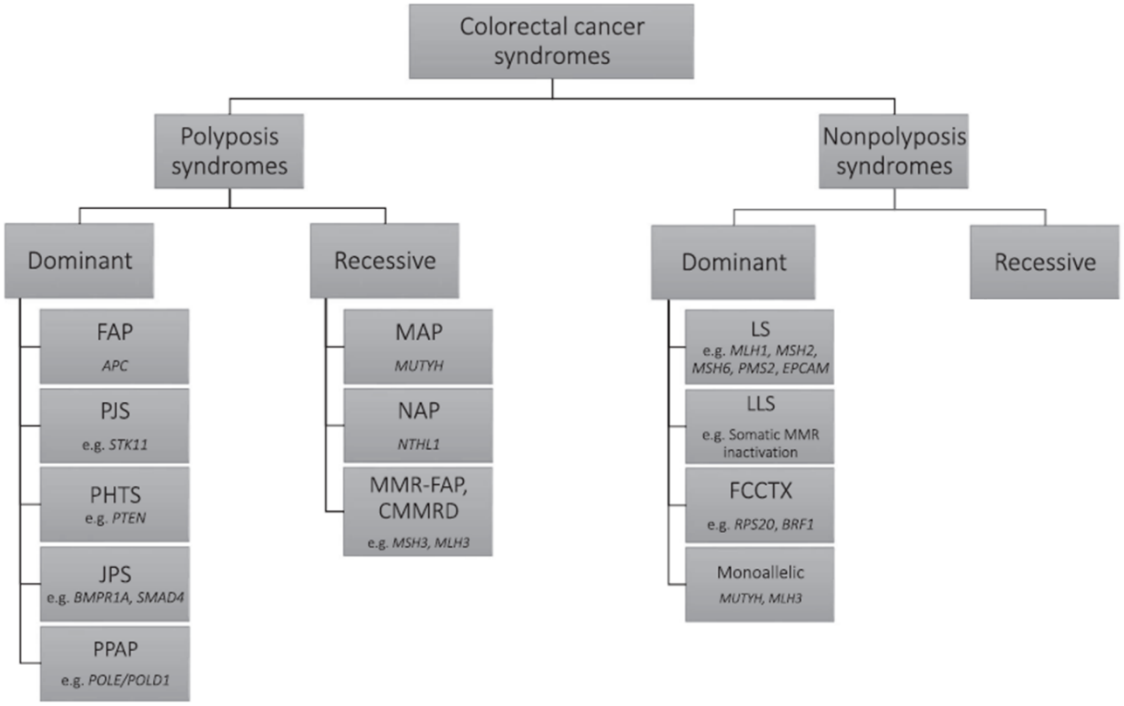


Figure 1.4. *Monogenic CRC syndromes classified in polyposis and non-polyposis ones, divided according to their pattern of inheritance. For each syndrome, the most common germline mutated genes are indicated. FAP: Familial Adenomatous Polyposis; PJS: Peutz-Jeghers syndrome; PHTS: PTEN Hamartoma Tumor Syndrome; JPS: Juvenile Polyposis Syndrome; PPAP: Polymerase Proofreading Associated Polyposis; MAP: MUTYH-associated polyposis; NAP: NTHL1-associated polyposis; CMMRD: constitutional mismatch repair deficiency syndrome; LS: Lynch syndrome; LLR: Lynch-like syndrome; FCCTX: Familial colorectal cancer type X syndrome (modified from Olkinuora et al., 2021).*

1.3.2.1 Lynch syndrome (LS) and constitutional MMR deficiency syndrome (CMMRD)

Lynch syndrome (formerly known as hereditary nonpolyposis colorectal cancer, HNPCC) is an autosomal dominant inherited syndrome that represents 3% of CRC cases. The average age of onset of CRC in LS patients is approximately 45 years and the tumor usually arise in the proximal (right-sided) colon with a rapid carcinogenic process. Individuals affected by LS have an increased risk for extra-colonic malignancies such as endometrial and ovarian cancers[25]. LS is caused by germline monoallelic loss-of-function mutations in MMR genes, i.e. *MLH1* (42% of variants), *MSH2* (33%), *MSH6* (18%), and *PMS2* (7.5%)[22], or by deletion of the gene *EPCAM* that causes the silencing of its neighboring gene *MSH2*[26]. *MLH1* may also be silenced by constitutional epimutation (*MLH1* promoter methylation)[27].

Pathogenic variants (PVs) in *MLH1* are associated with the highest risk for CRC and the risk of developing CRC is significantly higher for *MLH1* and *MSH2* than for *MSH6* or *PMS2* PV carriers. PVs in *MSH2* are associated with the highest risk for extracolonic cancers, especially endometrial cancer[28].

The heterozygous germline mutations in MMR genes convey vulnerability to a second somatic mutation in the wild-type allele, leading to deficient mismatch repair (dMMR) tumors characterized by microsatellite instability (MSI)[29]. LS is associated with an up to 80% risk of developing microsatellite unstable cancer. dMMR tumors can be also of sporadic origin[30].

Biallelic germline pathogenic variants in the four MMR genes result in a rare inherited cancer predisposition syndrome named constitutional MMR deficiency syndrome (CMMRD), characterized by high risk of malignancies in childhood and adolescence. A large proportion of CMMRD patients develop multiple (up to>100) synchronous polyps[24].

1.3.2.2 Familial adenomatous polyposis (FAP)

Familial adenomatous polyposis (FAP) is responsible for less than 1% of CRC cases and results in the development of hundreds to thousands of polyps throughout the colon (mostly in the left side) and rectum in young age (7-36 years of age). About 95% of individuals with FAP develop polyps by 35 years of age[31,32].

An attenuated form of FAP, called attenuated familial polyposis (aFAP), is characterized by a smaller number of polyps (< 100) that occur around 50–55 years of age[31]. Both FAP and aFAP are caused by germline heterozygous PVs in different domains of *APC*, a tumor suppressor gene that regulates cell growth, cellular adhesion, cytoskeleton stabilization[32].

1.3.2.3 Hamartomatous polyposis syndromes

Hamartomatous polyposis syndromes are a heterogeneous group of rare autosomal dominant hereditary syndromes, that induce the development of hamartomatous polyps in the gastrointestinal tract. They include Peutz-Jeghers syndrome (PJS), juvenile polyposis syndrome (JPS), and PTEN hamartoma tumor syndrome (PHTS)[33]. The histology of hamartomatous polyps permit to distinguish between Peutz-Jeghers polyps, that are typically multilobulated and covered by hyperplastic glandular mucosa, and juvenile polyps, that are spherical and include inflammatory cells[34].

Most cases of Peutz-Jeghers (94%) syndrome are due to germline mutations in the *STK11* tumor suppressor gene involved in the mTOR pathway. Instead, juvenile polyposis syndrome is caused by germline heterozygous mutations in *BMPR1A* (~28%) and in *SMAD4* (~27%), while for the 45% of the cases the genetic cause is not established[34,35]. PHTS includes various clinical entities characterized by the overgrowth of multiple hamartomas in several organs, due to a heterozygous germline pathogenic variant in *PTEN*[36].

1.3.2.4 MUTYH-associated polyposis (MAP)

MUTYH-associated polyposis is an autosomal recessive hereditary syndrome caused by biallelic (homozygous or compound heterozygous) mutations in *MUTYH,* which encodes a DNA glycosylase involved in base excision repair (BER) by repairing 8-oxoG caused by oxidative damage[37]. The phenotypes associated with MAP are highly variable, ranging from 1-10 colonic adenomas before 40 years of age, to 10-100 colonic adenomas and/or hyperplastic polyps, to more than 100 colonic polyps in the absence of a germline *APC* mutation[37]. Monoallelic *MUTYH* mutation carriers with a family history of CRC are at increased risk of colorectal, gastric, endometrial and possibly liver cancers[38].

1.3.2.5 NTHL1-associated polyposis (NAP)

Similarly to *MUTYH*, biallelic germline mutations in *NTHL1,* another BER gene, are responsible for NTHL1-associated polyposis, characterized by an increased risk for colorectal polyposis (1-100 polyps), CRC and breast cancer[39,40].

It has not been yet established if carriers of a heterozygous germline *NTHL1* mutations have an increased cancer risk[24], but some heterozygote individuals developed cancer and showed loss of heterozygosity in the tumor tissue[40].

1.3.2.6 Polymerase proofreading-associated polyposis (PPAP)

Heterozygous pathogenic missense variants in *POLE* and *POLD1* exonuclease domains have been reported to predispose to polymerase proofreading-associated polyposis (PPAP), characterized by multiple colorectal adenomas and CRC. The mutations map in the exonuclease domain (that has a proofreading function) and cause a defect in the correction of misincorporated bases inserted during DNA replication[24,41].

1.3.2.7 Serrated polyposis syndrome (SPS)

The serrated polyposis syndrome (SPS) is characterized by the presence of multiple serrated polyps throughout the colon. Serrated polyps are considered the precursors of up to 15–30% CRC through the serrated pathway. SPS is caused by heterozygous mutations in the RING-type E3 ubiquitin ligase *RNF43*, an inhibitor of the Wnt pathway[24].

1.3.2.8 Familial colorectal cancer type X syndrome (FCCTX)

The familial colorectal cancer type X (FCCTX) describes a group of MMR-proficient cases with a heterogeneous genetic background[22]. Some genes that have been reported to be potentially associated with FCCTX are *BMPR1A, RPS20, SEMA4A, SETD6, BRCA2, OGG1*, *FAN1, CENPE, CHD18, GREM1, BCR, KIF24, GALNT12, ZNF367, HABP4, GABBR2, BMP4, APC, NTS, TP53, SMAD4*[42,43]. Some of those genes (*APC, BMPR1A, SMAD4*) are also associated with polyposis syndromes.

These genes are involved in several functions, including DNA repair. *OGG1* encodes 8-oxoguanine glycosylase, an enzyme that belongs to the base excision repair pathway, repairing oxygen reactive DNA lesions[43]. Both BRCA2, a member of the Fanconi anemia pathway, and the nuclease FAN1 are involved in interstrand DNA cross-links repair[44,45].

RPS20 encodes a component of the small ribosome subunit. A frameshift variant has been identified in a four-generation FCCTX Finnish family co-segregating with CRC and all studied tumors were MMR proficient[46]. Other two cases were part of a panel that included 863 early onset/familial CRC patients[47]. Loss-of-function heterozygous variants in ribosomal protein (RP) genes cause Diamond Blackfan anemia (DBA)[48]. Registry data indicate that individuals with DBA have an increased risk of CRC, that is the most prevalent solid tumor in young adults DBA patients[49,50].

GALNT12 encodes the enzyme N-acetylgalactosaminyltransferase-type 12 involved in the O-glycosylation of mucin-type glycans. *GALNT12* loss of function variants confer a moderate susceptibility for CRC with an autosomal dominant pattern of inheritance[22,51–53], as well as *SEMA4A* that encodes for a semaphorin receptor with immunomodulatory effects and growth regulatory functions[54].

1.3.3 Sedentary lifestyle

Physical inactivity and a sedentary lifestyle are well-known colon cancer risk factors[55]. Sedentary time is increasing due to the rising prevalence of office work and to the changing in lifestyle habits, such as spending many hours sitting and watching TV[56]. A sedentary lifestyle can contribute to colorectal carcinogenesis by adiposity accumulation and metabolic dysfunction[57].

The American Cancer Society recommends moderate-intensity activity for at least 150 min or vigorous-intensity for 75 min throughout the week[6]. Physical activity result in a >20% reduction of the risk of CRC[58]. The beneficial effect of physical activity may be due to positive effects on gut motility, on metabolic hormone regulation, on immune system, on tissue oxygenation and on basal metabolism[6]. Standing and moderate physical activities improve blood flow and skeletal muscle function and result in improved glucose regulation[56].

1.3.4 Cigarette smoking

According to the World Cancer Research Fund (WCRF) and American Institute of Cancer Research, smoking 40 cigarettes per day increases CRC risk by up to 40% and doubles CRC mortality rate compared to non-smokers.

Ex-smokers show a higher risk to develop CRC compared to non-smokers, even if they have stopped to smoke over the past 25 years[59].

Cigarette smoke contains a mixture of toxic chemical compounds, including polynuclear aromatic hydrocarbons, nitrosamines and aromatic amines. They reach the colorectal mucosa through the circulatory system or through direct ingestion and induce genetic and epigenetic aberrations that lead to an increased risk for CRC[60].

1.3.5 Obesity

Obesity is defined by the World Health Organization (WHO) as "abnormal or excessive fat accumulation that may impair health". Chronic inflammation and excessive adipose tissue expansion are central characteristics of obesity[61].

Adiposity is measured with anthropometric parameters, usually body mass index (BMI, calculated as body weight in kilograms divided by the square of height in meters), and waist circumference (WC)[62].

According to WHO, overweight individuals have a BMI $\geq$ 25 kg/m^2, while obese $\geq$ 30 kg/m^2. It has been reported that each 5 kg/m^2 increase in BMI is associated with an increased risk for CRC of 5%[3].

The adipose tissue is divided into two distinct compartments: subcutaneous adipose tissue (SAT) and visceral adipose tissue (VAT), that is more linked to CRC[6,62].

The BMI represents the overall body fatness, while the WC reflects visceral fatness and is a stronger risk factor for CRC than BMI.

The adipose tissue is an active endocrine and metabolic organ that affects the physiology of other tissues through the release of free fatty acids, adipokines and cytokines[63].

The two most important endocrine adipokines produced by adipose tissue are leptin – a proinflammatory hormone that suppresses appetite, increases basal metabolism and has levels proportional to the adipose tissue volume – and adiponectin – an insulin sensitizing hormone, with anti-inflammatory effects and inversely correlated with BMI[61,63]. Leptin has been reported to stimulate cell proliferation, migration and invasion through PI3K/AKT/mTOR pathway, supporting tumor development[61]. Conversely, adiponectin inhibits cell proliferation activating the AMPK pathway and its reduction is related to CRC development[61].

In addition, VAT is heavily infiltrated with immune cells (such as lymphocytes and macrophages) that cause chronic low-grade systemic inflammation and insulin resistance[6], conditions associated with oxidative stress and cancer initiation and progression[61]. In the lean state, the healthy adipose tissue is enriched with anti-inflammatory immune cells that attenuate inflammation. As obesity develops, the secretion of inflammatory cytokines, including IL-1β (interleukin 1β), IL-6 (interleukin 6), and TNFα (tumor necrosis factor), increases and adipose inflammation is promoted[61].
Obesity-related inflammation causes insulin resistance, characterized by reduced metabolic response of tissues to insulin and subsequent hyperinsulinemia, a compensatory response to maintain normal blood glucose level[63]. In obese patients, hyperinsulinemia leads to the activation of the insulin-like growth factor (IGF) receptors, that result in activation of PI3K, mTOR and MAPK pathways, leading to promotion of cell proliferation and inhibition of apoptosis of cancerous colon epithelial cells[64].

1.3.6 Western diet

An unhealthy dietary pattern, such as the Western diet, is characterized by a high intake of red and processed meat, added sugar, sugar-sweetened beverages, desserts, refined grains, and potatoes[6]. This unhealthy dietary pattern is associated with an increased risk of CRC, tumor recurrence and mortality[65–67]. Conversely, a healthy/prudent dietary pattern is characterized by a high intake of fruits, vegetables, whole grains, low-fat dairy products, and fish[6].

1.3.6.1 Red and processed meat

CRC is associated with high red and processed meat consumption[3,68]. To understand which is the real risk factor for CRC, it is important to consider different aspects of meat consumption such as the amount, the type (fresh, red, or white), and processing and cooking methods. Different mechanisms of action for red and processed meat consumption in increasing CRC risk have been proposed:
- Heterocyclic aromatic amines (HAAs) and polycyclic aromatic hydrocarbons (PAHs) are considered as genotoxic compounds because they act directly on DNA, causing mutations[69]. These compounds are produced when the meat is cooked at high temperatures[70].
- N-nitrous compounds (NOCs), including nitrosamines, nitrosamides and nitrosoguanidines, can induce DNA GC→AT transitions and are synthesized both exogenously during meat or food processing and endogenously by reactions catalyzed by the intestinal microbiota[71,72]. They are present in bacon, cured meats, sausages, ham, smoked fish, and smoked cheeses, stored under humid conditions, smoked in air saturated with nitrogen, dried at high temperatures and cured with nitrate

and/or nitrite[73]. Endogenous NOCs production occurs within the colon after ingestion of red meat containing heme[74].

High content of heme iron in meat is associated with increased proliferation of colonocytes[69]. Through N-nitrosation, heme iron contributes to the endogenous formation of NOCs[71,72]. Dietary intake of heme iron increases lipid peroxidation and free radicals formation[75], causing DNA strand breaks and oxidative DNA damage[76].

- Trimethylamine-N-oxide (TMAO) is a gut metabolite that induces inflammation and cell proliferation[77]. TMAO precursors derive from animal-origin foods[78]. TMAO is produced from choline, phosphatidylcholine, and L-carnitine by microbiota–mediated formation of trimethylamine (TMA). TMA is transported into the liver where is oxidized into TMAO[79].
- Increased meat intake is associated with an increased fat intake and favors insulin resistance and a higher production of secondary bile acids[69].

High fat diet and consumption of red and processed meat increase secretion of primary bile acids (BAs) that can be metabolized by the gut bacteria to secondary BAs[3]. The secondary BAs induce oxidative DNA damage, metabolic stress, membrane perturbation, production of ROS and reactive nitrogen species, that create an altered microenvironment favoring CRC development[80].

1.3.6.2 Added sugar and sugar-sweetened beverages

Added sugar indicates sweeteners, such as high fructose corn syrup (HFCS) or sucrose, added to food and beverages. Sugar-sweetened beverages (SSBs) include soft drinks, fruit drinks, and sports drinks[81]. SSBs and high sugar intake, especially in adolescence, may alter the insulin-like growth factor axes causing insulin resistance, obesity, type 2 diabetes and contribute to inflammation, thus promoting colorectal carcinogenesis[82,83].

Fructose is a major ingredient of SSBs, and, unlike glucose, is metabolized predominantly in the liver. High levels of fructose intake (as in HFCS-sweetened beverages) trigger hepatic lipogenesis[81]. Apc defective mice treated with low doses of fructose, that do not induce obesity or metabolic dysfunction, develop a higher number of colon tumors and of higher grade compared to controls[84]. Fructose leads to activation of glycolysis and increased synthesis of fatty acids in the tumor, supporting its growth[84]. Moreover, in high-fat diet fed mice dietary fructose promotes intestinal cell survival and villous hypertrophy, increasing nutrient absorption and adiposity[85].

1.3.6.3 Alcohol consumption

Alcohol consumption is associated with an increased CRC risk in a dose-dependent manner[5]. Consuming two or three alcoholic beverages per day (~30 g/day) is associated with an increased risk of 20%, while for higher doses the risk is increased to 40%[5].

Ethanol is an established risk factor for CRC due to its first metabolite, acetaldehyde, that is considered as a carcinogen (Group 1) by the International Agency for Research (monographs.iarc.who.int).

Ingested alcohol reaches colonocytes through systematic circulation and probably diffuses into the lumen. Here, ethanol is metabolized by alcohol dehydrogenase into acetaldehyde, which causes mucosal injury and cell proliferation. The toxic acetaldehyde also enters the intestinal epithelial cells and can promote colorectal carcinogenesis by inducing DNA damage and reducing absorption of folate, required for proper DNA synthesis and methylation[6,86].

Other mechanisms that are involved in alcohol-induced carcinogenesis are:

- Increased gut permeability: excessive ethanol consumption causes alteration in the gut barrier function[87], facilitating the passage of other environmental carcinogens (e.g., aflatoxins, benzene, asbestos, etc.)[88].
- Induction of inflammation: ethanol causes an inflammation in murine mucus and submucous colonic layers and an increased pro-inflammatory cytokines production (IL-1α, IL-6, and TNF-α)[89]. TNF-α and IL-6 secretion following ethanol consumption is favored by synergistic interaction with gut microbiota. A study conducted on CaCo2 cells has shown that high ethanol concentration increases pro-inflammatory cytokines production in synergy with *E. coli*[87].
- DNA adduct formation: this process is mediated by cytochrome P450 2E1 (CYP2E1)[88]. This enzyme stimulates ethanol metabolism resulting in reactive oxygen species (ROS) production that causes lipid peroxidation and DNA adducts formation[90].
- Epigenetic alterations: alcohol consumption may increase epigenetic alterations because it reduces folates (vitamin B9) absorption and metabolism[91]. Reduced B9 levels cause aberrant methylation of DNA, contributing to carcinogenesis[92].

1.4 Protective factors

Protective factors for CRC include physical activity, prudent diet (high consumption of fruit, vegetables, fish, and whole-grain products), and non-steroidal anti-inflammatory drugs (NSAIDs)[93].

1.4.1 Fruit and vegetables

According to the WCRF, fruits and vegetables reduce CRC risk with "limited" evidence. The CUP (Continuous Update Project) has identified 17 studies on the association between fruits and vegetables, and CRC. Ten of these studies have been included in a dose-response meta-analysis showing inverse association for each 100g/day of fruits.

Fruits and vegetables are rich of bioactive compounds like polyphenols, flavonoids, and soluble fiber, vitamins, and minerals. A recent meta-analysis supports the hypothesis that these food categories represent CRC protective factors due to a high flavonoids intake[94]. This is probably due to their anti-inflammatory, antioxidant, and pro-apoptotic properties[95].

Anthocyanins, hydrosoluble pigments, seem to have a potential protective role in CRC thanks to their ability to negatively control some inflammatory pathways (NF-kB, MAPK, JNK), and inhibit Wnt signaling pathway, with an anti-proliferative role[96].

Fruits and vegetables also contain vitamins, especially vitamin Bs which are involved in DNA synthesis, repair, and methylation[97]. In mice models, vitamin B6 reduces the number of polyps in the colon and suppresses cell proliferation; both mechanisms are involved in CRC prevention and reduction[98,99].

Some studies have reported that high consumption of vitamin B9, or folic acid, is associated with reduced CRC and adenoma risks, but only if they come from the diet (especially from deep green leafy vegetables), and not by supplements[100].

Niacin, or vitamin B3, has hypolipidemic and anti-inflammatory effects. Studies in mice suggest that niacin supplementation can prevent colitis and CRC[101,102].

1.4.2 Fish and omega-3

Fish intake has a potentially positive impact on CRC, as it is an important dietary source of vitamin D and omega-3 fatty acids[69]. Polyunsaturated fatty acids are classified into omega-3 and omega-6 groups. Omega-3 (α-linolenic acid (ALA), eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA)) have anti-inflammatory and hypotrigliceridemic properties. In particular, EPA and DHA ameliorate the inflammation in adipose tissue and decrease insulin resistance[103].

A 12% reduction of CRC risk thanks to fish consumption has been reported[104].

The omega-3 fatty acids exert their antitumor actions through different mechanisms (regulation of Wnt/β-catenin and Hippo pathways, regulation of oxidative stress and the expression of Granzyme B), involving proliferation, apoptosis, and migration[103].

1.4.3 Fiber and whole grains

Since 2017, the WCRF supports the thesis that fiber and whole grains "probably" reduce CRC risk. A fiber consumption of at least 10 g/day is associated with a reduced CRC risk by up 10%[105].

Among protective mechanisms associated with fiber consumption, the increased fecal volume and the accelerated gut passage reduce the possibility that carcinogens interact with the gut mucosa[106]. Moreover, fibers reduce weight gain[107] and type 2 diabetes risk[108] and are fermented by the gut microbiota producing metabolites with anti-proliferative effects as short chain fatty acids[109].

One of the main dietary fiber sources are the whole grain cereals, which are beneficial for human health compared to their refined counterparts. Refined grains are obtained when bran and germ are removed, resulting in a reduction in fiber and micronutrients content[110]. Moreover, whole grains are an important source of polyphenols and flavonoids. These phenolic compounds have antioxidant properties with probable anti-tumoral properties[110].

Overall, whole grains represent protective factors against gastrointestinal cancers. Conversely, refined cereals are associated with an increased risk of CRC[110,111].

1.4.4 Non-steroidal anti-inflammatory drugs (NSAIDs)

The association between aspirin – and other nonsteroidal anti-inflammatory drugs (NSAIDs) – and colorectal cancers has been widely studied[112]. A chemopreventive effect of NSAIDs has been mainly attributed to cyclooxygenase COX-2 inhibition, that leads to the reduction of tumor cell proliferation and angiogenesis through the suppression of prostaglandin E2 production[113].

1.5 Gut microbiota and metabolome

The collection of microorganisms (bacteria, viruses, fungi, protozoa) colonizing the gastrointestinal (GI) tract is called the 'gut microbiota', which counts for more than 10^{14} cells[114]. Bacteria that colonize gut microbiota belong mainly to the phyla Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria[115], and follow a different distribution along the GI tract due to modification of the oxygen gradient[116].

When these microorganisms live in an equilibrium state – i.e. eubiosis – they perform different beneficial functions for the host. This equilibrium may be compromised resulting in an altered microbial composition, known as dysbiosis[117].

Among their benefic functions, gut microbes are able to ferment complex carbohydrates producing metabolites such as short-chain fatty acids (SCFAs), of which propionate, butyrate and acetate are

predominant. These SCFAs are rapidly absorbed by epithelial cells in the GI tract where they are involved in the regulation of cellular processes such as gene expression, differentiation, chemotaxis, proliferation, and apoptosis. These metabolites also play an important role in immune system regulation and in the inflammatory response, by influencing the secretion of anti-inflammatory cytokines[118].

In addition, the gut microbiota can synthesize vitamins for the host. In fact, lactic acid bacteria are key organisms in the production of vitamin B12, while Bifidobacteria are the main producers of folate. Further vitamins like vitamin K, riboflavin, biotin, nicotinic acid, pantothenic acid, pyridoxine, and thiamine may be synthesized by gut microbiota[119].

The presence of microbial species in the human GI tract influences also possible pathogen colonization, by competing for attachment or nutrients, and by producing antimicrobial proteins, which kill bacteria through direct interaction disrupting the bacterial cell wall via enzymatic attack[120].

Moreover, the gut microbiota helps to maintain the integrity of the mucosal barrier. The GI epithelium provides a physical barrier facilitated by tight junction proteins that interact with the cytoskeleton forming a complex structure that limits the gut permeability. The intestinal epithelium provides also a chemical barrier thanks to the outer mucus layer secreted by the goblet cells of the epithelium itself. Commensal bacteria producing anti-microbial proteins are found on the intestinal epithelium. In the lumen, commensal bacteria produce bacteriocins, modify the lumen pH, and compete for nutrients avoiding pathogenic bacteria colonization[121]. The physical and chemical barrier functions of the intestinal epithelium help to maintain gut integrity and homeostasis. If the protective capacity of the gut barrier is compromised, bacterial translocation and entry of toxic microbial products, such as pro-inflammatory endotoxins and metabolites, across the colonic epithelium occur[121].

A primary role of microbial dysbiosis in the development of different diseases is the gut barrier weaknesses, which causes a shift in microbial communities with commensal bacteria that may become pathogenic. The barrier weakness is due to a disturbance in the epithelium architecture with a destabilization of tight junctions and increase bacterial invasiveness[121]. This altered condition could induce a further shift in the microbiota composition, which may cause chronic inflammation and colorectal cancer onset.

1.5.1 Bacteria associated with CRC

The gut dysbiosis is associated with different types of cancer, including CRC. Specific changes in the microbiome occur during different stages of colorectal cancer, from adenomatous polyps to early-stage cancer to metastatic disease, supporting an etiological role of the gut microbiota[3].

Several bacterial species may have a direct role in CRC like *Fusobacterium nucleatum*, *Bacteroides fragilis*, and *Escherichia coli*, all species found enriched in lumen-associated microbiota (LAM) or mucosa-associated microbiota (MAM) of CRC patients compared to healthy controls[122–125].

F. nucleatum is a gram-negative anaerobe bacterium that colonize the oral cavity. This species is enriched in colorectal cancer tissue compared to normal colon mucosa[126,127]. *F. nucleatum* is present in approximately 10%-15% of colorectal tumors and has been associated with advanced disease stage, poor survival, and a higher risk of recurrence[3,127,128]. In addition, the presence of this species in tumor tissue has been associated with a reduced infiltration of T cells and a reduced anti-tumor immune response[3,129].

The enterotoxigenic *B. fragilis* (ETBF), which secretes the *B. fragilis* toxin, plays an important role in CRC initiation. This toxin is a zinc-dependent metalloprotease that targets epithelial tight junctions causing E-cadherin cleavage and increasing gut barrier permeability. In addition, the toxin induces colon inflammation and DNA damage[130–134].

Another pathogen involved in CRC carcinogenesis is a particular *Escherichia coli* strain, which is called *pks*$^+$ *E. coli* since is equipped with polyketide synthase gene complex (*pks*)[135]. *pks* encodes for colibactin, a genotoxin able to induce DNA double-strand breaks, cell cycle arrest and chromosomal aberrations[136,137].

1.5.1.1 The "driver-passenger" model

The "driver-passenger" model has been proposed as a model of the interaction between bacteria and CRC development[138] (Fig. 1.5). It is the bacterial counterpart of the genetic model according to which driver genetic and epigenetic mutations lead to epithelial dysplasia in the colon and then initiated cells accumulate other different mutations that promote tumor progression. Similarly, driver bacteria contribute to CRC initiation, while passenger bacteria are opportunistic pathogens that take advantage of alterations induced by tumorigenesis and proliferate[138].

Driver bacteria are defined as bacteria with pro-carcinogenic characteristics that may initiate CRC development. These bacteria produce genotoxic compounds that damage colonic epithelial cells DNA; stimulate cleavage of some tumor suppressor proteins causing cell proliferation and permeabilization of the gut barrier; induce a chronic inflammatory response, pushing colonic epithelium from inflammation towards carcinogenesis[138]. Candidate drivers include the enterotoxigenic *Bacteroides fragilis (ETBF)* and *Escherichia coli pks*$^+$ [139].

Passenger bacteria are poor colonizers of the healthy intestinal tract and cannot breach the intact colon wall; however, when an adenoma or carcinoma is present, the damaged gut barrier may allow these bacteria to pass across the intestinal barrier. Their proliferation is favored by the tumor

microenvironment and may be involved in CRC progression[138]. For example, Tjalsma and colleagues considered as a candidate passenger *Fusobacterium nucleatum*[138], that is over-represented on colorectal tumors compared to adjacent normal tissue[126,127]. However, it has been reported that *F. nucleatum* may act both as driver and passenger[140]. Kostic *et al.* found that the introduction of *F. nucleatum* accelerated the onset of colonic tumors in mice[126] and lipopolysaccharide of *F. nucleatum* could trigger β-catenin and NF-κB pathways through the Toll-like receptor 4 cascade, suggesting the driver role of *F. nucleatum*[140]. *F. nucleatum* acts in CRC progression through two virulence factors: FadA, that activates the E-caderin/β-catenin pathway in colon cancerous cells[141,142], and Fap2, that interacts with Gal-GalNAc residues, overexpressed on the surface of CRC cells, and with the TIGIT receptor present NK cells and on various T cells, inhibiting immune cells activity[126,143].

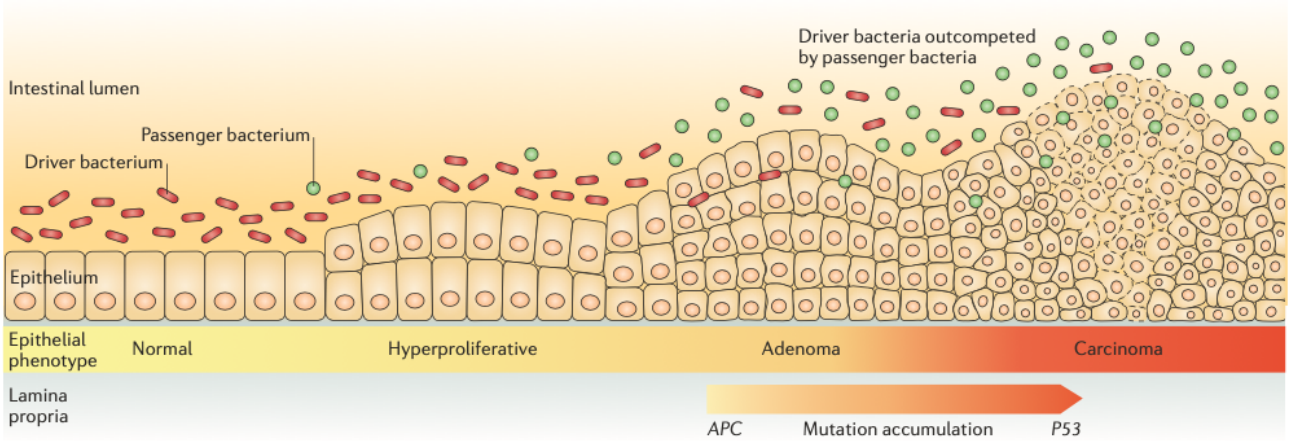


Figure 1.5. *The bacterial driver–passenger model for colorectal cancer (Tjalsma et al., 2012).*

1.5.2 Metabolome and CRC

The gut microbiota synthesizes a large variety of metabolites (volatile small molecules, lipids, proteins and peptides, sugars, secondary bile products or terpenoids, biogenic amines, oligosaccharides, glycolipids, organic acids and amino acids) that may affect the host physiology and contribute to pathogenesis of diseases[144]. The total collection of metabolites in a biological sample (e.g. feces, biopsies, urine, serum) is known as metabolome, and its analysis has individuated many metabolic biomarkers for CRC[145].

1.5.2.1 Short chain fatty acids (SCFAs)

The most studied gut microbiota-derived metabolites are short-chain fatty acids (SCFAs), that have been shown to reduce the risk of CRC by reducing levels of pro-inflammatory cytokines and increasing levels of anti-inflammatory cytokines and T-reg cells[3,146]. Non-absorbable dietary fibers are fermented by members of the intestinal microbial community to obtain SCFAs (e.g. propionate,

butyrate, and acetate)[146]. Members of Bacteroidetes in the proximal colon mainly produce acetate and propionate, while butyrate is produced by members of Firmicutes[139].

Butyrate has different effects on carcinogenesis according to the host genetic background, as extensively explained in Chapter 6. Briefly, although it increases the expression of pro-apoptotic and tumor-suppressor genes in cancerous colonocytes, it accelerates hyperproliferation in mice with germline mutations of *Apc* and MMR genes[147,148].

1.5.2.2 Trimethylamine N-oxide (TMAO)

An important gut microbe-dependent metabolite is Trimethylamine N-oxide (TMAO), that is produced after the metabolization of L-carnitine, betaine and dietary choline to trimethylamine (TMA) by gut microbiota[146]. As previously mentioned, red meat has a high content of choline and carnitine, and its consumption is therefore associated with increased levels of the pro-inflammatory metabolite TMAO[146,149].

1.5.2.3 Secondary bile acids

Bile acids (BAs) are classified as primary (e.g. cholic acid, CA), secondary (e.g. deoxycholic acid, DCA), and tertiary (e.g. taurocholic acid, TCA). Gut bacteria contribute to the production of secondary BA through the metabolism of primary BAs, that have been synthesized from cholesterol in the liver. Tertiary BAs are formed via hepatocyte metabolism of reabsorbed primary BAs. High levels of secondary BAs have been correlated with an increased risk of CRC[139]. DCA causes DNA damage in epithelial cells leading to apoptosis, induces oxidative stress and activates NF-kB promoting inflammation[150,151].

1.5.2.4 Polyamines

Polyamines are biosynthesized from the amino acids arginine and ornithine and can derive from diet or biosynthesis by host and gut microbiota[152]. For example, spermidine is a metabolite produced either by host or microbiota or consumed with diet. Enterotoxigenic *B. fragilis* through its spermin oxidase (SMO) generates H_2O_2 by the conversion of spermine to spermidine, promoting cellular oxidative stress[133].

Polyamines induce intracellular oxidative stress and DNA damage, accelerating carcinogenesis and facilitating cell proliferation and tumor metastasis[133,146].

Chapter 2

Objectives and study design

This study was part of two large projects (the FOHN Project to the Department of Health Sciences of Università del Piemonte Orientale funded by the Italian Ministry of University and Research under the "Departments of Excellence 2018-2022" program, and the ID. 25886 project funded by AIRC under IG 2021), through which we have obtained a wide range of data that allowed us to evaluate the role of various CRC risk factors: gut microbiota and metabolites, diet, lifestyle, obesity, inherited predisposition.

We recruited a large panel of 306 patients with colon polyps who underwent colonoscopy at the Ospedale Maggiore della Carità (Novara) and the Azienda Ospedaliera Universitaria Città della Salute e della Scienza (Torino). Histological analyses were conducted in the Pathology Unit of the Hospitals. We collected blood samples for genetic analyses on germline DNA, and two different swabs brushed on freshly removed polyps for mucosa-associated microbiota (MAM) and metabolome analyses, using a novel procedure that we developed[153]. Fecal samples for lumen-associated microbiota (LAM) analyses were collected fifteen days after colonoscopy. At the same time, anthropometric parameters (height and weight to calculate Body Mass Index (BMI), waist circumference (WC) and skinfolds) were collected. Two questionnaires validated by EPIC (European Prospective Investigation into Cancer), the first about lifestyle and the second about nutrition and dietary habits, were also administered. Nutrient amount and frequencies were then analyzed through the EPIC website. A third questionnaire concerned personal and familial history of cancer.

Germline genetic analyses were performed by targeted-NGS using a custom panel that includes 107 genes involved in inherited cancer syndromes. Both MAM and LAM were analyzed initially by 16S rDNA sequencing and then by shotgun sequencing, that reaches the bacterial strain level of resolution. Mucosa-associated metabolome was analyzed using bidimensional chromatography and mass spectrophotometry (GCxGC-MS).

Since bacteria in close contact with enterocytes may play an important role in colon carcinogenesis, we focused our attention on MAM and mucosa-associated metabolites to identify key carcinogenic pathogens, whereas LAM, although still essential as clinical marker, only seems to recapitulate the general gut environment.

Our objective was to shed light on the role of the intestinal microbiota on colon carcinogenesis and its interaction with obesity and inherited predisposition syndromes. Thus, we stratified patients according to the grade of dysplasia of their polyps, to their BMI and WC, or to their genetic background. Our aims were the following:

1. Firstly, we wanted to explore the relationship between MAM, mucosa-associated metabolome and CRC initiation and progression. We followed the hypothesis that some bacteria – found on low-grade polyps – may act as candidate drivers, involved in CRC initiation, and other bacteria

– found on high-grade polyps – may act as candidate passengers, possibly involved in CRC progression. We also considered that metabolite profiles accompanied the changes in bacterial composition associated to tumor stage, and we performed a microbiota-metabolome integrated analysis. The results obtained by the comparison of patients with high-grade *vs* low-grade dysplastic polyps are described in Chapter 4 [154].

2. Secondly, we wanted to explore the role of body weight, diet and lifestyle cofactors in colon carcinogenesis. Our hypothesis was that unhealthy diet and obesity alter gut microbiota favoring colorectal carcinogenesis. We reasoned that different microbiota components could be involved in colon carcinogenesis in obese and normal-weight patients. Thus, we analyzed polyp-associated microbiota and metabolites in obese and normal-weight patients, also evaluating the patient's dietary habits. The different bacterial and metabolome signatures obtained comparing normal-weight and obese patients, and their dietary habits, are shown in Chapter 5 (manuscript in preparation).
3. Thirdly, we wanted to analyze genetic predisposition and gut microbiota relationship. Thus, we first identified germline pathogenic variants (PVs) in cancer predisposing genes in individuals with colon polyps and classified patients as mutated or sporadic. Then, we performed MAM shotgun sequencing analysis to compare patients with or without inherited predisposition to cancer (i.e. mutated *vs* sporadic). These preliminary data are reported in Chapter 6 (preliminary data).

Chapter 3

A new method for investigating microbiota-produced small molecules in adenomatous polyps

In recent years, intestinal microbiota has become an increasingly popular research topic. Most studies have analyzed microbiota-produced metabolites on fecal samples[155,156], but this analysis may not accurately reflect the small molecules present on the intestinal mucosa surface. Therefore, we developed a high-throughput metabolomics method to capture the metabolites adherent to adenomatous polyps and adenocarcinoma. The new approach consists in gently brushing a swab on the adenoma surface to collect mucosa-associated metabolome after polyp removal, without jeopardizing tissue integrity that is essential for histological analysis. The presented method enables the simultaneous quantification of almost 300 metabolites, was analytically validated and had excellent performances in terms of recovery, linearity, specificity, intra- and inter-day precision, limits of detection, and quantification. Furthermore, as clinical validation, the method was shown to discriminate patients with low-grade from patients with high-grade dysplastic polyps, including adenocarcinomas[153].

Analytica Chimica Acta 1179 (2021) 338841

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

A new method for investigating microbiota-produced small molecules in adenomatous polyps

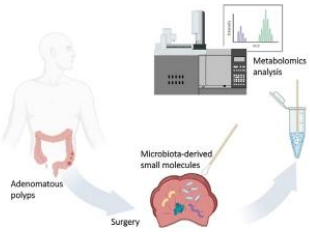
Elettra Barberis [a, b], Soni Joseph [c], Elia Amede [a, b], Michela Giulia Clavenna [c], Marta La Vecchia [c], Marika Sculco [c], Anna Aspesi [c], Pietro Occhipinti [d], Elisa Robotti [e], Renzo Boldorini [c], Emilio Marengo [e], Irma Dianzani [c, 1], Marcello Manfredi [a, b, *, 1]

[a] Department of Translational Medicine, University of Piemonte Orientale, Novara, Italy
[b] Center for Translational Research on Autoimmune and Allergic Diseases, University of Piemonte Orientale, Novara, Italy
[c] Department of Health Sciences, University of Piemonte Orientale, Novara, Italy
[d] Department of Gastroenterology, 'Maggiore Della Carità' Hospital, Novara, Italy
[e] Department of Sciences and Technological Innovation, University of Piemonte Orientale, Novara, Italy

HIGHLIGHTS

- Validated non-invasive metabolomics method for quantitative analysis of microbiota-produced small molecules.
- Ability to extract small molecules directly produced by the microbiota adherent to the surface of an adenoma.
- Simultaneous identification and quantification of almost 300 metabolites.
- Ability to investigate the microbiota-produced small molecules associated with histologic grade of dysplasia.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 April 2021
Received in revised form
30 June 2021
Accepted 4 July 2021
Available online 7 July 2021

Keywords:
Microbiota-derived metabolome
Metabolomics
Non-invasive
Microbiome
Colorectal cancer

ABSTRACT

The intestinal microbiota is composed of a large number of different bacteria that produce a variety of metabolites. Colorectal cancer, which typically develops from adenomatous polyps, is highly influenced by microbiota. Since a variety of molecular changes may occur as these polyps transform from benign tumor to malignant carcinoma, the ability to study the microbiota-produced metabolites can lead to new discoveries about the development and progression of this cancer. However, to address the complexity of the microbiota-produced molecules, novel methods are needed. To this aim, in the present work, we developed a high-throughput metabolomics method to capture the metabolic complexity of the microbiota metabolome adherent to adenomatous polyps and adenocarcinoma. For the first time, the method enables the simultaneous quantification of almost 300 metabolites, while preserving the integrity of the original sample. The metabolomics approach was analytically validated and had excellent performances in terms of recovery, linearity, specificity, intra- and inter-day precision, limits of detection,

* Corresponding author. author. Department of Translational Medicine, University of Piemonte Orientale, Novara, Italy.
E-mail address: marcello.manfredi@uniupo.it (M. Manfredi).
[1] co-last author.

https://doi.org/10.1016/j.aca.2021.338841

and quantification. Furthermore, the clinical potential of the method was demonstrated in adenoma collected for a colorectal adenoma study.

1. Introduction

In recent years, intestinal microbiota has become an increasingly popular research topic due to its numerous important functions [1]. In particular, recent publications have revealed that intestinal dysbiosis may generate several disorders, including cancer [2–4].

According to GLOBACON estimates, in 2018, colorectal cancer (CRC) was the third most frequently diagnosed cancer, globally characterized by a high rate of death [5,6]. While in recent decades the prognosis for patients with CRC has slowly improved, the 5-year relative survival rate is still 50–65% despite the decline in the mortality rate due to screening aimed at early detection of lesions.

In most of the cases, CRC develops from adenomatous polyps through the acquisition of dysplastic changes [7]. These changes, occurring during different stages of the adenoma-carcinoma sequence, promote the carcinogenic metabolite production that leads to neoplasia [8]. To understand CRC development, several studies have analyzed microbiota-produced metabolites on fecal samples [2,3]. This analysis, however, may not completely reflect the small molecules present on the intestinal mucosa, and in particular on the surface of benign or cancerous tissues. This is particularly important in CRC because it is well-known that the microbiota can influence CRC development through their metabolites [8,9] (e.g., short-chain fatty acids [SCFAs] such as butyrate), which are produced by bacterial fermentation of undigested fibers in the colon [10]. Brown et al. developed an interesting method for cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies by means of methanol and water extractions [11,12]. However, the fixation of tissue samples in alcohol is not the standard method, and more validation is still needed to ensure that biopsies can be used for other analyses through which the pathologist reaches a diagnosis.

No other methods are available for analyzing microbiota-produced metabolites on polyps while maintaining tumor tissue integrity. The aim of the present work was to accomplish for the first time the development of a high-throughput metabolomic method to capture the metabolic complexity of the microbiota-produced small molecules that are adherent to adenomatous polyps and adenocarcinoma. To overcome the limitations of the existing methods, which require sample destruction or are not completely suitable for subsequent histological analysis, we developed and validated a new analytical workflow to perform metabolomic analysis and histological evaluation on the same biopsy sample. This metabolomic approach was subsequently applied to samples collected in a panel of patients with colorectal adenomas as a case study to perform validation of the methodology in a clinical setting.

A bi-dimensional gas chromatography coupled to mass spectrometry instrument (GCxGC-TOFMS) was used for the quantification of both SCFAs and polar metabolites. SCFAs are usually quantified by gas chromatography (GC) and liquid chromatography (LC), but LC methods require specific derivatizations and interfering peaks are often reported [13,14].

2. Experimental section

2.1. Materials

Methyl *tert*-butyl ether, N,O-Bis(trimethylsilyl)trifluoro acetamide(BSTFA), methoxamine, hexadecane, amino acids standards, propanoic acid d2 and tridecanoic acid, acetic acids d4, glycine d5 and stearic acids d35 were purchased from Merck (Darmstad, Germany). Pyridine, water and acetonitrile were from VWR (Milano, Italy). Methanol and isopropanol were from Scharlab (Barcelona, Spain). Commercial mix standard of free fatty acids including acetic acid, propanoic acid, butanoic acid, isobutyric acid (propanoic acid, 2-methyl), isovaleric acid (butanoic acid, 3-methyl) and valeric acid (pentanoic acid) were from Restek (Bellefonte, PA). Swabs were from Copan (Brescia, Italy).

2.2. Sampling of microbiota-produced small molecules

To eliminate possible contaminants, before sample collection, sterile swabs were washed for 1 h in water, placed in an oven at 60 °C for 2 h, and then conserved in a methanol-water cleaned tube filled with nitrogen gas. In order to reduce the sampling variability, the rubbing was always performed by the same nurse that was previously taught. The following protocol was used: after the endoscopic removal of adenomatous polyp, the biopsy was held by a sterile forceps and the swab was gently rubbed on the surface of the biopsy by performing a complete rotation of the swab on the entire surface of the adenomatous polyp. The sampling was performed taking care to sample the same surface only once. In addition, metabolites concentrations/abundances can be further normalized for the size of the polyps. The procedure allows the absorption of small molecules adherent to the adenoma surface. The swab was then kept at −80 °C until the analysis.

2.3. Extraction of short-chain fatty acids (SCFAs)

SCFAs were extracted from the swab using water and sonication, followed by a liquid–liquid extraction with methyl *tert*-butyl ether (MTBE). Briefly, the swab, without the wood, was placed in a tube, then 300 μL of water and 15 μL of the internal standard propanoic acid d2 (20.4 ppm) were added, and the sample was vortexed for 20 s, followed by centrifugation for 15 s. The tube was then placed in a cold sonication bath for 30 min. The swab was squeezed with cleaned tweezers and 200 μL of water extract was placed in a new tube, bringing pH to 2 using 6 M HCl. For the extraction of SCFAs, 140 μL of MTBE was added, and the tube was placed on a rotator for 15 min, followed by centrifugation for 10 min at 4 °C and 21.1×g. Then, 100 μL of the organic phase containing the SCFAs was analyzed with bi-dimensional gas chromatography coupled to mass spectrometry (GCxGC-TOFMS). The remaining water phase solution was then subjected to a second extraction as described next.

2.4. Extraction of metabolites from the aqueous phase

The remaining aqueous phase, which contained other small molecules such as amino acids, sugars, medium fatty acids, and long fatty acids, was extracted again using methanol–isopropanol–acetonitrile. Briefly, 200 μL of water from the first liquid–liquid extraction was subjected to a second extraction using a 1 mL mixture of acetonitrile (ACN), methanol (MeOH), and water 3:3:2, with 5 μL of tridecanoic acid (0.5 μg/mL) as the internal standard. The sample was vortexed for 15 s and centrifuged for 15 min at 20 °C at 14.5×*g*. Next, 1 mL of the supernatant was placed in a new tube, dried in a speed vacuum, and stored at −20 °C until derivatization. The samples were finally derivatized with methoximation (20 μL of methoxamine, 80 °C, 20 min) and silylation (30 μL of N,O-Bis(trimethylsilyl)trifluoroacetamide, 80 °C, 20 min) prior to the GCxGC-TOFMS analysis.

2.5. SCFAs and amino acid calibration curves

The calibration curves of SCFAs were carried out in a concentration range from 5 ppb to 80 ppm, while the calibration curves for amino acids were calculated in a range from 11 ppb to 20 ppm. The reproducibility and the recovery of the method were determined by depositing the standard solutions directly on the swab, then conserving the swab at −80 °C for one night, followed by extraction of the molecules using the protocol already discussed. Quality control samples (QC) containing spiked standards were also acquired before and after the analysis of the clinical samples.

2.6. GCxGC-TOFMS analysis

For the instrumental analysis, a LECO Pegasus BT 4D GCxGC-TOFMS instrument (Leco Corp., St. Josef, MI, USA) equipped with a LECO dual stage quad jet thermal modulator was used. The GC part of the instrument was an Agilent 7890 gas chromatograph (Agilent Technologies, Palo Alto, CA), equipped with a split/splitless injector. The first dimension column was a 30 m Rxi-5Sil (Restek Corp., Bellefonte, PA) MS capillary column, with an internal diameter of 0.25 mm and a stationary phase film thickness of 0.25 μm for metabolite from aqueous solution, while for SCFAs analysis the column was a 30 m DB-FATWAX-UI (Agilent Technologies, Santa Clara, CA) with a diameter of 0.25 mm and a film thickness of 0.25 μm, and the second dimension chromatographic columns was a 2 m Rxi-17Sil MS (Restek Corp., Bellefonte, PA) with a diameter of 0.25 mm and a film thickness of 0.25 μm. High-purity helium (99,9999%) was used as the carrier gas with a flow rate of 1.4 mL/min 1 μL of sample was injected in splitless mode at 250 °C. The temperature program for metabolites analysis was as follows: the initial temperature was set at 70 °C for 2 min, then ramped 6 °C/min up to 160 °C, 10 °C/min up to 240 °C, 20 °C/min to 300 and then held at this value for 6 min. The secondary column was maintained at +5 °C relative to the GC oven temperature of the first column. The temperature program for SCFAs was as follows: the initial temperature was 40 °C for 2 min, then ramped 7 °C/min up to 165 °C, 25 °C/min up to 240 °C, maintained for 5 min. The secondary column was maintained at +5 °C relative to the GC oven temperature of the first column. Electron impact ionization was applied (70 eV). The ion source temperature was set at 250 °C, the mass range was 40–300 m/z with an extraction frequency of 32 kHz, for the SCFAs analysis, while for the metabolites the mass range was 25–550 m/z. The acquisition rates were 200 spectra/s. The modulation periods for both programs were 4s for the entire run. The modulator temperature offset was set at +15 °C relative to the secondary oven temperature, while the transfer line was set at 280 °C [15,16].

2.7. Data analysis

The chromatograms were acquired in TIC (total ion current) mode. Peaks with signal-to-noise (S/N) value lower than 500.0 were rejected. ChromaTOF version 5.31 was used for raw data processing. Mass spectral assignment was performed by matching with NIST MS Search 2.3 libraries adding Fiehn Library. Commercial mix standard of free fatty acids composed by acetic acid, propanoic acid, propanoic acid 2-methyl, butanoic acid, butanoic acid 3-methyl and pentanoic acid was run individually and EI spectra were matched against the NIST library. The calibration curves of the SCFAs were obtained using Excels, while the analytical results were processed and compared with the open source software MetaboAnalyst5.0 (www.metaboanalyst.org) [17].

Partial least square discriminant analysis (PLS-DA) was then applied to identify candidate biomarkers with a variable selection strategy in backward elimination, allowing the selection of the most discriminant variables according to the smallest % classification error rate in cross-validation (leave-more-out procedure with 5 cancellation groups of 20% of the samples randomly selected each time, procedure repeated 1000 times): at each iteration of the variable selection algorithm, the variable with the lowest VIP score was eliminated [18]. The classification performances were evaluated on the basis of several parameters: accuracy %, Non-Error-Rate % (NER %), sensitivity, specificity and precision. PLS-DA was carried out by Matlab R2014a (The Mathworks, Natick, MA, USA), exploiting in-house developed routines and the Classification Toolbox from Milano Chemometrics [19,20].

2.8. Application of the protocol in a colon adenoma study

For clinical validation of the new metabolomics method, we analyzed the microbiota-produced small molecules from adenomatous polyps collected in the gastroenterology department of Maggiore della Carità Hospital in Novara. Twenty patients carrying at least one adenomatous polyp with a size ≥1 cm were enrolled in the present study.

Table 1 reports demographic and histopathologic data.

The study was performed in accordance with the principles contained in the Declaration of Helsinki. The international ethical committee of the Novara Hospital (Novara, Italy) approved the study protocol, and all the participants provided written informed consent.

3. Results

3.1. Analytical validation

The microbiota metabolome can be characterized through the analysis of fecal samples or intestinal biopsies: the fecal metabolites are produced by the overall intestinal microbiome community

Table 1
Demographic and histopathologic features of the patients included in the pilot application of the new method.

Clinical features	Patients $N = 30$ (male; female)
Gender	
Male	21
Female	9
Age	
Mean ± SD	62.3 ± 9.3
Histology	
Low-grade dysplasia	15 (10 M; 5F)
High-grade dysplasia	9 (7 M; 2F)
Adenocarcinoma	6 (4 M; 2F)

via intestinal fermentation and also cell metabolism. The ability to study the microbiota-derived small molecules that are adherent to adenomatous polyps or to the adenocarcinoma collected during colonoscopy could be of great help in understanding the development of CRC because gut metabolites are related to different CRC stages [21]. One of the most important issues in the analysis of biopsies is sample degradation, which is necessary to extract small molecules for analysis. However, tissue destruction invalidates the possibility to perform histologic diagnosis and this can not be done in the clinical practice. To overcome this limitation, we developed a non-invasive method that is able to sample the small molecules from the surface of a tissue sample. A very efficient method of analysis should enable the extraction, identification, and reliable quantitative estimation of major and trace components while leaving unchanged the original sample for subsequent histological analysis. The developed method not only has all of those features but also allows for the analysis of metabolites that are directly produced by the microbiota of adenomatous polyps and adenocarcinoma and which are thus present on the surface of a specific tissue. The protocol enables the analysis of SCFAs, which is the most important class of molecules for microbiome study, and also sugars, medium- and long-chain fatty acids, amino acids, and organic acids, and others.

3.2. SCFA analysis

The validation of the method was carried out on the main SCFAs (acetic acid, propanoic acid, propanoic acid 2-methyl, butanoic acid, butanoic acid 3-methyl, and pentanoic acid). Mixed standard solutions were deposited on sterile and cleaned swabs and extracted with the developed method after being kept for one night at −80 °C. SCFAs were then analyzed with GCxGC-TOFMS. For each analyte, a calibration curve reporting the peak area of the *m/z* transition signal (y) versus the concentration of the standard solution (x) was created. Concentration levels in the following ranges were considered: from 5 ppm to 80 ppm for acetic acid (5 ppm, 10 ppm, 20 ppm, 40 ppm, 80 ppm), from 5 ppb to 2 ppm for propanoic acid and 2-methyl propanoic acid (5 ppb, 10 ppb, 20 ppb, 50 ppb, 100 ppb, 200 ppb, 500 ppb, 1 ppm, 2 ppm), from 5 ppb to 0.2 ppm for butanoic acid (5 ppb, 10 ppb, 20 ppb, 50 ppb, 100 ppb, 200 ppb), from 5 ppb to 0.1 ppm for 3-methyl butanoic acid (5 ppb, 10 ppb, 20 ppb, 50 ppb, 100 ppb), and from 5 ppb to 0.5 ppm for pentanoic acid (5 ppb, 10 ppb, 20 ppb, 50 ppb, 100 ppb, 200 ppb, 500 ppb). Moreover, to overcome possible memory effects, the standard solutions were injected in randomized order. For all the analytes, a linear regression fit with a weighting factor of 1/x was used, and good linearity was obtained. The regression coefficients, reported in Table 2, were the following: 0.98 for acetic acid, 0.99 for propanoic acid, 0.99 for propanoic acid 2-methyl, 0.96 for butanoic acid, 0.99 for butanoic acid 3-methyl, and 0.99 for pentanoic acid. Table 2 also summarizes the intercept, slope, and linearity range.

The limit of detection (LOD) and limit of quantitation (LOQ) were calculated considering signal-to-noise ratios of 3 and 10, respectively, and the values obtained were the following for LOD and LOQ, respectively: 6 and 8 ppm for acetic acid, 5.6 and 7 ppb for propanoic acid, 6.2 and 9 ppb for 2-methyl propanoic acid, 5.6 and 7 ppb for butanoic acid, 5.6 and 7 ppb for 3-methyl butanoic acid, and 5.5 and 6.5 ppb for pentanoic acid. The LODs and LOQs with the proposed method were very good. Table 3 reports the inter-day precisions of the concentration evaluated by analyzing each analyte every day (three replicates) for 3 successive days. The results showed that the inter-day precisions were 9.94% for acetic acid, 3.07% for propanoic acid, 7.87% for 2-methyl propanoic acid, 6.40% for butanoic acid, 6.63% for 3-methyl butanoic acid, and 2.25% for pentanoic acid. To evaluate the recovery R (as a percentage) of each analyte and to verify its possible dependence on the concentration, the analyte standard solutions at the different concentration levels used for the calibration curve were deposited on a new swab and then extracted and analyzed using the developed protocol. The recovery values were calculated as C_{obs}/C_{ref}, where C_{obs} is the difference between the concentration determined for the spiked sample and the native concentration in the same sample, and C_{ref} is the spiked concentration. For all the analytes, a percentage recovery R was calculated at low (5 ppm for acetic acid, 10 ppb for other SCFAs), medium (20 ppm for acetic acid, 100 ppb for other SCFAs), and high (50 ppm for acetic acid, 1000 ppb for other SCFAs) concentrations (reported in Table 4). As can be observed, the R ranged in the following intervals: from 91.44% to 117.44% for acetic acid, from 81.49% to 114.71% for propanoic acid, from 45.96% to 100.65% for 2-methyl propanoic acid, from 53.89% to 80.94% for butanoic acid, from 23.32% to 119.78% for 3-methyl butanoic acid, and from 94.82% to 90.93% for pentanoic acid. Because the recovery is also correlated to the concentration, a correction must be applied when analyzing real samples. In addition, the validation of the curve and of the method was also performed by analyzing two QC samples, prepared by spiking SCFAs standards at 1 ppm concentration, before and after the analysis of real samples from colorectal polyps reported in section 3.4. The results showed a very good accuracy (2.2% for acetic acid, 4.0% for propanoic acid, 3.4% for propanoic acid, 2-methyl, 5.1% for butanoic acid, 5.7% for butanoic acid, 3-methyl and 4.7% for pentanoic acid).

Because we used an untargeted approach for the SCFA quantification, the analysis of real samples allowed the identification and relative quantification of other molecules from different classes as benzene and substituted derivatives, benzothiazoles, fatty acyls, carboxylic acids and derivatives, fatty alcohols, fatty aldehydes, cinnamaldehydes, keto acids and derivatives, furanones and

Table 3
Intra-day and inter-day precision of the method calculated as a percentage of the coefficient of variation (CV%) for each analyte.

Molecule	Intra-Day CV (%) $n = 3$	Inter-Day CV (%) $n = 3$
Acetic acid	5.55%	9.94%
Propanoic acid	6.69%	3.07%
Propanoic acid, 2-methyl	7.31%	7.88%
Butanoic acid	4.31%	6.40%
Butanoic acid, 3-methyl	3.64%	6.64%
Pentanoic acid	1.05%	2.25%

Table 2
Coefficient of correlation, slope, intercept, and linearity range for the SCFA standards used for validation of the method.

Molecule	Coefficient of correlation	Slope	Intercept	Linearity range
Acetic acid	0.98	1×10^8	3×10^8	8ppm–80 ppm
Propanoic acid	0.99	4×10^8	-9×10^5	7ppb–2 ppm
Propanoic acid, 2 methyl	0.99	8×10^8	1×10^8	9ppb–2 ppm
Butanoic acid	0.96	4×10^8	1×10^7	7ppb–0.2 ppm
Butanoic acid, 3 methyl	0.99	1×10^9	6×10^6	7ppb–0.1 ppm
Pentanoic acid	0.99	3×10^8	-4×10^4	6.5ppb–0.5 ppm

Table 4
Percentage of recovery R calculated at low, medium, and high concentrations for each analyte.

Molecule	Low $n = 3$	Medium $n = 3$	High $n = 3$
Acetic acid	91.44 ± 0.72	114.15 ± 1.19	117.44 ± 1.34
Propanoic acid	81.49 ± 0.98	103.63 ± 1.08	114.71 ± 1.26
Propanoic acid, 2-methyl	45.96 ± 0.94	89.59 ± 0.97	100.65 ± 1.25
Butanoic acid	53.89 ± 0.39	73.13 ± 0.95	80.94 ± 0.99
Butanoic acid, 3-methyl	23.32 ± 0.58	79.53 ± 0.91	119.78 ± 1.56
Pentanoic acid	94.82 ± 0.83	101.43 ± 1.23	90.93 ± 1.05

dihydrofurans, cinnamic acids and derivatives, and phenols. The complete list of the identified and quantified molecules in the adenomatous polyps is reported in Supplementary Table 1.

3.3. Aqueous phase metabolites analysis

We have further validated the aqueous phase extraction using amino acids standards. Mixed standard solutions were deposited on sterile and cleaned swabs and were extracted with the developed method. Amino acids were then analyzed with GCxGC-TOFMS. For each analyte, a calibration curve reporting the peak area of the "quantifier" transition signal (y) versus standard concentration (x) was built with a concentration range from 10 ppb to 20 ppm (10 ppb, 20 ppb, 50 ppb, 100 ppb, 200 ppb, 500 ppb, 1 ppm, 2 ppm, 5 ppm, 10 ppm, 20 ppm). The regression coefficients, intercept, slope, and linearity range are reported in Table 5. The LOD and LOQ, respectively, of the amino acids were as follows: 180 ppt and 14 ppb for phenylalanine, 1 ppb and 11 ppb for aspartic acid, 700 ppt and 55 ppb for isoleucine, 200 ppt and 25 ppb for methionine, 1 ppb and 48 ppb for proline, 100 ppt and 20 ppb for valine, and 14 ppb and 150 ppb for lysine. In addition, the inter-day and intra-day reproducibility were also calculated. The intra-day variations ranged from 3.37% to 10.58%, while the inter-day variations ranged from 2.90% to 13.32% (Table 6). The recoveries percentages reported in Table 7 were good and directly correlated to the concentration used. The validation of the curve was also performed by analyzing two QC samples, prepared by spiking amino acids standards at 10 ppb concentration, before and after the analysis of real samples from colorectal adenomatous polyps reported in section 3.4. The results showed a very good accuracy that ranges from 1.3% of proline to 9.84% of aspartic acid.

We also investigated the identity of metabolites extracted with the second liquid–liquid extraction. The complete list of identified molecules in real samples collected from adenomatous polyps is reported in Supplementary Table 2. More than 200 molecules were identified: medium (8–10 C-atoms), long (14–20 C-atoms), and very long (22 or more) chain fatty acids, including subclasses of monounsaturated with omega-9 (MUFAs), polyunsaturated fatty acids (PUFAs) with omega-3 and omega-6; saturated fatty acids; essential fatty acids (EFAs); amino acids; phenols, sugars, sterols, and so on.

Table 6
Intra-day, inter-day precision, and recovery percentage of the method, calculated as a percentage of the coefficient of variation (CV) for each analyte.

Molecule	Intra-Day CV (%) $n = 3$	Inter-Day CV (%) $n = 3$
Phenylalanine	4.37%	6.52%
Aspartic acid	9.95%	11.32%
Isoleucine	10.10%	10.09%
Methionine	9.37%	3.46%
Proline	3.37%	2.90%
Valine	10.58%	13.32%
Lysine	9.67%	9.28%

3.4. Application to a colorectal adenomatous polyps

Then the developed method was applied for the analysis of the small molecules produced by the gut microbiota in twenty patients with low-grade dysplasia (n = 10) and high-grade dysplasia/adenocarcinoma (n = 10). Adenomatous polyps removed by the gastroenterologist were sampled using the swab, and then small molecules adsorbed from the surface of the biopsy were analyzed to quantify SCFAs and other small molecules. The metabolomics analysis not only allowed for the absolute quantification of the main SCFAs but also the identification and relative quantification of 208 and 81 relevant molecules in the aqueous and organic phases, respectively (Supplementary Tables 1–2). Since the polyps had the same sizes, metabolites concentrations/abundances were not normalized for the size. The hierarchical clustering heat map and the partial least square discriminant analysis (PLS-DA) reported in Fig. 2A–B shows the presence of a microbiota metabolic signature associated with low-grade dysplasia (green) and high-grade dysplasia (red) adenomas and adenocarcinoma (red). The analysis of SCFAs showed a higher concentration of 3-methyl butanoic acid (isovaleric) in high-grade polyps and in polyps with adenocarcinoma transformation (Fig. 2C).

PLS-DA was applied with a variable selection procedure in backward elimination: at each iteration, one variable was eliminated according to the minimum error in cross-validation (leave-more-out with 5 cancellation groups). The variable with the smallest VIP score in cross-validation was eliminated. The final model included 29 biomarkers (Supplementary Table 3) and provided the perfect classification of all the samples in the two classes in fitting (non-error-rate = 100%) and very good results in cross-validation (random selection of 5 cancellation groups; 1000 iterations), with a non-error-rate% of 99.95%.

We have extended the present approach for the analysis of microbiota-produced small molecules to ten more patients as further validation, 5 with low-grade dysplasia and 5 with high-grade dysplasia/adenocarcinoma polyps. As reported in Fig. 2C (light green and light red dots), the results confirmed that high-grade/adenocarcinoma polyps are characterized by higher concentration of 3-methyl butanoic acid (light red dots).

Table 5
Coefficient of correlation, slope, intercept, and linearity range of the amino acid standards used for validation of the method.

Molecule	Coefficient of correlation	Slope	Intercept	Linearity range
Phenylalanine	0.99	5×10^7	2×10^7	14 ppb–14 ppm
Aspartic acid	0.99	2×10^7	8×10^6	11 ppb–6 ppm
Isoleucine	0.99	5×10^7	1×10^7	55 ppb–6 ppm
Methionine	0.97	9×10^7	8×10^6	25 ppb–1 ppm
Proline	0.99	7×10^7	5×10^6	48 ppb–5 ppm
Valine	0.99	3×10^7	7×10^7	20 ppb–20 ppm
Lysine	0.99	6×10^6	-3×10^6	150 ppb–15 ppm

Table 7
Percentage of recovery R calculated at low, medium, and high concentrations for each amino acid.

Molecule	Low (15 ppb) $n = 3$	Medium (100 ppb) $n = 3$	High (1 ppm) $n = 3$
Phenylalanine	45.03 ± 0.87	61.65 ± 0.31	84.04 ± 0.56
Aspartic acid	60.04 ± 1.33	69.37 ± 2.49	95.55 ± 0.66
Methionine	10.63 ± 0.84	39.57 ± 1.82	44.04 ± 0.82
Proline	41.74 ± 0.92	91.34 ± 0.43	130.72 ± 1.38
Lysine	72.08 ± 0.48	73.03 ± 0.66	75.67 ± 0.33
Valine	13.45 ± 0.75	21.55 ± 1.23	73.54 ± 0.54
Isoleucine	60.56 ± 0.98	63.08 ± 0.44	60.20 ± 0.28

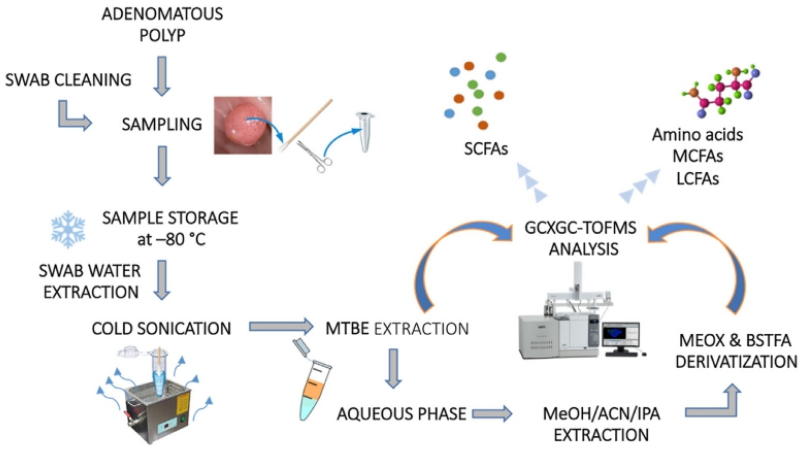


Fig. 1. Experimental workflow of the developed method. The adenomatous polyps were sampled using a cleaned swab and then kept at −80 °C until the analysis. SCFAs were first extracted using MTBE and then analyzed with GCxGC-TOFMS, while small molecules contained in the aqueous phase were extracted using MeOH/ACN/IPA and then derivatized before GCxGC-TOFMS analysis.

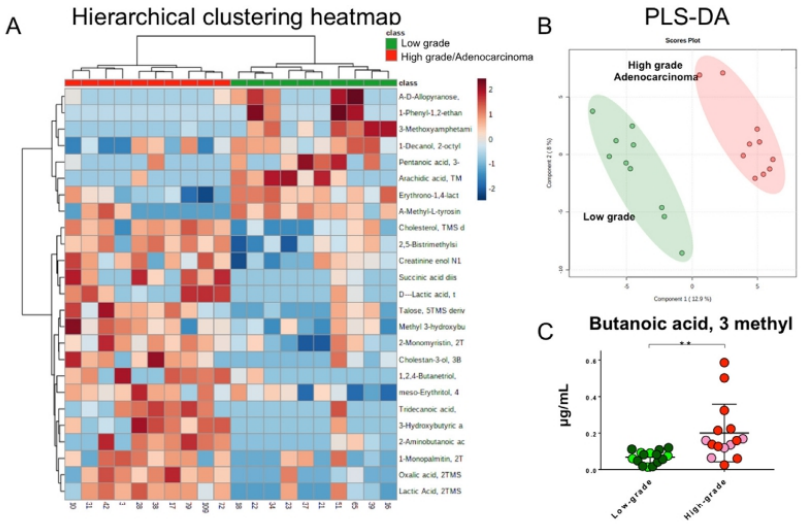


Fig. 2. Metabolomics profiles of microbiota-derived molecules in adenomatous polyps. (A) Hierarchical clustering heat map, (B) PLS-DA analysis, and (C) box plot of butanoic acid 3-methyl in low-grade (discovery and validation samples, respectively in green and light green) and high-grade dysplasia or adenocarcinoma (discovery and validation samples, respectively in red and light red). *P*-value = 0.01. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

This work presents the development of a novel method for studying microbiota-produced small molecules, as illustrated in Fig. 1. Adenomatous polyps were removed by a gastroenterologist and then small molecules, in particular SCFAs, were sampled using a swab. The swab was stored at −80 °C until the analysis. SCFAs and other small molecules were extracted from the swab using two different liquid−liquid extractions and analyzed with GCxGC-TOFMS. The method was fully validated for the absolute quantification of acetic acid, propanoic acid, butanoic acid, isobutyric (propanoic, 2-methyl) acid, isovaleric (butanoic, 3-methyl) acid, and valeric acid (pentanoic acid). The most important SCFAs were satisfactorily validated in terms of specificity, linearity, recovery, intra-day and inter-day precision, LOQ and LOD. In addition, almost 300 molecules were detected, including not only fatty acids (short, medium, and long), which are well-known players in microbiota studies, but also other molecules such as amino acids, steroids, alcohols, organic and keto acids, and so on (Supplementary Table 4). These results highlight the wide range of applicability of the methodology in clinical research.

Current protocols of metabolite extraction from biopsy involve destruction of the sample, precluding additional uses of the tissue for histologic diagnosis and this is not admissible in the clinical practice. A mix of solvents extracts the metabolites, and after a few steps, the tissue molecules are ready to be analyzed, but the sample is forever lost. Using that protocol, the metabolomics of mucosal biopsies of patients with ulcerative colitis, collected during colonoscopies, were compared to biopsies from healthy subjects, revealing changes in the abundance of lysophosphatidylcholine, acyl carnitine, and amino acid profiles [22]. Another study showed that the metabolomic profile of routine needle biopsies was able to identify specific tumor type metabolic signatures for breast cancer stratification [23]. However, the metabolome identified was not actually associated with the microbiota but, rather, mostly with the analyzed tissues. To the best of our knowledge, the metabolomics method developed in this study represents the first approach that allowed for the characterization of microbiota-derived small molecules from adenomatous polyps and adenocarcinoma, providing a biochemical snapshot of the tissue microenvironment. Furthermore, it is also noteworthy that our method consists of two liquid−liquid extractions, which allows the SCFAs and polar metabolites to be studied in a unique and single sample, unlike all the conventional metabolomic approaches that enable the analysis of only one class of molecules. In fact, SCFAs are usually extracted using acidified water, diethyl ether, dichloromethane, ethyl acetate, or MTBE [24], and no methods are available to extract small molecules remaining in the aqueous phase. This thus emphasizes the great potential of the metabolomic approach developed here to obtain a comprehensive and realistic overview of the microbiota-produced metabolome.

Finally, in this work, the methodology was applied for the analysis of small molecules produced by the microbiota in colon adenomatous polyps. The statistical analyses showed that the metabolome profile present on the surface of the polyp tissue is correlated with the grade of dysplasia and with the adenocarcinoma transformation. For example, the concentrations of 3-methyl butanoic acid were increased in patients with adenomas with high grade of dysplasia and in adenocarcinoma compared with healthy controls, and concentrations of 3-methyl butanoic acid had already been found to be greater in the stool of patients with adenocarcinoma [25]. Although preliminary, our results, seem to indicate that these altered metabolite levels might be useful to distinguish adenomatous polyps with low grade dysplasia from polyps with high-grade dysplasia and adenocarcinoma. Finally, this technology may provide information on the mechanism of colorectal cancer formation.

These results therefore demonstrate the high potentiality of the developed method in the characterization of microbiota in a reliable, accurate, and sensitive manner, with short- and medium-chain fatty acids predominantly detected in the organic phase, whereas the aqueous phase provided more insights into amino acids, amines, and carboxylic acids.

5. Conclusion

In summary, this work describes for the first time the development, validation, and application of a metabolomic method for the comprehensive and quantitative analysis of microbiota-produced small molecules. The main strengths of this novel approach are (i) the ability to extract small molecules directly produced by the microbiota adherent to the surface of an adenoma and (ii) the simultaneous identification and quantification of almost 300 metabolites, (iii) leaving unchanged the original samples, which therefore facilitates its implementation in large-scale clinical studies. Conversely, the most important limitation of this methodology is the high variability of the adenoma sampling, which may depend on the tumor volume, morphology, localization, patient preparation, and colonoscopy instrument used. Furthermore, this new method was tested in adenomatous polyps collected in a colon adenoma cancer study with the aim of assessing the performance of the technique on real samples and exploring its ability to investigate the microbiota-produced small molecules associated with histologic grade of dysplasia as a case study. In this regard, future studies are needed to assess the clinical potential of this new method in larger sample cohorts.

CRediT authorship contribution statement

Elettra Barberis: Design of the study, Methodology, Formal analysis, Writing. **Soni Joseph:** Sampling, Review. **Elia Amede:** Methodology, Writing. **Michela Giulia Clavenna:** Sampling, Design of the study, Review. **Marta La Vecchia:** Sampling, Design of the study, Review. **Marika Sculco:** Sampling, Design of the study, Review. **Anna Aspesi:** Sampling, Design of the study, Review. **Pietro Occhipinti:** Sampling, Design of the study, Review. **Elisa Robotti:** Formal analysis, Review. **Renzo Boldorini:** Sampling, Design of the study, Review. **Emilio Marengo:** Formal analysis, Review. **Irma Dianzani:** Design of the study, Supervision, Project administration, Funding acquisition, Writing – review & editing. **Marcello Manfredi:** Design of the study, Supervision, Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Italian Ministry of Education, University and Research (MIUR) program “Departments of Excellence 2018−2022”, FOHN to ID. This study was (partially) funded by the AGING Project—Department of Excellence—DIMET, Università del Piemonte Orientale, MIUR ITALY to MM.

Appendix A. Supplementary data

Supplementary data to this article can be found online at

https://doi.org/10.1016/j.aca.2021.338841.

References

[1] S.V. Lynch, O. Pedersen, The human intestinal microbiome in health and disease, N. Engl. J. Med. 375 (2016) 2369–2379.

[2] M.X. Chen, S.Y. Wang, C.H. Kuo, I.L. Tsai, Metabolome analysis for investigating host-gut microbiota interactions, J. Formos. Med. Assoc. 118 (2019) 10–22.

[3] B. Routy, V. Gopalakrishnan, R. Daillère, L. Zitvogel, J.A. Wargo, G. Kroemer, The gut microbiota influences anticancer immunosurveillance and general health, Nat. Rev. Clin. Oncol. 15 (2018) 382–396.

[4] D. Ternes, J. Karta, P. Tsenkova, P. Wilmes, S. Haan, E. Letellier, Microbiome in colorectal cancer: how to get from meta-omics to mechanism? Trends Microbiol. 28 (2020) 401–423.

[5] F. Bray, J. Ferlay, I. Soerjomataram, R.L. Siegel, L.A. Torre, A. Jemal, Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries, Cancer J Clin 68 (2018) 394–424.

[6] H. Brenner, M. Kloor, C.P. Pox, Colorectal cancer, Lancet 383 (2014) 1490–1502.

[7] C.N. Arnold, A. Goel, H.E. Blum, C.R. Boland, Molecular pathogenesis of colorectal cancer: implications for molecular diagnosis, Cancer 104 (2005) 2035–2047.

[8] P. Louis, G.L. Hold, H.J. Flint, The gut microbiota, bacterial metabolites and colorectal cancer, Nat. Rev. Microbiol. 12 (2014) 661–672.

[9] C. Hassan, A. Gimeno-García, M. Kalager, C. Spada, A. Zullo, G. Costamagna, D. Senore, D.K. Rex, Systematic review with meta-analysis: the incidence of advanced neoplasia after polypectomy in patients with and without low-risk adenomas, Aliment. Pharmacol. Ther. 39 (2014) 905–912.

[10] H.M. Hamer, D. Jonkers, K. Venema, S. Vanhoutvin, F.J. Troost, R.-J. Brummer, Review article: the role of butyrate on colonic function, Aliment. Pharmacol. Ther. 27 (2008) 104–119.

[11] M.V. Brown, J.E. McDunn, P.R. Gunst, E.M. Smith, M.V. Milburn, D.A. Troyer, K.A. Lawton, Cancer detection and biopsy classification using concurrent histopathological and metabolomic analysis of core biopsies, Genome Med. 4 (2012) 33.

[12] T. Huan, D.A. Troyer, L. Li, Metabolite analysis and histology on the exact same tissue: comprehensive metabolomic profiling and metabolic classification of prostate cancer, Sci. Rep. 6 (2016) 32272.

[13] G. Liebisch, J. Ecker, S. Roth, S. Schweizer, V. Öttl, H.F. Schött, H. Yoon, D. Haller, 5 Holler, E.R. Burkhardt, S. Matysik, Quantification of fecal short chain fatty acids by liquid chromatography tandem mass spectrometry—investigation of pre-analytic stability, Biomolecules 9 (2019) 121.

[14] H.E. Song, H.Y. Lee, S.J. Kim, S.H. Back, H.J. Yoo, A facile profiling method of short chain fatty acids using liquid chromatography-mass spectrometry, Metabolites 9 (2019) 173.

[15] E. Barberis, S. Timo, E. Amede, V.V. Vanella, C. Puricelli, G. Cappellano, D. Raineri, M.G. Cittone, E. Rizzi, A.R. Pedrinelli, V. Vassia, F.G. Casciaro, S. Priora, I. Nerici, A. Galbiati, E. Hayden, M. Falasca, R. Vaschetto, P.P. Sainaghi, U. Dianzani, R. Rolla, A. Chiocchetti, G. Baldanzi, E. Marengo, M. Manfredi, Large-scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2, Int. J. Mol. Sci. 21 (2020) 8623.

[16] M. Manfredi, E. Conte, E. Barberis, A. Buzzi, E. Robotti, V. Caneparo, D. Cecconi, J. Brandi, E. Vanni, M. Finocchiaro, M. Astegiano, M. Gariglio, E. Marengo, M. DeAndrea, Integrated serum proteins and fatty acids analysis for putative biomarker discovery in inflammatory bowel disease, J. Proteomics. 195 (2019) 138–149.

[17] Z. Pang, J. Chong, S. Li, J. Xia, MetaboAnalystR 3.0: toward an optimized workflow for global metabolomics, Metabolites 10 (2020) 186.

[18] A. Oussama, F. Elabadi, S. Platikanov, F. Kzaiber, R. Tauler, Detection of olive oil adulteration using FT-IR spectroscopy and PLS with variable importance of projection (VIP) scores, J. Am. Oil Chem. Soc. 89 (10) (2012) 1807–1812.

[19] M. Manfredi, E. Robotti, F. Quasso, E. Mazzucco, G. Calabrese, E. Marengo, Fast classification of hazelnut cultivars through portable infrared spectroscopy and chemometrics, Spectrochim. Acta Mol. Biomol. Spectrosc. 189 (2018) 427–435.

[20] D. Ballabio, V. Consonni, Classification tools in chemistry. Part 1: linear models. PLS-DA, Anal. Methods 5 (16) (2013) 3790–3798.

[21] S. Yachida, S. Mizutani, H. Shiroma, S. Shiba, T. Nakajima, T. Sakamoto, H. Watanabe, K. Masuda, Y. Nishimoto, M. Kubo, F. Hosoda, H. Rokutan, M. Matsumoto, H. Takamaru, M. Yamada, T. Matsuda, M. Iwasaki, T. Yamaji, T. Yachida, T. Soga, H. Kurokawa, A. Toyoda, Y. Ogura, T. Hayashi, M. Hatakeyama, H. Nakagama, Y. Saito, S. Fukuda, T. Shibata, T. Yamada, Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer, Nat. Med. 25 (2019) 968–976.

[22] J. Diab, T. Hansen, R. Goll, H. Stenlund, E. Jensen, T. Moritz, J. Florholmen, G. Forsdahl, Mucosal metabolomic profiling and pathway analysis reveal the metabolic signature of ulcerative colitis, Metabolites 9 (2019) 291.

[23] N. Harada-Shoji, T. Soga, H. Tada, M. Miyashita, M. Harada, G. Watanabe, Y. Hamanaka, A. Sato, T. Suzuki, A. Suzuki, T. Ishida, A metabolic profile of routine needle biopsies identified tumor type specific metabolic signatures for breast cancer stratification: a pilot study, Metabolomics 15 (2019) 147.

[24] C. Lotti, J. Rubert, F. Fava, K. Tuohy, F. Mattivi, U. Vrhovsek, Development of a fast and cost-effective gas chromatography–mass spectrometry method for the quantification of short-chain and medium-chain fatty acids in human biofluids, Anal. Bioanal. Chem. 409 (2017) 5555–5567.

[25] T.L. Weir, D.K. Manter, A.M. Sheflin, B.A. Barnett, A.L. Heuberger, E.P. Ryan, Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults, PloS One 8 (2013), 70803.

Chapter 4

Distinct signatures of tumor-associated microbiota and metabolome in low-grade vs high-grade dysplastic colon polyps: inference of their role in tumor initiation and progression

According to the driver-passenger model for colorectal cancer[138], bacterial species have distinct temporal associations with colorectal tissues according to their role in tumor pathogenesis. We reasoned that driver bacteria, which are involved in CRC initiation and are predominant in the initial stages of tumorigenesis, would be enriched in low-grade dysplasia polyps, whereas passenger bacteria would mainly be found in advanced CRC stages.

In this manuscript, we have characterized the mucosa- and lumen-associated microbiota (MAM and LAM, respectively) and their metabolites in 78 patients with colon polyps, using our new technique that allows the collection of bacteria and metabolites from the tumor surface and maintains tissue integrity[153].

Our results show that MAM and LAM differ in composition and diversity, even at the phylum level. Analysis of mucosa-associated microbiome and metabolome in low- and high-grade dysplasia adenoma patients revealed different signatures that discriminate between these two patient groups. Importantly, we show that high-grade dysplastic adenomas are colonized by candidate passenger genera, whereas in low-grade dysplastic polyps we found enriched *Bacteroides fragilis,* that is one of the most studied driver bacteria since its enterotoxigenic strain can facilitate the initiation of pre-malignant lesions through the release of enterotoxins. These data support the hypothesis that bacteria are involved in CRC in a tumor-stage specific manner. Moreover, we revealed that differences in metabolite profiles accompanied the changes in bacterial composition associated to tumor stage.

Finally, we integrated microbiota and metabolome data, and identified some positive and negative correlations among metabolite classes and bacterial genera. Integrated microbiota-metabolome analysis suggests the involvement of the gut microbiota in the production and consumption of these metabolites, supporting the involvement of not only MAM, but also its metabolome, in CRC initiation and progression[154].

cancers

MDPI

Article

Distinct Signatures of Tumor-Associated Microbiota and Metabolome in Low-Grade vs. High-Grade Dysplastic Colon Polyps: Inference of Their Role in Tumor Initiation and Progression

Michela Giulia Clavenna [1,†], Marta La Vecchia [1,†], Marika Sculco [1], Soni Joseph [1], Elettra Barberis [2,3], Elia Amede [2,3], Marta Mellai [1,3], Silvia Brossa [1], Giulia Borgonovi [1], Pietro Occhipinti [4], Renzo Boldorini [1], Elisa Robotti [5], Barbara Azzimonti [1,3], Elisa Bona [6], Edoardo Pasolli [7,8], Daniela Ferrante [2], Marcello Manfredi [2,3], Anna Aspesi [1,‡] and Irma Dianzani [1,*,‡]

[1] Department of Health Sciences, Università del Piemonte Orientale, 28100 Novara, Italy; michela.clavenna@uniupo.it (M.G.C.); marta.lavecchia@uniupo.it (M.L.V.); marika.sculco@uniupo.it (M.S.); soni.joseph@uniupo.it (S.J.); marta.mellai@uniupo.it (M.M.); renzo.boldorini@med.uniupo.it (R.B.); barbara.azzimonti@med.uniupo.it (B.A.); anna.aspesi@med.uniupo.it (A.A.)

[2] Department of Translational Medicine, Università del Piemonte Orientale, 28100 Novara, Italy; elettra.barberis@uniupo.it (E.B.); elia.amede@uniupo.it (E.A.); daniela.ferrante@med.uniupo.it (D.F.); marcello.manfredi@uniupo.it (M.M.)

[3] Center for Translational Research on Autoimmune and Allergic Diseases, University of Piemonte Orientale, 28100 Novara, Italy

[4] Department of Gastroenterology, "Maggiore della Carità" Hospital, 28100 Novara, Italy

[5] Department of Sciences and Technological Innovation, Università del Piemonte Orientale, 15121 Alessandria, Italy; elisa.robotti@uniupo.it

[6] Department for Sustainable Development and Ecological Transition, Università del Piemonte Orientale, 13100 Vercelli, Italy; elisa.bona@uniupo.it

[7] Department of Agricultural Sciences, University of Naples Federico II, 80055 Portici, Italy; edoardo.pasolli@unina.it

[8] Task Force on Microbiome Studies, University of Naples Federico II, 80055 Portici, Italy

* Correspondence: irma.dianzani@med.uniupo.it

† These authors contributed equally to this work.

‡ These authors are joint senior authors.

check for updates

Citation: Clavenna, M.G.; La Vecchia, M.; Sculco, M.; Joseph, S.; Barberis, E.; Amede, E.; Mellai, M.; Brossa, S.; Borgonovi, G.; Occhipinti, P.; et al. Distinct Signatures of Tumor-Associated Microbiota and Metabolome in Low-Grade vs. High-Grade Dysplastic Colon Polyps: Inference of Their Role in Tumor Initiation and Progression. *Cancers* **2023**, *15*, 3065. https://doi.org/10.3390/cancers15123065

Academic Editor: M. Canan Arkan

Received: 20 April 2023
Revised: 31 May 2023
Accepted: 1 June 2023
Published: 6 June 2023

Simple Summary: A growing body of research has shown the connection between gut microbiota and colorectal cancer. However, most studies analyze fecal microbiota, which do not reliably represent the bacterial populations associated with colon mucosa. We analyzed the microbiota and metabolome directly collected from the surface of colon polyps, and showed different bacterial and metabolite signatures that discriminate between patients with low- and high-grade dysplastic polyps. We identified bacterial genera and species that are enriched in the early stages of tumor development and may act as drivers of carcinogenesis. Moreover, we revealed that differences in metabolite profiles accompanied the changes in bacterial composition associated with tumor stage, and that gut bacteria are involved in the production and consumption of significantly altered metabolites. Our findings pave the way for future mechanistic investigations to elucidate the role of specific bacteria in colon carcinogenesis and to design preventative measures based on microbiota modulation.

Abstract: According to the driver–passenger model for colorectal cancer (CRC), the tumor-associated microbiota is a dynamic ecosystem of bacterial species where bacteria with carcinogenic features linked to CRC initiation are defined as "drivers", while opportunistic bacteria colonizing more advanced tumor stages are known as "passengers". We reasoned that also gut microbiota-associated metabolites may be differentially enriched according to tumor stage, and be potential determinants of CRC development. Thus, we characterized the mucosa- and lumen-associated microbiota (MAM and LAM, respectively) and mucosa-associated metabolites in low- vs. high-grade dysplastic colon polyps from 78 patients. We show that MAM, obtained with a new biopsy-preserving approach, and LAM differ in composition and α/β-diversity. By stratifying patients for polyp histology, we found

that bacteria proposed as passengers by previous studies colonized high-grade dysplastic adenomas, whereas driver taxa were enriched in low-grade polyps. Furthermore, we report altered "mucosa-associated metabolite" levels in low- vs. high-grade groups. Integrated microbiota-metabolome analysis suggests the involvement of the gut microbiota in the production and consumption of these metabolites. Altogether, our findings support the involvement of bacterial species and associated metabolites in CRC mucosal homeostasis in a tumor-stage-specific manner. These distinct signatures may be used to distinguish low-grade from high-grade dysplastic polyps.

Keywords: mucosa-associated microbiota; lumen-associated microbiota; microbiota-derived metabolites; gut; colon polyp; colorectal cancer; driver bacteria; passenger bacteria

1. Introduction

Colorectal cancer (CRC) is among the most prevalent cancers worldwide, with more than 1.9 million cases and 935,000 deaths in 2020 [1]. Its multifactorial etiology includes genetic, environmental, and life-style factors. In most cases, CRC develops from adenomatous polyps, which show different grades of dysplasia during tumor progression.

Mounting evidence indicates that changes in the gut microbiota play an important role in colon carcinogenesis [2].

The bacterial driver–passenger model for CRC proposes that bacterial species have distinct temporal associations with colorectal tissues according to their role in CRC pathogenesis [3–5]. Driver bacteria are found in the initial stages of carcinogenesis and are, therefore, thought to play a role in CRC initiation by different mechanisms: production of genotoxic substances, disruption of the function of tumor suppressor proteins, such as E-cadherin, production of metabolites that increase the proliferation of enterocytes and opportunistic microbial pathogens, and induction and maintenance of the inflammatory process [3,4,6]. Passenger bacteria are instead opportunistic pathogens involved in CRC progression [3,7]. One of the most studied driver bacteria is the enterotoxigenic strain of *Bacteroides fragilis* (ETBF) that can facilitate the initiation of pre-malignant lesions through the release of enterotoxins (BFTs) [3]. The epithelial response to *B. fragilis* toxins induces E-cadherin cleavage, resulting in enhanced barrier permeability, Wnt/β-catenin, and NF-κB signaling [8]. Importantly, tumor-susceptible mice ($Apc^{Min/+}$) colonized by ETBF strains are used as a model of microbial-induced colon tumorigenesis [9]. Another example of driver bacteria is represented by *Escherichia coli* strains harboring the polyketide synthase (PKS) genomic island and capable of synthesizing the genotoxic virulence factor colibactin, which induces double-strand DNA breaks [10].

Given that the tumor microenvironment (TME) changes during the oncogenic process, pathogenic driver bacteria can be numerically overwhelmed and gradually replaced by passenger bacteria that acquire a growth advantage in the tumor context. These opportunistic pathogens, which otherwise would not be able to colonize healthy colorectal tissues, exploit the altered metabolism of the tumor colonocytes to proliferate [3,4]. Passenger bacteria may also be actively involved in cancer progression, though their relevance is still unclear [3,4,11].

Another mechanism through which microbiota can influence CRC development is the production of various metabolites [12,13]. It is, in fact, widely accepted that microbiota-derived metabolites play a crucial role in host physiology and disease development, and that their abundances may vary according to tumor stage. Among the most heavily studied microbiota-derived metabolites in CRC, are short-chain fatty acids (SCFAs) and bile acids (BAs) [14]. In particular, deoxycholic acid, which is a secondary bile acid, can cause inflammation and promote intestinal tumorigenesis in $Apc^{Min/+}$ mice [15]. Bile acid is involved in tumor progression through the activation of the NF-κB pathway, which promotes cell growth and survival [16]. Moreover, gut-associated metabolites may directly

alter the gut microbiota composition by promoting the proliferation of specific bacteria, in particular that of passengers [17].

A number of studies have performed integrated analyses on lumen-associated microbiota (LAM) and metabolome from fecal or serum samples to characterize the role of bacteria and associated metabolites in the pathogenesis of CRC [18–20]. However, data are scant on gut mucosa-associated microbiota (MAM) and mucosa-associated metabolome.

The emerging consensus is that the composition of gut MAM differs from that of LAM, with only a few species being present in both compartments [21–25].

Based on the aforementioned evidence, the aim of this study was to characterize the temporal association of MAM and their metabolites with low-grade vs. high-grade dysplastic colon polyps—which represent two distinct stages of the adenoma–carcinoma sequence—and to ascertain their role in neoplastic development. For this purpose, we have devised a new technique that allows the collection of bacteria and metabolites from the adenoma's surface without jeopardizing tissue integrity.

The analysis of MAM and mucosal-associated metabolome, according to the histological classification of colon polyps, reveals the preponderance of potential driver bacteria in low-grade dysplastic polyps, while potential passenger bacteria are enriched in high-grade dysplastic ones. The bacteria classification is based on previous reports, where candidate drivers or passengers were proposed [3,4], albeit no functional experiments were performed to show that these bacteria play a direct role in malignancy. We also report differences in the metabolite relative abundances between the two study groups, suggesting that these signatures may be used to distinguish colon polyps according to histology. Finally, integrated analysis of MAM and metabolites shows either positive or negative correlations between enriched bacteria and specific classes of metabolites, supporting the involvement of the gut microbiota in the production and consumption of these metabolites.

2. Materials and Methods

2.1. Patients Enrollment

Patients (n = 78, males = 45; females = 33) were recruited before colonoscopy at the Gastroenterology Unit of Maggiore della Carità University Hospital (Novara, Italy). All the patients undergoing colonoscopy signed an informed consent form. After colonoscopy, only patients with polyps larger than 10 mm, older than 18 years were included in this study. The other exclusion criteria were prebiotic, probiotic or antibiotic consumption within one month before fecal sample (LAM) collection. Previous gastro-intestinal conditions that could modify the gut microbiota were evaluated, including diverticula, cholecystectomy and previous polyp occurrence and reported in Tables 1 and S1. All the patients used laxatives before colonoscopy, as required by the procedure. A team that includes 13 medical operators of the same Unit, all using the same working procedures, performed all the colonoscopies, and a single nurse collected all the microbiome and the metabolome samples.

2.2. Sample Collection

To collect MAM and associated metabolites, e-NAT™ (COPAN, Brescia, Italy) swabs or dry swabs, respectively, were used to gently brush the polyp surfaces without compromising their tissue integrity. Samples were stored at −80 °C until 16S rRNA tag sequencing and metabolite extraction. Bowel preparation could alter gut microbiota and metabolome composition [26]. Nagata and colleagues in 2019 showed that after 14 days, both microbiota and metabolome are completely restored [27]. Therefore, for LAM analyses, fecal samples were collected from the patients 14 days after colonoscopy, aliquoted and stored at −80 °C until microbial DNA isolation and sequencing.

Table 1. Clinical features of the patients. Statistically significant *p*-values ($p < 0.05$) are in bold.

Clinical Features	Patients n = 78 (%)	Patients with Low-Grade Dysplastic Polyps n = 44 (%)	Patients with High-Grade Dysplastic Polyps n = 34 (%)	*p*-Value
Gender				
Female	33 (42.3%)	21 (47.7%)	12 (35.3%)	0.3
Male	45 (57.7%)	23 (52.3%)	22 (64.7%)	
BMI (body mass index)				
Normal weight	34 (43.6%)	20 (45.4%)	14 (41.2%)	0.7
Overweight or obese	44 (56.4%)	24 (54.6%)	20 (58.8%)	
Age				
Median (IQR)	61 (58–70)	61 (58–68)	62 (56–70)	0.9
Polyp size mm				
Median (IQR)	14 (10–23)	12 (10–16)	15 (12–25)	**0.02**
Type of polyp				
Tubular	28 (35.9%)	21 (47.7%)	7 (20.6%)	**0.005**
Villous	3 (3.8%)	1 (2.3%)	2 (5.9%)	
Tubulo-villous	40 (51.3%)	16 (36.4%)	24 (70.6%)	
Others	7 (9.0%)	6 (13.6%)	1 (2.9%)	
Previous gastrointestinal conditions				
Diverticulitis	26 (33.3%)	19 (43.2%)	7 (20.6%)	0.2
Previous polyp occurrence	8 (10.2%)	6 (13.6%)	2 (5.9%)	
IBD	1 (1.3%)	0	1 (2.9%)	
Previous cholecystectomy	5 (6.4%)	4 (9.1%)	1 (2.9%)	
Slight mucosal inflammation	1 (1.3%)	0	1 (2.9%)	
Polyp localization				
Right colon	18 (23.1%)	11 (25.0%)	7 (20.6%)	0.8
Left colon	52 (66.7%)	28 (63.6%)	24 (70.6%)	
Transversal colon	8 (10.2%)	5 (11.4%)	3 (8.8%)	

BMI: body mass index; IQR: interquartile range; IBD: inflammatory bowel disease.

Patient nutritional habits were evaluated with the validated European Prospective Investigation into Cancer and nutrition (EPIC) questionnaire on nutrition [28]. The questionnaire is composed of 16 categories and questions about 266 different items, including simple foods and recipes, to understand the food frequency intake. The questionnaire was completed online and analyzed, and the intake frequency was transformed in grams/day. We analyzed and compared the consumption of the most important nutrients in low- vs. high-grade dysplasia groups (Table S2).

2.3. Histology

After removal, polyps were included in neutral buffered formalin for at least 24 h and then included in paraffin. Sections were cut at 4-µm thickness and stained using hematoxylin-eosin. Polyps were oriented using a stereo microscope and cut alongside the major axis, identifying, if possible, the base implant. All polyps (i.e., tubular, villous, tubulovillous, sessile-serrated) were evaluated by an expert pathologist (R.B.) at the University Hospital Pathology Unit in Novara, Italy. Patients with low-grade dysplastic adenomas were included in the "low-grade" group, while patients with high-grade dysplastic adenomas were included in the "high-grade" group (Table 1).

2.4. MAM and LAM Analyses

Microbial DNA for MAM analyses were extracted from e-NAT™ swabs with QIAamp® DNA Microbiome kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

LAM analyses were performed on microbial DNA extracted from fecal samples using QIAamp® PowerFecal® Pro DNA kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. The yield and quality of microbial DNA was determined using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The quantity was assessed with Invitrogen™ Qubit™ 1X dsDNA HS Assay Kit (Invitrogen Co., Thermo Fisher Scientific Inc.) using a Qubit 4 fluorometer (Invitrogen).

To avoid contaminations, microbial DNA extraction for MAM and LAM was performed in sterile conditions, using a laminar flow cabinet and sterile reagents and materials. E-NAT™ swabs not brushed on any tissue were used as negative controls for microbial DNA extraction and 16S rRNA sequencing.

MAM and LAM samples were subjected to 16S rRNA amplicon sequencing analysis using Microbiota Solution B Kit, a next-generation sequencing (NGS) in vitro molecular test, CE-IVD marked (Arrow Diagnostics Srl, Genoa, Italy). Polymerase chain reaction (PCR) amplification of the V3-V4-V6 hypervariable regions of bacterial 16S rRNA was obtained by using the patented degenerate primer sets within the Arrow Microbiota Solution B kit (cod. AD-002.024), according to the manufacturer's instructions. PCR products were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter Inc., Brea, CA, USA), and indexes were added in a subsequent step. The hypervariable V3-V4-V6 regions of the bacterial 16S rRNA were amplified according to the manufacturer's instructions.

The DNA concentration of the libraries was fluorometrically measured and samples were pooled in equimolar concentrations. The final 16S rRNA amplicon libraries were sequenced on a MiSeq Illumina® sequencing platform (Illumina, San Diego, CA, USA) using a MiSeq Reagent Nano Kit v2 cartridge for a 2 × 250 paired-end sequencing.

2.5. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt)

Functional abundances were predicted using the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt2) software 2.0 [29]. Pathways differentially abundant between low- and high-grade dysplastic polyps were detected using the STAMP software [30]. Pathways with p-value < 0.05 were identified as significant after false discovery rate (FDR) correction. We consulted the MetaCyc website (https://metacyc.org/, accessed on 26 September 2022) to identify the products of each pathway which emerged from the PICRUSt analysis.

2.6. Mucosa-Associated Metabolome

Small molecules were extracted and analyzed as reported in our previous validated method [31]. Briefly, short chain fatty acid (SCFA) extraction from dry swabs was performed first using water and sonication and then liquid-liquid extraction with methyl tert-butyl ether (MTBE). Methanol-isopropanol-acetonitrile was then used to extract other metabolites (i.e., amino acids, sugars, long fatty acids, and medium fatty acids) from the aqueous phase. The internal standards deuterated propanoic acid (1 ppm), tridecanoic acid (0.5 ppm) and hexadecane (1 ppm) were also added. SCFAs and small molecules were analyzed by bidimensional gas chromatography mass spectrometry GCXGC/TOFMS (BT 4D, Leco Corp., St. Josef, MI, USA), as described in our previous work [31]. The samples were analyzed using both targeted and untargeted approaches. Briefly, SCFAs were quantified using a targeted analysis performed with internal standards and external calibration curves, as previously reported [31]. For the untargeted analysis, peaks with signal-to-noise (S/N) value lower than 500.0 were rejected. ChromaTOF version 5.31 was used for raw data processing and mass spectral assignment was performed by matching with NIST MS Search 2.3 libraries adding Fiehn Library. Identification of molecules was also performed using an in-house library built with commercial mix standards that contain hundreds of molecules. As the polyp mean area was different between the low- and high-grade dysplasia groups (median (IQR) 12 (10–16) mm vs. 15 (12–25) mm; p-value $\leq$ 0.05), normalization was performed by dividing the metabolites' abundances by the value of the area of each analyzed polyp, with the limit of the type and shape of the polyps. Measurements were

performed at the time of colonoscopy with graph paper. The internal standards that were spiked in each sample, were used for instrument stability monitoring and data normalization. In addition, small molecule levels from untargeted analysis were also normalized by total sum of abundances. To study a possible correlation between polyp-associated microbiota and its metabolites, the metabolome analysis was integrated with MAM using M^2IA, an open-source web server. The hierarchical clustering heat map analysis was performed through MetaboAnalyst software 5.0 (www.metaboanalyst.ca, accessed on 17 December 2021) using the Euclidean distance as distance measure and the Ward method as clustering method. Only modulated metabolites (p-value < 0.05 and fold change > 1.3 or <0.769) were used. Metabolomics data are shown in Table S3.

2.7. Raw Sequence Processing

Raw sequences obtained from MAM and LAM DNA were processed using the software MicrobAT Suite v1.2.1 (SmartSeq srl, Novara, Italy), based on the Ribosomal Database Project (RDP) database. MicrobAT (SmartSeq s.r.l.) is a standalone software based on client/server system. Through a graphical interface developed in Java, the user can load the FASTQ files, download the raw data of the analysis and print the reports of the samples. The first step is a cleaning of the reads obtained from the FASTQ file using custom algorithms that remove the short sequences (read length < 200 nt) and sequences with a low quality (average Phred quality score [32] < 25). High-quality sequences are then aligned with the reference database, i.e., RDP database release 11-update 5 [33]. During this taxonomic assignment process, only the reads with minimum sequence length that align with reference ≥80% and similarity threshold ≥97% were associated, by the analysis system, with the species taxonomic level. Finally, the software generates absolute abundance tables and three files (OTU, taxonomy, metadata) used as input for the subsequent analyses [34,35].

Statistical analysis regarding variations within the bacterial communities was performed using MicrobiomeAnalyst software 1.0 (Comprehensive Statistical, Visual, and Meta-Analysis of Microbiome data) [36].

Firstly, a data integrity check was performed by the online software to show the information collected. Secondly, taxa having zero reads across all the samples or appearing in only one sample were removed by default. Finally, a low-count filter was applied to remove taxa containing less than 30 (LAM vs. MAM comparison) or 10 reads (low- vs. high-grade dysplasia comparison) in at least 20% of samples.

2.8. Statistical Analysis

Fisher's exact test was used to compare the groups, as reported in Table 1.

Heat tree analysis was used to compare statistically significant differences between the groups, i.e., MAM vs. LAM, low- vs. high-grade dysplastic polyps. This method, performed through R metacoder package [37], uses hierarchical structure of taxonomic classifications to quantitatively (median abundance) and statistically (non-parametric Wilcoxon Rank Sum test) depict taxon differences among communities, using color and size of nodes.

To compare MAM with LAM, data were summarized using α- or β-diversity indexes. Three α-diversity metrics were used: the observed number, the Shannon index, and the Simpson index. The first index evaluates the number of unique taxa observed in each sample, considering only richness. The last two are based on not only richness but also evenness, which represents the abundance of a given microorganism. We performed α-diversity analysis using the phyloseq package [38], and results were plotted across samples and depicted as box plots for each group.

Beta diversity analysis, used to compare the different composition between the analysis groups, was calculated by Bray–Curtis distance, and the results were visualized in two plots through principal coordinate analysis (PCoA). In the plots, each point represents the entire microbiome of a single sample. The statistical significance of the differences in

β-diversity between groups (MAM vs. LAM; low- vs. high-grade dysplastic polyps) was evaluated using permutational ANOVA (PERMANOVA).

Linear discriminant analysis effect size (LDA-LEfSe) was used to identify signatures at different taxonomic levels, characterizing each different group (MAM vs. LAM, low- vs. high-grade dysplastic polyps). This method estimates both statistical significance and biological consistency (effect size). Firstly, it uses the Kruskal–Wallis sum-rank test to identify taxa that are statistically different between groups. Subsequently, LEfSe applies LDA to calculate the effect size of each differentially abundant feature. Features with $p < 0.05$ and an LDA score > or <2 were considered taxa able to discriminate between the two groups. For mucosal-adherent and luminal microbiota comparison, the false discovery rate (FDR) was used to correct for multiple testing, and taxa with p-values < 0.05 were considered statistically significant. The p-values adjusted for the FDR are indicated as q-values.

For the correlation analyses between polyp-associated microbiota and metabolites, Spearman's rank correlation coefficient was calculated using the M^2IA web server with default settings.

3. Results

3.1. Characterization of LAM and MAM in Patients with Colon Polyps

Contrary to other sample collection protocols that cause the degradation of the biopsies to obtain microbiota and metabolites, we have developed a new approach that allows us to analyze the microbiota and metabolome adherent to the polyp surface without compromising the integrity of the biopsies, a key requisite to perform an accurate histological analysis.

Seventy-eight patients (45 males, 33 females) with polyps larger than 10 mm were recruited before colonoscopy at the Gastroenterology Unit of the University Hospital Maggiore della Carità in Novara, Italy. The clinical features of this study population are shown in Table 1. We did not find any statistically significant difference between low- and high-grade groups regarding previous gastro-intestinal conditions and polyp localization (Table 1). Since diet can influence the gut microbiota composition, we compared nutrient consumption between patients with low- and high-grade dysplastic polyps. We did not find any statistically significant difference in the daily consumption of fiber, lipids, red and processed meat, fruit and vegetables, as shown in Table S2.

MAM samples were collected by gently brushing the surface of the resected polyp with an e-NAT™ swab, which allows the preservation of the nucleic acids until extraction, whereas LAM-containing specimens were isolated from feces using a standard approach (see Methods). Subsequently, MAM and LAM samples were subjected to 16S rRNA sequencing, yielding an average number of reads of 53,028.73 and 67,479.96, respectively. After applying a low-count filter to remove taxa showing less than 10 reads, we obtained 165 taxa from the MAM samples and 202 from the LAM ones. These genomic sequences were included in the BioProject MIMEC Project_Swab PRJNA783496 and MIMEC Project_Fecal PRJNA783535 available in the NCBI database https://submit.ncbi.nlm.nih.gov/subs/sra/SUB11427238/overview, accessed on 17 December 2021 and https://submit.ncbi.nlm.nih.gov/subs/sra/SUB11420448/overview, accessed on 17 December 2021, respectively. The α-diversity indexes—which include the observed number (Figure 1a), the Shannon (Figure 1b) and the Simpson (Figure 1c) indexes—show that LAM is characterized by a significantly higher mean species diversity than that of MAM ($p < 0.05$).

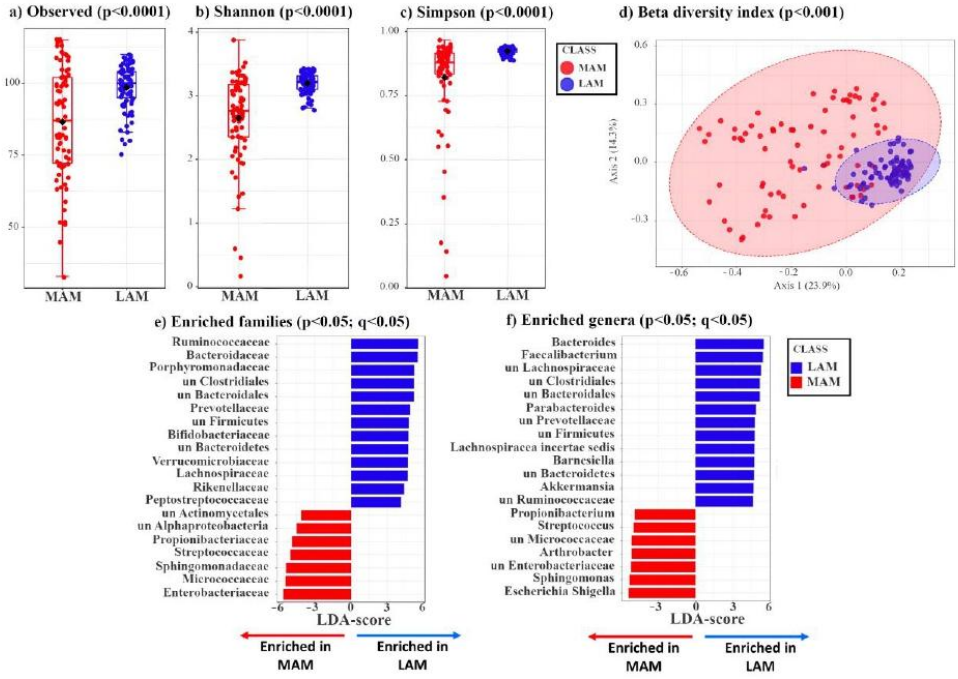


Figure 1. Box plots showing three different α-diversity indexes tested by Wilcoxon rank sum test. (**a**) The observed index is a richness-based measure showing the number of unique taxa present in each sample. The box plot shows the difference in observed species between MAM (red) and LAM (blue) samples ($p = 1.17 \times 10^{-4}$). (**b**,**c**) Shannon and Simpson indexes are based on richness (i.e., numbers of species) and evenness (i.e., abundance of microorganisms). The two different box plots show significant differences between MAM (red) and LAM (blue) samples (Shannon: $p = 3.49 \times 10^{-10}$, Simpson: $p = 1.05 \times 10^{-11}$). (**d**) β-diversity analysis shows a significant separation between MAM (red) and LAM (blue) samples ($p < 0.001$). Principal coordinates analysis (PCoA), based on Bray–Curtis distance matrix, shows the different microbial composition between the two groups. This is achieved by comparing the changes in presence/absence or abundance of thousands of species and by summarizing how "similar" or "dissimilar" they are. The X-axis explains 23.9% of the variability between samples, while the Y-axis explains 14.3%. The statistical significance of the clustering pattern was calculated using permutational ANOVA (PERMANOVA). (**e**,**f**) Linear discriminant analysis effect size (LDA-LEfSe) showing families (**e**) and genera (**f**) enriched in LAM (blue, on the right, LDA-score > 3) or MAM (red, on the left, LDA-score < −3).

Next, we assessed the β-diversity indexes of MAM vs. LAM by principal coordinates analysis (PCoA) using the Bray–Curtis distance matrix. As shown in Figure 1d, LAM displays a tighter clustering compared to that of MAM. The statistical significance of the clustering pattern was confirmed by permutational ANOVA (PERMANOVA) ($p < 0.001$, Figure 1d).

Compared to MAM, LAM shows a phylum enrichment of Firmicutes (Bacillota) (51.30% vs. 39.03%), Bacteroidetes (Bacteroidota) (22.03% vs. 7.30%), and Verrucomicrobia (Verrucomicrobiota) (1.72% vs. 1.00%) ($p < 0.05$) (Figure S1a,b). Conversely, MAM displays a phylum enrichment of Proteobacteria (Pseudomonadota) (15.91% vs. 2.52%) and Actinobacteria (Actinomycetota) (11.79% vs. 6.27%) (Figure S1a,b). The two groups (MAM vs. LAM) show statistically significant differences at phylum, class, order (Figure S1c–e), and at family and genus level (Figure 1e,f).

LDA LEfSe analysis was carried out to identify the bacterial genera that were enriched in the LAM or MAM samples (Table S4). In agreement with the phylum results, all genera included in the Bacteroidetes (Bacteroidota) and Verrucomicrobia phyla and most genera included in the Firmicutes phylum (Bacillota) are significantly enriched ($p < 0.05$) in LAM compared with MAM, whereas most genera included in Proteobacteria (Pseudomonadota) and Actinobacteria (Actinomycetota) are more abundant in MAM samples (Table S4).

Overall, these data suggest that patients' MAM differs from LAM, possibly because MAM is more related to the localized changes occurring near the polyps and more dependent of tumor stage, while LAM is representative of all bacterial species present in the gut. Thus, even though LAM has a higher number of species (α-diversity), most of these latter are shared among patients regardless of tumor stage.

3.2. Identification of Mucosa-Associated Bacterial Signatures Distinguishing Low-Grade from High-Grade Dysplastic Colorectal Polyps

Since bacteria in close contact with enterocytes may play an important role in colon carcinogenesis, we focused our attention on polyp-associated microbiota. In particular, we asked whether the MAM characterizing low-grade dysplastic polyps would be enriched in driver bacteria, which influence the initial stages of carcinogenesis, whereas MAM of high-grade dysplastic polyps would be enriched in passenger species.

By stratifying patients according to histology (low-grade vs. high-grade dysplasia), we identified two genera (*Pelomonas* and *Phascolarctobacterium*) enriched in low-grade, while the potential passenger genus *Anaerococcus* [4] was enriched in high-grade dysplastic polyps (Figure 2a). Moreover, we found the potential driver species, *Bacteroides fragilis* [3] and five other species (i.e., *Bacteroides* spp., *Beta proteobacterium*, unclassified *Phascolarctobacterium*, unclassified *Erysipelotrichaceae incertae sedis*, and *Phascolarctobacterium faecium*) enriched in low-grade dysplastic polyps ($p < 0.05$), whereas the two potential passenger species, unclassified *Anaerococcus* and *Streptococcus anginosus* [4], were enriched in high-grade dysplastic polyps (Figure 2b). Thus, as we hypothesized, known candidate driver taxa are only enriched in MAM of low-grade dysplastic polyps, while known candidate passenger taxa are only enriched in MAM of high-grade dysplastic polyps. It must be considered that candidate driver or passenger classification is based on previously suggested classifications, lacking validation by functional experiments.

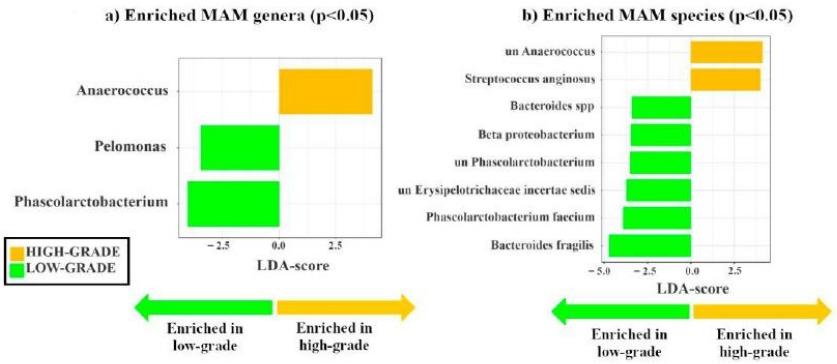


Figure 2. Linear discriminant analysis effect size (LDA-LEfSe) showing (**a**) bacterial genera and (**b**) species enriched in MAM high- (yellow, LDA score > 2) vs. low-grade dysplastic adenomas (green, LDA score < −2). This method incorporates statistical significance (Kruskal–Wallis) with biological consistency (effect size). The length of the bar represents a log10 transformed LDA score. This value is positive if the bacterial species is enriched in the first compared to the second group and negative if the second group shows enrichment compared to the first group. A significance level of $p < 0.05$ and an LDA score of 2 are used to determine the species best characterizing each phenotype.

Similar differences in MAM between low- and high-grade dysplastic polyps were observed by analyzing the phylogenetic heat tree (Figure S2).

It is noteworthy that the analysis of LAM also revealed different signatures for genera and species in high- and low- grade dysplastic groups (Figure S3).

3.3. Comparison of Mucosa-Associated Metabolome between High-Grade and Low-Grade Dysplastic Colorectal Polyps

Next, we asked whether there was an association between tumor stage and the composition of the mucosa-associated metabolome. To answer this question, we identified the metabolome adherent to the polyps through a high-throughput metabolomics approach recently described by our group [31]. Of note, metabolites in the gut can derive from bacteria, endogenous compounds, or exogenous dietary components [39–41].

Metabolome analysis of 59 (34 low- vs. 25 high-grade) out of 78 patients (19 samples were unavailable) uncovered 41 metabolites that allowed us to distinguish between high- and low-grade dysplastic polyps (fold change, FC > 1.3 enriched in high-grade or FC < 0.769 depleted in high-grade; $p < 0.05$). In high-grade polyps, we found a higher concentration of SCFAs, such as butyric acid (FC = 3.7; $p < 0.05$) and isobutyric acid (FC = 3.5; $p < 0.05$), lactic acid (FC = 1.9; $p < 0.05$), the nucleobase uracil (FC = 3.6; $p < 0.01$), and several amino acids, such as threonine (FC = 8.2; $p < 0.01$), serine (FC = 2.8; $p < 0.05$), α-aminobutanoic acid (FC = 5.6; $p < 0.05$). Other differentially abundant metabolites included erythronic acid (FC = 0.7; $p < 0.05$), L-threitol (FC = 1.9; $p < 0.05$), pyroglutamic acid (FC = 4.7; $p < 0.05$), and hydroquinone (FC = 2.8; $p < 0.01$) (Figures S4 and S5).

The hierarchical clustering heat map (Figures 3a and S6) shows the distribution of the metabolites that are statistically different between low- (green) and high-grade (yellow) dysplastic polyps. The partial least square discriminant analysis (PLS-DA) reported in Figure 3b shows the presence of a metabolic signature associated with low- (green) or high-grade (yellow) dysplastic polyps.

3.4. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) in Low-Grade vs. High-Grade

PICRUSt analysis was performed to predict the metabolic function of bacteria found differently enriched in the low- or the high-grade group. PICRUSt analysis on MAM identified 17 different pathways statistically enriched in low- or high-grade dysplastic polyps (4 enriched in low-grade and 13 enriched in high-grade). We show these pathways in Table 2. We consulted the MetaCyc website (https://metacyc.org/, accessed on 26 September 2022) to identify the products of each pathway which emerged from the PICRUSt analysis. In particular, in Figure S7, we show the mixed acid fermentation pathway and superpathway of the pyrimidine ribonucleosides salvage, which were enriched in the high-grade group.

3.5. Integration of MAM and Polyp-Adherent Metabolome Data

In order to investigate which bacterial taxa and small molecules/metabolite classes were mainly responsible for the overall associations with the histological grade of polyps, the individual correlations between genus level, bacterial abundance profile, class level, and individual metabolite level intensity profile were analyzed using the M2IA open-source web server.

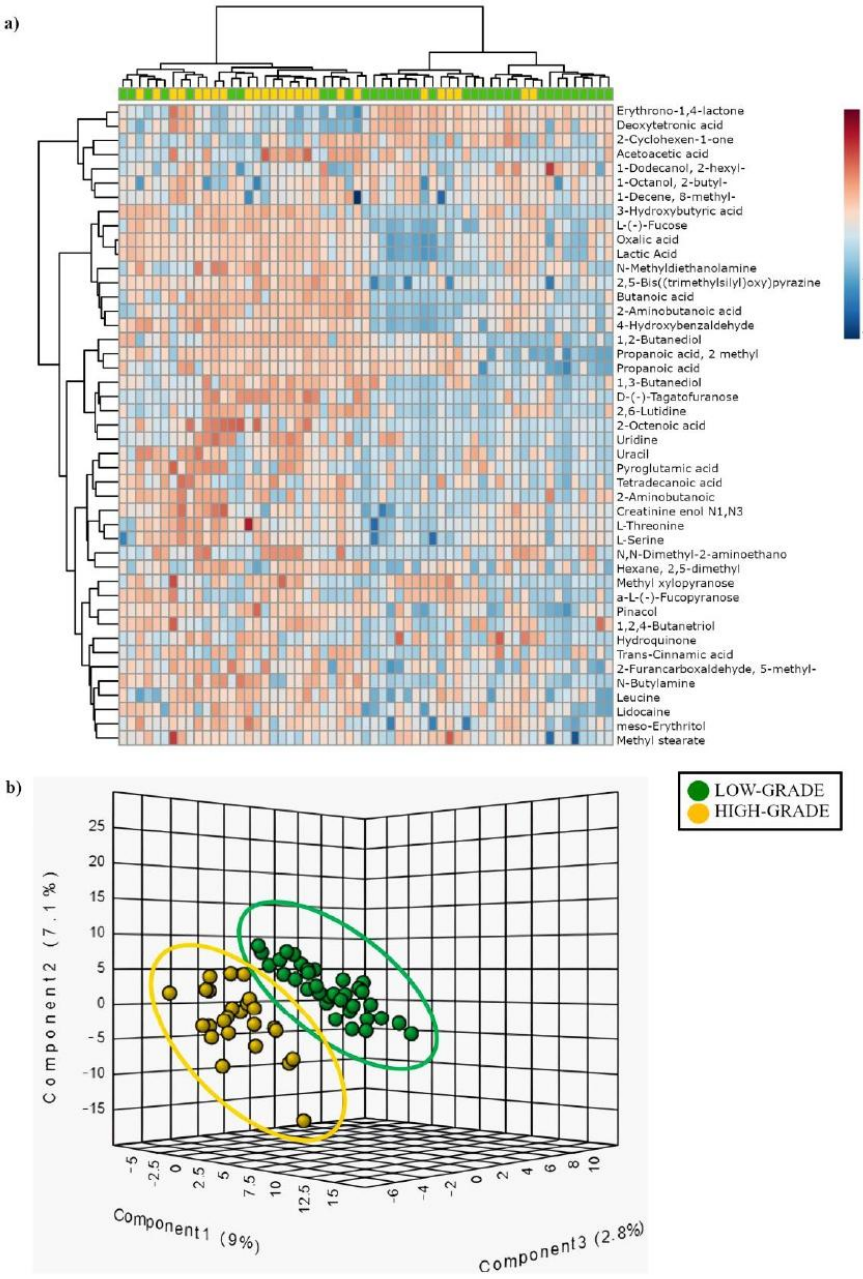


Figure 3. (**a**) Hierarchical clustering heat-map showing different metabolite distributions between low- (green) and high-grade (yellow) dysplastic polyps. All the metabolites listed show a statistically significant difference between low and high-grade dysplasia groups ($p < 0.05$). Higher concentrations are reported in red, while low levels are in blue (auto-scaled data). (**b**) Partial least square discriminant analysis (PLS-DA) showing different metabolite distribution between low- (green) and high-grade (yellow) dysplastic polyps.

Table 2. Pathways differently enriched in patients with low- (green) or high-grade (yellow) dysplastic polyps, obtained by PICRUSt analysis. * Pathway described in article Discussion.

Pathway	Description	*p*-Value
Superpathway of adenosylcobalamin salvage from cobinamide I (COBALSYN-PWY)	Vitamin biosynthesis	8.54×10^{-3}
Adenosylcobalamin biosynthesis from adenosylcobinamide-GDP I (PWY-5509)	Vitamin biosynthesis	0.016
Superpathway of adenosylcobalamin salvage from cobinamide II (PWY-6269)	Vitamin biosynthesis	2.78×10^{-3}
Sucrose degradation IV (sucrose phosphorylase) (PWY-5384)	Carbohydrate degradation	0.035
Mixed acid fermentation (FERMENTATION-PWY) *	Carbohydrate degradation	5.94×10^{-3}
Superpathway of tetrahydrofolate biosynthesis and salvage (FOLSYN-PWY)	Vitamin biosynthesis	0.036
Superpathway of tetrahydrofolate biosynthesis (PWY-6612)	Vitamin biosynthesis	0.034
Superpathway of thiamine diphosphate biosynthesis II (PWY-6895)	Vitamin biosynthesis	0.034
Superpathway of purine nucleotides de novo biosynthesis II (DENOVOPURINE2-PWY)	Nucleotides synthesis	6.73×10^{-3}
Superpathway of guanosine nucleotides de novo biosynthesis II (PWY-6125)	Nucleotides synthesis	0.028
Superpathway of pyrimidine ribonucleosides salvage (PWY-7196) *	Nucleotides synthesis	7.02×10^{-3}
Superpathway of guanosine nucleotides de novo biosynthesis I (PWY-7228)	Nucleotides synthesis	0.036
Superpathway of purine nucleotides de novo biosynthesis I (PWY-841)	Nucleotides synthesis	0.038
Superpathway of pyrimidine ribonucleotides de novo biosynthesis (PWY0-162)	Nucleotides synthesis	0.031
Incomplete reductive TCA cycle (P42-PWY)	Reductive TCA cycle	0.017
PreQ0 biosynthesis (PWY-6703)	Secondary metabolite biosynthesis	0.049
Pyrimidine deoxyribonucleotides de novo biosynthesis II (PWY-7187)	Nucleoside and nucleotide synthesis	0.019

As shown in Figure 4, PLS-DA revealed the presence of specific microbiota and metabolic signatures associated with low- (green) and high-grade dysplasia (yellow).

The correlations between modulated genera, significant metabolite classes, and their relative individual metabolites (FC > 1.3 or <0.769, $p < 0.05$) were performed using 16S sequencing and metabolomic data obtained from the analysis of low- and high-grade polyps. Spearman's rank correlations were calculated between the relative concentration of metabolite classes and the abundance of bacterial taxonomic groups. More than two hundred significant bacteria-metabolite class correlations were identified at the genus level. Fifty-six of these were positive correlations, while 158 were negative correlations (Figure 5). Aromatic compounds were negatively correlated with *Pelomonas*, *Phascolarctobacterium*, and *Bacteroides*, to which *B. fragilis* belongs (Figure 5). *Pelomonas* and *Phascolarctobacterium* were also negatively correlated with organonitrogen compounds.

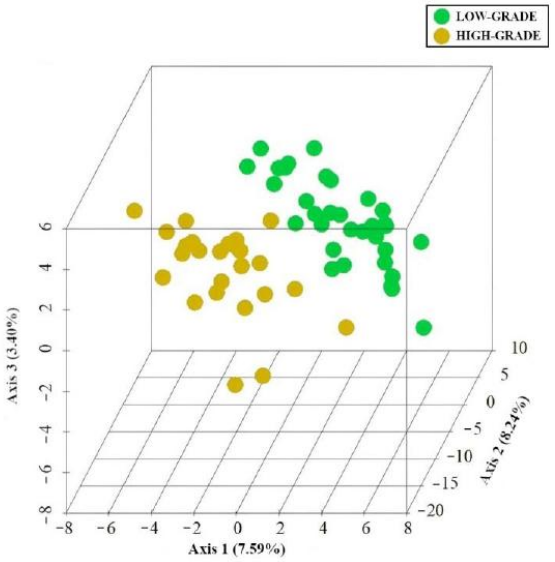


Figure 4. PLS-DA. Specific MAM and mucosa-associated metabolite signatures. PLS-DA analysis of low-grade (green) and high-grade dysplasia (yellow). The three-dimensional score plot indicates the degree of group differences using the first three principal components. Each dot indicates a single sample.

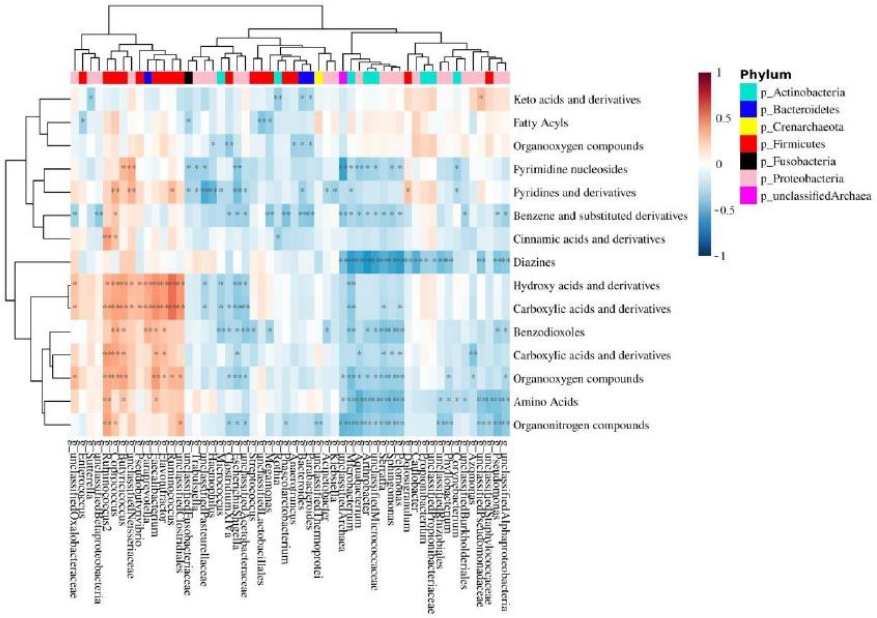


Figure 5. Correlation heat map between MAM genus and mucosa-associated metabolite class. The red color indicates a positive correlation, while the blue color indicates a negative correlation. The darker color indicates the most highly correlated variables. The asterisks indicate the correlation coefficient *p*-value: * $p < 0.05$ and ** $p < 0.01$.

4. Discussion

In this study, we have analyzed 78 colon polyp patients with the aim of correlating the composition of their gut microbiota and associated metabolome with tumor development. To this end, we devised a novel sampling strategy that enables the collection of mucosa-associated microbiota and metabolome without jeopardizing tumor integrity. By integrating these data, we identified bacteria and metabolites involved in colorectal cancer in a tumor-stage-specific manner.

It is well known that the intestinal microbiota may compromise the mucosal barrier, cross the epithelium, and interact with immune cells, causing local inflammation, cancer induction and progression [2]. Thus, colon cancer microbiota has been generally characterized by using samples collected during surgery to look for bacteria that infiltrate the tumor and shape its microenvironment. We reasoned that this approach would identify mostly passenger bacteria, including not only bacteria with a role in cancer progression, but also those that simply thrive in the cancer microenvironment. Because we wanted instead to identify driver bacteria, we decided to focus on colon adenomas, i.e., benign tumors that have just began the adenoma–carcinoma sequence. Since adenomas are usually excised at colonoscopy to undergo the necessary histological analyses and diagnostic procedures, we devised a strategy that preserved tumor integrity. Our aim was, thus, to investigate whether the surface of intestinal adenomas hosts bacteria that influence cell transformation, given that the bacterial species and metabolites in contact with enterocytes may play an important role in colon carcinogenesis [42,43]. A drawback of this approach is that we cannot compare our samples with healthy neighboring mucosa, since the healthy mucosa is not removed during colonoscopy and therefore is not available for the collection of microbiota and metabolites by brushing.

By comparing the composition of MAM with that of LAM—the former obtained from swabs brushed against the polyp surface, while the latter isolated from fecal samples—we show that the α-diversity indexes (i.e., observed, Shannon, and Simpson) of LAM are significantly higher than those of MAM (Figure 1a–c), and that LAM displays tighter clustering compared to MAM (Figure 1d), in good agreement with previous studies on biopsies from healthy individuals [22,24,44,45].

Despite having higher diversity levels, LAM appears to be more homogeneous than MAM, as shown in Figure 1d ($p < 0.001$). Indeed, LAM displays enrichment of the phyla Firmicutes (Bacillota), Bacteroidetes (Bacteroidota), and Verrucomicrobia, while MAM mainly consists of Proteobacteria (Pseudomonadota) and Actinobacteria (Actinomycetota) (Figure S1). Overall, we found 49 genera that were increased in LAM or MAM samples ($q < 0.05$) (Table S4), consistent with our data at the phylum level. Our findings are also in good agreement with a study by Tang and colleagues [45] showing that, among individuals without gastrointestinal symptoms undergoing routine screening colonoscopies, the phyla Firmicutes (Bacillota) and Bacteroidetes (Bacteroidota) were enriched in LAM, whereas Proteobacteria (Pseudomonadota) were more abundant in biopsy samples. Moreover, analyzing healthy subjects, Ringel and colleagues found that LAM was enriched with Firmicutes (Bacillota)—in agreement with our results and those of Eckburg et al. [23]—and Actinobacteria (Actinomycetota) and less populated by Bacteroidetes (Bacteroidota) and Proteobacteria (Pseudomonadota)—consistent with our results—compared to MAM [22]. In agreement with our data, the mucosal samples analyzed by Sun and colleagues showed an enrichment of *Propionibacterium* (phylum Actinobacteria) and *Escherichia* (phylum Proteobacteria) compared to stool samples [24]. The discrepancy in abundance of some phyla between these literature data and our results can be explained by the fact that our analyses were performed on patients with low- or high-grade dysplastic colorectal polyps, whereas the published data were on healthy subjects. Moreover, differences in experimental procedures such as sampling and extraction protocols, or data analysis could have influenced the results [24,45]. It is also possible that our swab-brushing procedure may lead to discrepancy in MAM composition detection compared to the commonly used biopsies.

We have then characterized the composition of MAM and associated metabolome in colon polyp patients stratified according to their tumor histology (high- vs. low-grade dysplasia). We show that MAM from the low-grade dysplasia group has a larger number of *Pelomonas* and *Phascolarctobacterium* than that of patients with high-grade dysplastic polyps (Figure 2a). Interestingly, *Pelomonas* is associated with the onset of multifocal atrophic gastritis and intestinal metaplasia, well established premalignant gastro-intestinal lesions [46], and is enriched in LAM of CRC patients receiving chemotherapy and/or radiotherapy treatment [47], while *Phascolarctobacterium* is enriched in stool samples of patients with CRC compared to healthy subjects [48]. Overall, our results show that *Pelomonas* and *Phascolarctobacterium* are more abundant in low-grade vs. high-grade dysplasia, which supports the hypothesis that these microorganisms may function as driver bacteria during CRC pathogenesis; however, functional experiments are needed to support this possibility.

At the species level, we detected enrichment of *Bacteroides fragilis* and *Bacteroides* spp. on the surface of low-grade polyps compared to high-grade adenomas (Figure 2b). Thus, it is tempting to speculate that these species may play a role in CRC tumor initiation. Intriguingly, enterotoxigenic *B. fragilis* is a well-known driver of CRC [3] due to its oncogenic properties. Among the species found enriched in the low-grade dysplasia group, we also found *unclassified Erysipelotrichaceae incertae sedis*, a species that belongs to the *Erysipelotrichaceae* family. Interestingly, the levels of intestinal *Erysipelotrichaceae* are reduced in the LAM of patients with advanced colon adenomas compared to healthy subjects [49], and increased in hyperplastic polyps compared to adenocarcinomas [47]. Lastly, bacteria belonging to this family play an important role in inflammation [50] and are associated with increased levels of inflammatory markers involved in tumor growth, invasion, and metastasis [51]. The lack of *Fusobacteria* enrichment in our samples can be explained by the fact that *Fusobacterium* spp. colonize more advanced CRC tissues [52].

With regard to the metabolome, we succeeded in identifying 41 metabolites differentially associated with low- and high-grade dysplasia groups, with 29 of them previously found enriched in LAM of CRC patients compared to healthy subjects [48,53–58]. These findings confirm and extend our previous work on tumor-associated metabolites isolated from 20 patients [31]. More specifically, we found erythronic acid (FC = 0.7) (Figure S4) enriched in low- compared to high-grade dysplastic polyps, a metabolite produced by bacteria from the Actinobacteria (Actinomycetota) and Proteobacteria (Pseudomonadota) phyla [56]. Fittingly, we found enrichment of the aforementioned genus *Pelomonas*, which belongs to Proteobacteria (Pseudomonadota), in low-grade dysplastic polyps.

In contrast, microbiota analyses of high-grade dysplastic adenomas showed an enrichment of the genus *Anaerococcus* (Figure 2b), that was found significantly enriched in CRC tissues and is considered a potential passenger genus [4]. Further functional studies are needed to evaluate the role of *Anaerococcus* in cancer progression.

As expected, our results indicate a gradual replacement of the potential driver bacteria by the potential passenger bacteria in high-grade adenomas. These microorganisms are opportunistic pathogens—possibly involved in CRC progression—taking advantage of the changes occurring in the TME to colonize it even further. In fact, we did not identify any enrichment of potential passenger genera or species in low-grade dysplastic samples. In this regard, a limitation of the present work is that we cannot classify bacteria as driver or passenger based on our results, but rather must rely on previously suggested classifications, which, however, are often based on disease–bacteria correlations and not on functional experiments at a mechanistic level.

Upon analysis of the polyp-associated metabolome, we found L-serine (FC = 2.8) and threonine (FC = 8.2) enriched in high- vs. low-grade dysplastic polyps (Figure S4). Interestingly, Garza et al. have recently published a computational model explaining the association between passenger bacteria and CRC metabolites, such as L-serine and threonine [17]. These authors suggest that changes in metabolite composition may allow opportunistic passenger bacteria to colonize tumor sites [17]. Moreover, a more recent study has shown that L-serine is required for CRC cell proliferation, and that dysregu-

lation of serine metabolism is closely related to the occurrence and development of this tumor [59]. Two other metabolites found enriched in the high-grade dysplasia group were lactic acid (FC = 1.9) and butyric acid (FC = 3.7) (Figure S4), which may both derive from the aerobic glycolysis occurring in CRC cells and/or the metabolism of gut bacteria. In fact, cancer cells are able to perform aerobic glycolysis even under oxygen availability, thereby producing lactate from pyruvate [60,61], a phenomenon known as the Warburg effect [62]. Due to this metabolic switch to aerobic glycolysis, cancer cells are unable to efficiently metabolize butyrate, the primary energy source of normal colonocytes. On the other hand, as aforementioned, our samples from the high-grade dysplasia group displayed increased levels of *Anaerococcus* spp. (Figure 2b), which belongs to the phylum Firmicutes (Bacillota) known to be a major source of butyric acid and lactic acid [63].

PICRUSt analysis on MAM identified four pathways statistically enriched in low-grade and 13 in high-grade dysplasia groups. Interestingly, the high-grade group showed an enrichment of the mixed acid fermentation pathway and superpathway of pyrimidine ribonucleosides salvage. The first one leads to the production of lactic acid, while the second one of uracil, both metabolites enriched in high-grade dysplastic polyps (Figures S4 and S7). It is worth noting that the PICRUSt analysis provides only a prediction of the functions of the bacterial community.

The integration of bacterial genera and metabolite class data, crucial to understanding the relationship between bacterial genera and metabolite classes [64], shows that the two genera, *Pelomonas* and *Phascolarctobacterium*, enriched in the low-grade dysplasia group, are negatively correlated with organonitrogen compounds (Figure 5). Interestingly, betaine, carnitine, and choline, whose high levels in the plasma of human subjects have been shown to reduce CRC risk [65], are organonitrogen compounds. In particular, they all are trimethylamine (TMA) precursors that can be metabolized to trimethylamine-N-oxide (TMAO) by *Phascolarctobacterium* [66]. TMAO is known to promote CRC progression not only through N-Nitroso compounds formation [67], which leads to DNA damage, but also via the production of reactive oxygen species (ROS) [68]. Thus, it is possible that an increase in *Pelomonas* and *Phascolarctobacterium*, as demonstrated by the present study, may ultimately lead to the downregulation of betaine, carnitine, and choline as the latter may be used by the bacteria to produce TMA and consequently, TMAO. Unfortunately, the method used for the analysis of mucosal-adherent metabolome is not suitable for the quantification of TMAO, so further analyses are needed to confirm this hypothesis.

Lastly, the genera, *Pelomonas*, *Phascolarctobacterium*, and *Bacteroides*, the latter of which comprises the species *B. fragilis*, (Figure 5), showed a negative correlation with benzene and benzene derivatives, which include benzoic acid and its substituted derivatives endowed with HDAC inhibitory activity [69]. Thus, it is conceivable that the decrease in benzene and derivatives may be involved in CRC development.

5. Conclusions

In conclusion, our findings (summarized in Figure 6), based on a novel sampling strategy, support the hypothesis of a direct and indirect involvement of the gut microbiota and their metabolites in CRC initiation and early progression, since we found different signatures in low-grade dysplastic polyps compared with high-grade ones, which represent more advanced stages of the adenoma–carcinoma sequence. Our results also stress the importance of analyzing tumor-associated microbiota and metabolome to identify key carcinogenic pathogens, whereas luminal-derived data, although still essential as clinical markers, only seem to recapitulate the general gut environment.

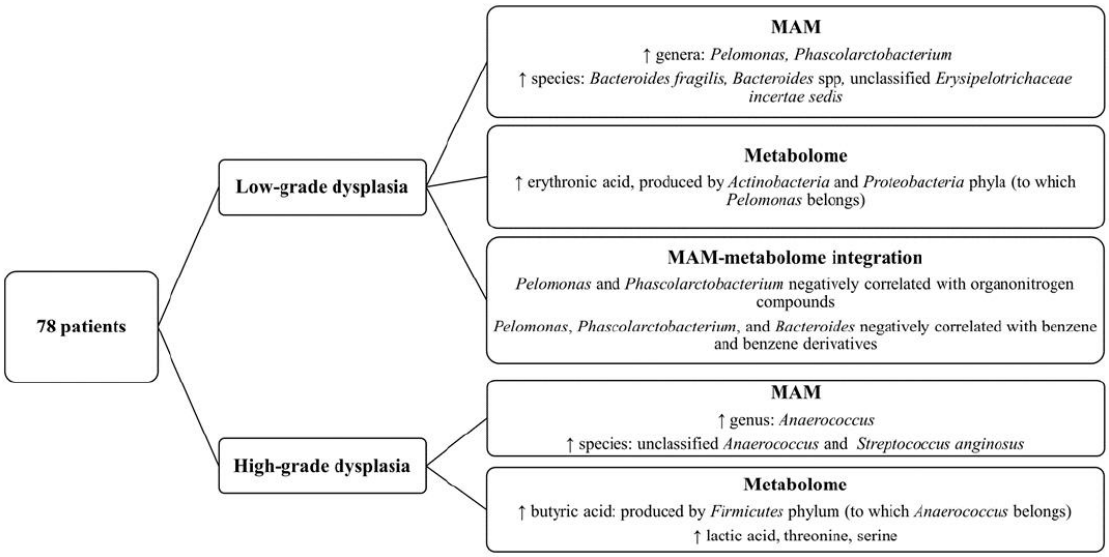


Figure 6. Diagram showing the key findings of this article about MAM and mucosa-associated metabolome of patients with high- or low-grade dysplastic polyps.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers15123065/s1, Table S1: Previous gastrointestinal conditions reported by the analyzed patients with low- (green) or high-grade (yellow) dysplastic polyps; Table S2: comparison of nutrient intake between patients with low- and high-grade dysplastic polyps; Table S3: metabolomics data; Table S4: Genera that distinguish MAM and LAM. Negative LDA score: enrichment in MAM. Positive LDA score: enrichment in LAM. Only statistically significant genera were included in the table (p-value and FDR < 0.05). The updated names covered by the International Code of Nomenclature for Prokaryotes are indicated in parentheses ([70]); Figure S1: pie chart and linear discriminant analysis effect size (LDA-LEfSe) of the comparison between MAM and LAM; Table S4: genera that distinguish MAM and LAM; Figure S2: phylogenetic heat tree showing differences in the relative abundance of MAM taxa between high-grade dysplastic (yellow) vs. low-grade dysplastic polyps (green); Figure S3: linear discriminant analysis effect size (LDA-LEfSe) showing (a) bacterial genera and (b) bacterial species enriched in LAM high- (yellow; LDS score > 2) vs. low-grade dysplastic adenomas (green; LDA score < −2); Figure S4: box plots of the most significant molecules discriminating low- (green) from high-grade (yellow) dysplastic adenomatous polyps; Figure S5: volcano plot of quantified metabolites; Figure S6: alternative representation of the hierarchical clustering heat-map reported in Figure 3a. Here, the patients with low-grade dysplastic polyps (green) are grouped on the right, while the patients with high-grade dysplastic polyps (yellow) are grouped on the left; Figure S7: products of mixed acid fermentation pathway (a) and superpathway of pyrimidine ribonucleosides salvage (b).

Author Contributions: Conceptualization, M.G.C., M.L.V., A.A. and I.D.; formal analysis, M.G.C., M.L.V., E.R., B.A., E.B. (Elisa Bona), E.P., D.F. and M.M. (Marcello Manfredi); investigation, M.G.C., M.L.V., M.S., E.B. (Elettra Barberis), E.A., M.M. (Marta Mellai), S.B., G.B. and M.M. (Marcello Manfredi); resources, S.J., P.O. and R.B.; writing—original draft preparation, M.G.C., M.L.V., M.M. (Marcello Manfredi), A.A. and I.D.; writing—review and editing, M.S., S.J., E.B. (Elettra Barberis), E.A., M.M. (Marta Mellai), S.B., G.B., P.O., R.B., E.R., B.A., E.B. (Elisa Bona), E.P. and D.F.; visualization, M.G.C., M.L.V., E.B. (Elettra Barberis), E.A., M.M. (Marcello Manfredi) and A.A.; supervision, A.A. and I.D.; project administration, A.A. and I.D.; funding acquisition, M.M. (Marcello Manfredi) and I.D. All authors have read and agreed to the published version of the manuscript.

Funding: The research leading to these results has received funding from AIRC under IG 2021-ID. 25886 project–P.I. Irma Dianzani, and from the Italian Ministry of Education, University and Research (MIUR) program "Departments of Excellence 2018–2022", FOHN Project–Department of Health Sciences, and AGING Project–Department of Translational Medicine, Università del Piemonte Orientale.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Maggiore della Carità Hospital (Novara, Italy. Study No. CE 78/19; 7 June 2019).

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The dataset comprising raw 16S rRNA sequences generated during the current study was deposited in the National Center for Biotechnology Information (NCBI) database. BioProject MIMEC Project_Swab PRJNA783496 (https://submit.ncbi.nlm.nih.gov/subs/sra/SUB10703567/overview, accessed on 17 December 2021) and MIMEC Project_Fecal PRJNA783535 (https://submit.ncbi.nlm.nih.gov/subs/sra/SUB10721767/overview, accessed on 17 December 2021).

Acknowledgments: We sincerely thank Santina Castriciano (the Scientific Affairs Director of COPAN Italia SpA) for providing eNAT™ samples to collect mucosal-adherent microbiota. We would like to thank Marcello Arsura (Abeschool) for reviewing the English language.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
2. Song, M.; Chan, A.T.; Sun, J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology* **2020**, *158*, 322–340. [CrossRef] [PubMed]
3. Tjalsma, H.; Boleij, A.; Marchesi, J.R.; Dutilh, B.E. A Bacterial Driver-Passenger Model for Colorectal Cancer: Beyond the Usual Suspects. *Nat. Rev. Microbiol.* **2012**, *10*, 575–582. [CrossRef] [PubMed]
4. Wang, Y.; Zhang, C.; Hou, S.; Wu, X.; Liu, J.; Wan, X. Analyses of Potential Driver and Passenger Bacteria in Human Colorectal Cancer. *Cancer Manag. Res.* **2020**, *12*, 11553–11561. [CrossRef]
5. Mori, G.; Pasca, M.R. Gut Microbial Signatures in Sporadic and Hereditary Colorectal Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 1312. [CrossRef] [PubMed]
6. Kosumi, K.; Mima, K.; Baba, H.; Ogino, S. Dysbiosis of the Gut Microbiota and Colorectal Cancer: The Key Target of Molecular Pathological Epidemiology. *J. Lab. Precis. Med.* **2018**, *3*, 76. [CrossRef]
7. Kostic, A.D.; Chun, E.; Robertson, L.; Glickman, J.N.; Gallini, C.A.; Michaud, M.; Clancy, T.E.; Chung, D.C.; Lochhead, P.; Hold, G.L.; et al. Fusobacterium Nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment. *Cell Host Microbe* **2013**, *14*, 207–215. [CrossRef]
8. Wu, S.; Powell, J.; Mathioudakis, N.; Kane, S.; Fernandez, E.; Sears, C.L. Bacteroides Fragilis Enterotoxin Induces Intestinal Epithelial Cell Secretion of Interleukin-8 through Mitogen-Activated Protein Kinases and a Tyrosine Kinase-Regulated Nuclear Factor-KappaB Pathway. *Infect. Immun.* **2004**, *72*, 5832–5839. [CrossRef]
9. Chung, L.; Thiele Orberg, E.; Geis, A.L.; Chan, J.L.; Fu, K.; DeStefano Shields, C.E.; Dejea, C.M.; Fathi, P.; Chen, J.; Finard, B.B.; et al. Bacteroides Fragilis Toxin Coordinates a Pro-Carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe* **2018**, *23*, 203–214.e5. [CrossRef]
10. Nougayrède, J.-P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. Escherichia Coli Induces DNA Double-Strand Breaks in Eukaryotic Cells. *Science* **2006**, *313*, 848–851. [CrossRef]
11. Avril, M.; DePaolo, R.W. "Driver-Passenger" Bacteria and Their Metabolites in the Pathogenesis of Colorectal Cancer. *Gut Microbes* **2021**, *13*, 1941710. [CrossRef] [PubMed]
12. Louis, P.; Hold, G.L.; Flint, H.J. The Gut Microbiota, Bacterial Metabolites and Colorectal Cancer. *Nat. Rev. Microbiol.* **2014**, *12*, 661–672. [CrossRef] [PubMed]
13. Hassan, C.; Gimeno-García, A.; Kalager, M.; Spada, C.; Zullo, A.; Costamagna, G.; Senore, C.; Rex, D.K.; Quintero, E. Systematic Review with Meta-Analysis: The Incidence of Advanced Neoplasia after Polypectomy in Patients with and without Low-Risk Adenomas. *Aliment. Pharmacol. Ther.* **2014**, *39*, 905–912. [CrossRef] [PubMed]
14. Peng, Y.; Nie, Y.; Yu, J.; Wong, C.C. Microbial Metabolites in Colorectal Cancer: Basic and Clinical Implications. *Metabolites* **2021**, *11*, 159. [CrossRef] [PubMed]

15. Cao, H.; Luo, S.; Xu, M.; Zhang, Y.; Song, S.; Wang, S.; Kong, X.; He, N.; Cao, X.; Yan, F.; et al. The Secondary Bile Acid, Deoxycholate Accelerates Intestinal Adenoma-Adenocarcinoma Sequence in $Apc^{Min/+}$ Mice through Enhancing Wnt Signaling. *Fam. Cancer* **2014**, *13*, 563–571. [CrossRef] [PubMed]
16. Payne, C.M.; Weber, C.; Crowley-Skillicorn, C.; Dvorak, K.; Bernstein, H.; Bernstein, C.; Holubec, H.; Dvorakova, B.; Garewal, H. Deoxycholate Induces Mitochondrial Oxidative Stress and Activates NF-KappaB through Multiple Mechanisms in HCT-116 Colon Epithelial Cells. *Carcinogenesis* **2007**, *28*, 215–222. [CrossRef]
17. Garza, D.R.; Taddese, R.; Wirbel, J.; Zeller, G.; Boleij, A.; Huynen, M.A.; Dutilh, B.E. Metabolic Models Predict Bacterial Passengers in Colorectal Cancer. *Cancer Metab.* **2020**, *8*, 3. [CrossRef]
18. Yachida, S.; Mizutani, S.; Shiroma, H.; Shiba, S.; Nakajima, T.; Sakamoto, T.; Watanabe, H.; Masuda, K.; Nishimoto, Y.; Kubo, M.; et al. Metagenomic and Metabolomic Analyses Reveal Distinct Stage-Specific Phenotypes of the Gut Microbiota in Colorectal Cancer. *Nat. Med.* **2019**, *25*, 968–976. [CrossRef]
19. Bisht, V.; Nash, K.; Xu, Y.; Agarwal, P.; Bosch, S.; Gkoutos, G.V.; Acharjee, A. Integration of the Microbiome, Metabolome and Transcriptomics Data Identified Novel Metabolic Pathway Regulation in Colorectal Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 5763. [CrossRef]
20. Chen, F.; Dai, X.; Zhou, C.-C.; Li, K.-X.; Zhang, Y.-J.; Lou, X.-Y.; Zhu, Y.-M.; Sun, Y.-L.; Peng, B.-X.; Cui, W. Integrated Analysis of the Faecal Metagenome and Serum Metabolome Reveals the Role of Gut Microbiome-Associated Metabolites in the Detection of Colorectal Cancer and Adenoma. *Gut* **2022**, *71*, 1315–1325. [CrossRef]
21. Sun, J.; Kato, I. Gut Microbiota, Inflammation and Colorectal Cancer. *Genes. Dis.* **2016**, *3*, 130–143. [CrossRef] [PubMed]
22. Ringel, Y.; Maharshak, N.; Ringel-Kulka, T.; Wolber, E.A.; Sartor, R.B.; Carroll, I.M. High Throughput Sequencing Reveals Distinct Microbial Populations within the Mucosal and Luminal Niches in Healthy Individuals. *Gut Microbes* **2015**, *6*, 173–181. [CrossRef] [PubMed]
23. Eckburg, P.B.; Bik, E.M.; Bernstein, C.N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S.R.; Nelson, K.E.; Relman, D.A. Diversity of the Human Intestinal Microbial Flora. *Science* **2005**, *308*, 1635–1638. [CrossRef] [PubMed]
24. Sun, S.; Zhu, X.; Huang, X.; Murff, H.J.; Ness, R.M.; Seidner, D.L.; Sorgen, A.A.; Blakley, I.C.; Yu, C.; Dai, Q.; et al. On the Robustness of Inference of Association with the Gut Microbiota in Stool, Rectal Swab and Mucosal Tissue Samples. *Sci. Rep.* **2021**, *11*, 14828. [CrossRef]
25. Avelar-Barragan, J.; DeDecker, L.; Lu, Z.N.; Coppedge, B.; Karnes, W.E.; Whiteson, K.L. Distinct Colon Mucosa Microbiomes Associated with Tubular Adenomas and Serrated Polyps. *NPJ Biofilms Microbiomes* **2022**, *8*, 69. [CrossRef]
26. O'Brien, C.L.; Allison, G.E.; Grimpen, F.; Pavli, P. Impact of Colonoscopy Bowel Preparation on Intestinal Microbiota. *PLoS ONE* **2013**, *8*, e62815. [CrossRef]
27. Nagata, N.; Tohya, M.; Fukuda, S.; Suda, W.; Nishijima, S.; Takeuchi, F.; Ohsugi, M.; Tsujimoto, T.; Nakamura, T.; Shimomura, A.; et al. Effects of Bowel Preparation on the Human Gut Microbiome and Metabolome. *Sci. Rep.* **2019**, *9*, 4042. [CrossRef]
28. Riboli, E.; Hunt, K.J.; Slimani, N.; Ferrari, P.; Norat, T.; Fahey, M.; Charrondière, U.R.; Hémon, B.; Casagrande, C.; Vignat, J.; et al. European Prospective Investigation into Cancer and Nutrition (EPIC): Study Populations and Data Collection. *Public Health Nutr.* **2002**, *5*, 1113–1124. [CrossRef]
29. Douglas, G.M.; Maffei, V.J.; Zaneveld, J.R.; Yurgel, S.N.; Brown, J.R.; Taylor, C.M.; Huttenhower, C.; Langille, M.G.I. PICRUSt2 for Prediction of Metagenome Functions. *Nat. Biotechnol.* **2020**, *38*, 685–688. [CrossRef]
30. Parks, D.H.; Tyson, G.W.; Hugenholtz, P.; Beiko, R.G. STAMP: Statistical Analysis of Taxonomic and Functional Profiles. *Bioinformatics* **2014**, *30*, 3123–3124. [CrossRef]
31. Barberis, E.; Joseph, S.; Amede, E.; Clavenna, M.G.; La Vecchia, M.; Sculco, M.; Aspesi, A.; Occhipinti, P.; Robotti, E.; Boldorini, R.; et al. A New Method for Investigating Microbiota-Produced Small Molecules in Adenomatous Polyps. *Anal. Chim. Acta* **2021**, *1179*, 338841. [CrossRef] [PubMed]
32. Ewing, B.; Hillier, L.; Wendl, M.C.; Green, P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Res.* **1998**, *8*, 175–185. [CrossRef] [PubMed]
33. Cole, J.R.; Wang, Q.; Fish, J.A.; Chai, B.; McGarrell, D.M.; Sun, Y.; Brown, C.T.; Porras-Alfaro, A.; Kuske, C.R.; Tiedje, J.M. Ribosomal Database Project: Data and Tools for High Throughput RRNA Analysis. *Nucleic Acids Res.* **2014**, *42*, D633–D642. [CrossRef]
34. Bona, E.; Massa, N.; Toumatia, O.; Novello, G.; Cesaro, P.; Todeschini, V.; Boatti, L.; Mignone, F.; Titouah, H.; Zitouni, A.; et al. Climatic Zone and Soil Properties Determine the Biodiversity of the Soil Bacterial Communities Associated to Native Plants from Desert Areas of North-Central Algeria. *Microorganisms* **2021**, *9*, 1359. [CrossRef] [PubMed]
35. Torre, E.; Sola, D.; Caramaschi, A.; Mignone, F.; Bona, E.; Fallarini, S. A Pilot Study on Clinical Scores, Immune Cell Modulation, and Microbiota Composition in Allergic Patients with Rhinitis and Asthma Treated with a Probiotic Preparation. *Int. Arch. Allergy Immunol.* **2022**, *183*, 186–200. [CrossRef]
36. Dhariwal, A.; Chong, J.; Habib, S.; King, I.L.; Agellon, L.B.; Xia, J. MicrobiomeAnalyst: A Web-Based Tool for Comprehensive Statistical, Visual and Meta-Analysis of Microbiome Data. *Nucleic Acids Res.* **2017**, *45*, W180–W188. [CrossRef]
37. Foster, Z.S.L.; Sharpton, T.J.; Grünwald, N.J. Metacoder: An R Package for Visualization and Manipulation of Community Taxonomic Diversity Data. *PLoS Comput. Biol.* **2017**, *13*, e1005404. [CrossRef]
38. McMurdie, P.J.; Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [CrossRef]

39. Agus, A.; Clément, K.; Sokol, H. Gut Microbiota-Derived Metabolites as Central Regulators in Metabolic Disorders. *Gut* **2021**, *70*, 1174–1182. [CrossRef]
40. Lamichhane, S.; Sen, P.; Dickens, A.M.; Orešič, M.; Bertram, H.C. Gut Metabolome Meets Microbiome: A Methodological Perspective to Understand the Relationship between Host and Microbe. *Methods* **2018**, *149*, 3–12. [CrossRef]
41. Yuan, C.; Graham, M.; Staley, C.; Subramanian, S. Mucosal Microbiota and Metabolome along the Intestinal Tract Reveal a Location-Specific Relationship. *mSystems* **2020**, *5*, e00055-20. [CrossRef] [PubMed]
42. Gao, R.; Kong, C.; Huang, L.; Li, H.; Qu, X.; Liu, Z.; Lan, P.; Wang, J.; Qin, H. Mucosa-Associated Microbiota Signature in Colorectal Cancer. *Eur. J. Clin. Microbiol. Infect. Dis.* **2017**, *36*, 2073–2083. [CrossRef]
43. Wang, Q.; Ye, J.; Fang, D.; Lv, L.; Wu, W.; Shi, D.; Li, Y.; Yang, L.; Bian, X.; Wu, J.; et al. Multi-Omic Profiling Reveals Associations between the Gut Mucosal Microbiome, the Metabolome, and Host DNA Methylation Associated Gene Expression in Patients with Colorectal Cancer. *BMC Microbiol.* **2020**, *20*, 83. [CrossRef]
44. Donaldson, G.P.; Lee, S.M.; Mazmanian, S.K. Gut Biogeography of the Bacterial Microbiota. *Nat. Rev. Microbiol.* **2016**, *14*, 20–32. [CrossRef] [PubMed]
45. Tang, M.S.; Poles, J.; Leung, J.M.; Wolff, M.J.; Davenport, M.; Lee, S.C.; Lim, Y.A.; Chua, K.H.; Loke, P.; Cho, I. Inferred Metagenomic Comparison of Mucosal and Fecal Microbiota from Individuals Undergoing Routine Screening Colonoscopy Reveals Similar Differences Observed during Active Inflammation. *Gut Microbes* **2015**, *6*, 48–56. [CrossRef] [PubMed]
46. Yang, I.; Woltemate, S.; Piazuelo, M.B.; Bravo, L.E.; Yepez, M.C.; Romero-Gallo, J.; Delgado, A.G.; Wilson, K.T.; Peek, R.M.; Correa, P.; et al. Different Gastric Microbiota Compositions in Two Human Populations with High and Low Gastric Cancer Risk in Colombia. *Sci. Rep.* **2016**, *6*, 18594. [CrossRef]
47. Mori, G.; Rampelli, S.; Orena, B.S.; Rengucci, C.; De Maio, G.; Barbieri, G.; Passardi, A.; Casadei Gardini, A.; Frassineti, G.L.; Gaiarsa, S.; et al. Shifts of Faecal Microbiota During Sporadic Colorectal Carcinogenesis. *Sci. Rep.* **2018**, *8*, 10329. [CrossRef] [PubMed]
48. Weir, T.L.; Manter, D.K.; Sheflin, A.M.; Barnett, B.A.; Heuberger, A.L.; Ryan, E.P. Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults. *PLoS ONE* **2013**, *8*, e70803. [CrossRef]
49. Zagato, E.; Pozzi, C.; Bertocchi, A.; Schioppa, T.; Saccheri, F.; Guglietta, S.; Fosso, B.; Melocchi, L.; Nizzoli, G.; Troisi, J.; et al. Endogenous Murine Microbiota Member Faecalibaculum Rodentium and Its Human Homologue Protect from Intestinal Tumour Growth. *Nat. Microbiol.* **2020**, *5*, 511–524. [CrossRef]
50. Clos-Garcia, M.; Garcia, K.; Alonso, C.; Iruarrizaga-Lejarreta, M.; D'Amato, M.; Crespo, A.; Iglesias, A.; Cubiella, J.; Bujanda, L.; Falcón-Pérez, J.M. Integrative Analysis of Fecal Metagenomics and Metabolomics in Colorectal Cancer. *Cancers* **2020**, *12*, 1142. [CrossRef]
51. Dinh, D.M.; Volpe, G.E.; Duffalo, C.; Bhalchandra, S.; Tai, A.K.; Kane, A.V.; Wanke, C.A.; Ward, H.D. Intestinal Microbiota, Microbial Translocation, and Systemic Inflammation in Chronic HIV Infection. *J. Infect. Dis.* **2015**, *211*, 19–27. [CrossRef] [PubMed]
52. Flanagan, L.; Schmid, J.; Ebert, M.; Soucek, P.; Kunicka, T.; Liska, V.; Bruha, J.; Neary, P.; Dezeeuw, N.; Tommasino, M.; et al. Fusobacterium Nucleatum Associates with Stages of Colorectal Neoplasia Development, Colorectal Cancer and Disease Outcome. *Eur. J. Clin. Microbiol. Infect. Dis.* **2014**, *33*, 1381–1390. [CrossRef]
53. Cheng, Y.; Xie, G.; Chen, T.; Qiu, Y.; Zou, X.; Zheng, M.; Tan, B.; Feng, B.; Dong, T.; He, P.; et al. Distinct Urinary Metabolic Profile of Human Colorectal Cancer. *J. Proteome Res.* **2012**, *11*, 1354–1363. [CrossRef] [PubMed]
54. Ni, Y.; Xie, G.; Jia, W. Metabonomics of Human Colorectal Cancer: New Approaches for Early Diagnosis and Biomarker Discovery. *J. Proteome Res.* **2014**, *13*, 3857–3870. [CrossRef] [PubMed]
55. Brown, D.G.; Rao, S.; Weir, T.L.; O'Malia, J.; Bazan, M.; Brown, R.J.; Ryan, E.P. Metabolomics and Metabolic Pathway Networks from Human Colorectal Cancers, Adjacent Mucosa, and Stool. *Cancer Metab.* **2016**, *4*, 11. [CrossRef]
56. Sinha, R.; Ahn, J.; Sampson, J.N.; Shi, J.; Yu, G.; Xiong, X.; Hayes, R.B.; Goedert, J.J. Fecal Microbiota, Fecal Metabolome, and Colorectal Cancer Interrelations. *PLoS ONE* **2016**, *11*, e0152126. [CrossRef] [PubMed]
57. Goedert, J.J.; Sampson, J.N.; Moore, S.C.; Xiao, Q.; Xiong, X.; Hayes, R.B.; Ahn, J.; Shi, J.; Sinha, R. Fecal Metabolomics: Assay Performance and Association with Colorectal Cancer. *Carcinogenesis* **2014**, *35*, 2089–2096. [CrossRef]
58. Wang, X.; Wang, J.; Rao, B.; Deng, L. Gut Flora Profiling and Fecal Metabolite Composition of Colorectal Cancer Patients and Healthy Individuals. *Exp. Ther. Med.* **2022**, *23*, 250. [CrossRef]
59. Zhao, X.; Fu, J.; Hu, B.; Chen, L.; Wang, J.; Fang, J.; Ge, C.; Lin, H.; Pan, K.; Fu, L.; et al. Serine Metabolism Regulates YAP Activity Through USP7 in Colon Cancer. *Front. Cell. Dev. Biol.* **2021**, *9*, 639111. [CrossRef]
60. Qian, J.; Gong, Z.-C.; Zhang, Y.-N.; Wu, H.-H.; Zhao, J.; Wang, L.-T.; Ye, L.-J.; Liu, D.; Wang, W.; Kang, X.; et al. Lactic Acid Promotes Metastatic Niche Formation in Bone Metastasis of Colorectal Cancer. *Cell. Commun. Signal.* **2021**, *19*, 9. [CrossRef]
61. Wei, Y.; Xu, H.; Dai, J.; Peng, J.; Wang, W.; Xia, L.; Zhou, F. Prognostic Significance of Serum Lactic Acid, Lactate Dehydrogenase, and Albumin Levels in Patients with Metastatic Colorectal Cancer. *BioMed Res. Int.* **2018**, *2018*, 1804086. [CrossRef] [PubMed]
62. Cruz, M.D.; Ledbetter, S.; Chowdhury, S.; Tiwari, A.K.; Momi, N.; Wali, R.K.; Bliss, C.; Huang, C.; Lichtenstein, D.; Bhattacharya, S.; et al. Metabolic Reprogramming of the Premalignant Colonic Mucosa Is an Early Event in Carcinogenesis. *Oncotarget* **2017**, *8*, 20543–20557. [CrossRef] [PubMed]
63. Van den Abbeele, P.; Belzer, C.; Goossens, M.; Kleerebezem, M.; De Vos, W.M.; Thas, O.; De Weirdt, R.; Kerckhof, F.-M.; Van de Wiele, T. Butyrate-Producing Clostridium Cluster XIVa Species Specifically Colonize Mucins in an in Vitro Gut Model. *ISME J.* **2013**, *7*, 949–961. [CrossRef]

64. Dalal, N.; Jalandra, R.; Sharma, M.; Prakash, H.; Makharia, G.K.; Solanki, P.R.; Singh, R.; Kumar, A. Omics Technologies for Improved Diagnosis and Treatment of Colorectal Cancer: Technical Advancement and Major Perspectives. *BioMed Pharmacother.* **2020**, *131*, 110648. [CrossRef]
65. Nitter, M.; Norgård, B.; de Vogel, S.; Eussen, S.J.P.M.; Meyer, K.; Ulvik, A.; Ueland, P.M.; Nygård, O.; Vollset, S.E.; Bjørge, T.; et al. Plasma Methionine, Choline, Betaine, and Dimethylglycine in Relation to Colorectal Cancer Risk in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Ann. Oncol.* **2014**, *25*, 1609–1615. [CrossRef] [PubMed]
66. Simó, C.; García-Cañas, V. Dietary Bioactive Ingredients to Modulate the Gut Microbiota-Derived Metabolite TMAO. New Opportunities for Functional Food Development. *Food Funct.* **2020**, *11*, 6745–6776. [CrossRef]
67. Oellgaard, J.; Winther, S.A.; Hansen, T.S.; Rossing, P.; von Scholten, B.J. Trimethylamine N-Oxide (TMAO) as a New Potential Therapeutic Target for Insulin Resistance and Cancer. *Curr. Pharm. Des.* **2017**, *23*, 3699–3712. [CrossRef]
68. Li, T.; Chen, Y.; Gua, C.; Li, X. Elevated Circulating Trimethylamine N-Oxide Levels Contribute to Endothelial Dysfunction in Aged Rats through Vascular Inflammation and Oxidative Stress. *Front. Physiol.* **2017**, *8*, 350. [CrossRef]
69. Anantharaju, P.G.; Reddy, B.D.; Padukudru, M.A.; Kumari Chitturi, C.M.; Vimalambike, M.G.; Madhunapantula, S.V. Naturally Occurring Benzoic Acid Derivatives Retard Cancer Cell Growth by Inhibiting Histone Deacetylases (HDAC). *Cancer Biol. Ther.* **2017**, *18*, 492–504. [CrossRef]
70. Oren, A.; Garrity, G.M. Valid publication of the names of forty-two phyla of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 005056. [CrossRef]

Supplementary Material

Low-grade	Previous gastrointestinal diseases or procedures	High-grade	Previous gastrointestinal diseases or procedures
8	No	7	Diverticulitis
16	Diverticulitis	10	Diverticulitis
18	Previous polyps occurrence	11	Diverticulitis
20	Diverticulitis	12	No
21	Diverticulitis; previous polyps occurrence	15	No
22	No	28	No
23	No	31	No
25	No	36	No
26	No	44	No
29	No	48	No
32	No	52	No
34	No	53	No
37	Diverticulitis	54	Cholecystectomy (65 years)
39	Cholecystectomy (57 years)	72	IBD at diagnosis, no treatment before colonoscopy and LAM collection
47	Diverticulitis	86	No
49	Previous polyps occurrence	97	No
50	No	99	No
51	No	105	No
57	Diverticulitis	110	Diverticulitis
58	Diverticulitis	112	Diverticulitis
59	Diverticulitis	119	No
60	Previous polyps occurrence	120	No
64	No	121	Previous polyps occurrence
65	No	122	No
66	Previous polyps occurrence	123	No
68	No	124	Diverticulitis; previous polyps occurrence
69	Diverticulitis	128	Diverticulitis
71	Diverticulitis	138	No
73	Diverticulitis	140	Slight mucosal inflammation
77	Diverticulitis	141	No
83	No	155	Diverticulitis
107	Diverticulitis	156	No
111	Diverticulitis; previous polyps occurrence	160	No
113	Cholecystectomy (79 years); diverticulitis	162	No
125	Cholecystectomy (73 years); diverticulitis		
130	Cholecystectomy (55 years)		
132	Diverticulitis		
137	No		
139	Previous polyps occurrence		
153	No		
154	Diverticulitis		
157	No		
159	Diverticulitis		
163	No		

Table S1. Previous gastrointestinal conditions reported by the analyzed patients with low- (green) or high-grade (yellow) dysplastic polyps.

Nutrient	Median g/day (IQR)	*p*-value
Red meat		
Low-grade	29.4 (15.7-60.4)	0.5
High-grade	42.6 (23.2-58.4)	
Red and processed meat		
Low-grade	49.5 (29.6-90.9)	0.4
High-grade	65.5 (45.6-84.6)	
Fruit and vegetables		
Low-grade	375.8 (328.1-525.4)	0.8
High grade	431.3 (292.3-560.3)	
Fibers		
Low-grade	18.2 (13.4-21.8)	0.6
High grade	19.5 (12.6-23.4)	
Lipids		
Low-grade	71.6 (47.5-85.6)	0.5
High-grade	73.8 (56.3-90.2)	

Table S2. Comparison of nutrient intake between patients with low- and high-grade dysplastic polyps. Table S2 shows the median value for the two groups and the interquartile range (IQR). The *p*-value was obtained using Mann-Whitney statistical test. A p-value < 0.05 was considered statistically significant.

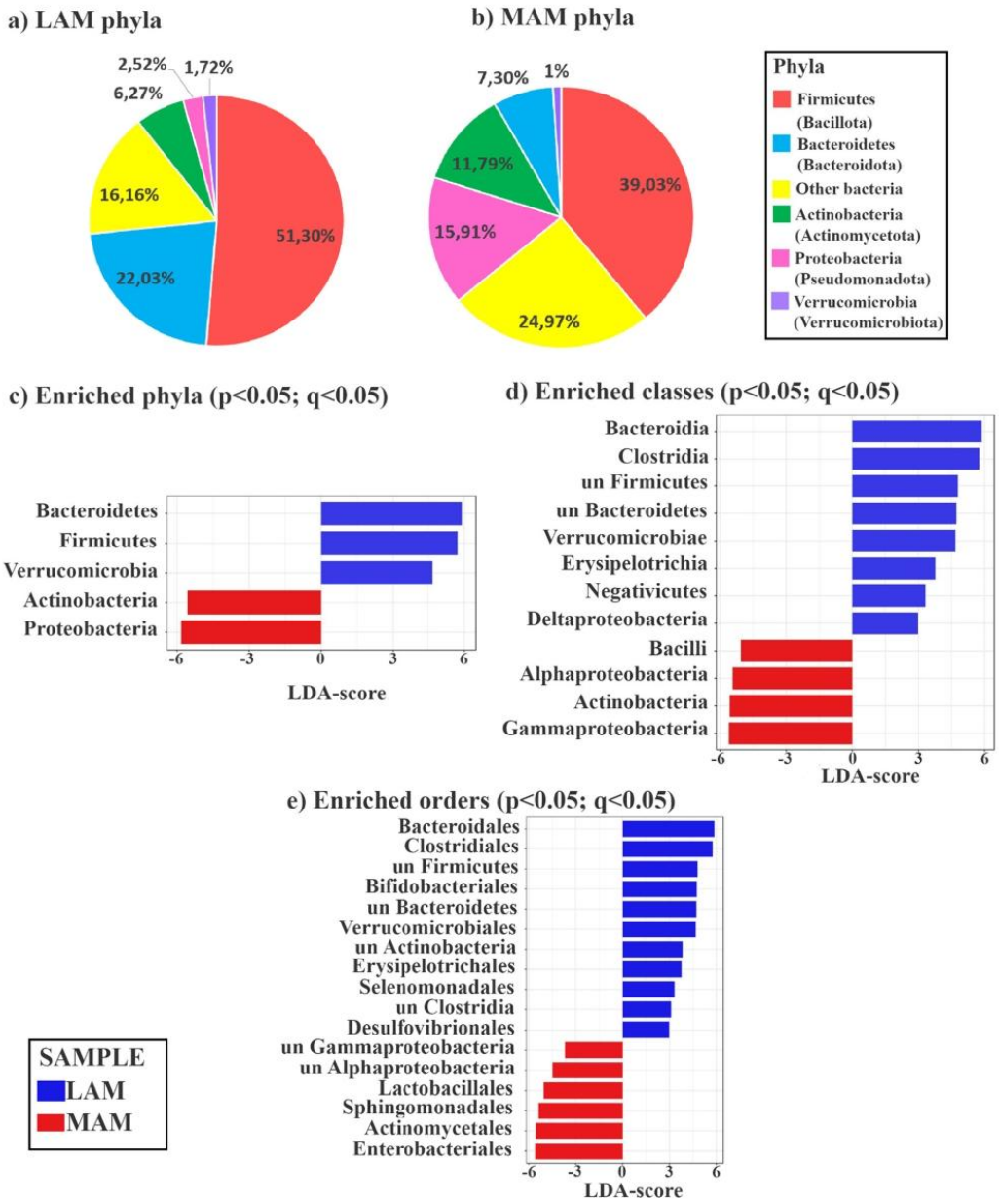


Fig. S1. Pie chart showing the relative abundance percentage of different phyla in LAM (a) *vs* MAM (b). LefSe representing phyla (c) classes (d) and orders (e) significantly enriched in LAM (blue) or MAM (red).

Phylum	*Genus*	*Pvalues*	*FDR*	*LDAscore*
Actinobacteria (Actinomycetota)	*Propionibacterium*	7,42E-24	4,75E-22	-4.89
Actinobacteria (Actinomycetota)	*unclassified_Actinomycetales*	2,38E-20	7,19E-18	-4.14
Actinobacteria (Actinomycetota)	*Arthrobacter*	3,30E-13	2,35E-12	-5.14
Actinobacteria (Actinomycetota)	*unclassified_Micrococcaceae*	4,94E-13	3,16E-12	-5.14
Actinobacteria (Actinomycetota)	*unclassified_Bifidobacteriaceae*	1,40E-01	2,89E-01	4.39
Actinobacteria (Actinomycetota)	*unclassified_Coriobacteriaceae*	0.0043271	0.0067545	-2.89
Actinobacteria (Actinomycetota)	*unclassified_Actinobacteria*	0.0084067	0.01281	3.84
Bacteroidetes (Bacteroidota)	*unclassified_Bacteroidales*	3,37E-19	7,19E-18	5.19
Bacteroidetes (Bacteroidota)	*unclassified_Bacteroidetes*	9,98E-18	1,38E-16	4.72
Bacteroidetes (Bacteroidota)	*unclassified_Porphyromonadaceae*	1,58E-16	1,68E-15	4.32
Bacteroidetes (Bacteroidota)	*unclassified_Rikenellaceae*	2,03E-12	1,09E-11	3.54
Bacteroidetes (Bacteroidota)	*Barnesiella*	1.43e-13	6,54E-09	4.75
Bacteroidetes (Bacteroidota)	*Bacteroides*	5,47E-09	2,34E-08	5.51
Bacteroidetes (Bacteroidota)	*Alistipes*	2,55E-08	1,02E-08	4.31
Bacteroidetes (Bacteroidota)	*Parabacteroides*	4,10E-09	1,54E-08	4.89
Bacteroidetes (Bacteroidota)	*Odoribacter*	2,65E-06	8,49E-06	3.66
Bacteroidetes (Bacteroidota)	*Butyricimonas*	3,93E-05	1,20E-04	3.64
Bacteroidetes (Bacteroidota)	*unclassified_Prevotellaceae*	5,41E-02	1,24E-01	4.81
Bacteroidetes (Bacteroidota)	*Prevotella*	0.017516	0.023851	4.04
Firmicutes (Bacillota)	*unclassified_Clostridiales*	2,11E-14	1,93E-13	5.19
Firmicutes (Bacillota)	*Faecalibacterium*	5,93E-11	2,92E-10	5.44
Firmicutes (Bacillota)	*unclassified_Firmicutes*	1,90E-07	6,77E-07	4.79
Firmicutes (Bacillota)	*Roseburia*	2,90E-07	9,76E-07	4.52
Firmicutes (Bacillota)	*Flavonifractor*	5,67E-05	1,65E-04	3.54
Firmicutes (Bacillota)	*Oscillibacter*	1,60E-04	4,46E-05	3.76
Firmicutes (Bacillota)	*Lachnospiracea_incertae_sedis*	1,41E-03	3,77E-03	4.76
Firmicutes	*unclassified_Lachnospiraceae*	2,94E-02	7,23E-02	5.31

(Bacillota)				
Firmicutes (Bacillota)	*unclassified_Peptostreptococcaceae*	4,27E-02	1,01E-01	3.93
Firmicutes (Bacillota)	*Ruminococcus*	8,30E-03	1,83E-01	4.54
Firmicutes (Bacillota)	*unclassified_Clostridia*	9,47E-02	2,02E-01	3.1
Firmicutes (Bacillota)	*Streptococcus*	3,53E-01	7,06E-01	-4.98
Firmicutes (Bacillota)	*Butyricicoccus*	7,33E-01	0.00014221	2.84
Firmicutes (Bacillota)	*Clostridium_XlVa*	0.00034832	0.00064562	3.87
Firmicutes (Bacillota)	*unclassified_Erysipelotrichaceae*	0.00035307	0.00064562	4.15
Firmicutes (Bacillota)	*unclassified_Ruminococcaceae*	0.00073058	0.0012637	4.65
Firmicutes (Bacillota)	*unclassified_Selenomonadales*	0.0034355	0.0056377	3.32
Firmicutes (Bacillota)	*Clostridium_XlVb*	0.0036883	0.0059012	-3.08
Firmicutes (Bacillota)	*unclassified_Streptococcaceae*	0.010488	0.015255	-3.79
Firmicutes (Bacillota)	*unclassified_Veillonellaceae*	0.012696	0.018057	3.12
Firmicutes (Bacillota)	*Clostridium_XI*	0.015786	0.021963	3.66
Proteobacteria (Pseudomonadota)	*unclassified_Sphingomonadaceae*	1,08E-17	1,38E-16	-4.4
Proteobacteria (Pseudomonadota)	*unclassified_Alphaproteobacteria*	1,10E-13	8,76E-13	-4.49
Proteobacteria (Pseudomonadota)	*Sphingomonas*	8,49E-13	4,94E-12	-5.33
Proteobacteria (Pseudomonadota)	*unclassified_Gammaproteobacteria*	4,08E-03	1,04E-02	-3.68
Proteobacteria (Pseudomonadota)	*Bilophila*	0.00059991	0.0010665	2.98
Proteobacteria (Pseudomonadota)	*Escherichia_Shigella*	0.00081832	0.0013782	-5.39
Proteobacteria (Pseudomonadota)	*Gemmiger*	0.019303	0.025737	-3.73
Proteobacteria (Pseudomonadota)	*unclassified_Enterobacteriaceae*	0.026979	0.035238	-5.18
Verrucomicrobia (Verrucomicrobiota)	*Akkermansia*	0.010152	0.015109	4.68

Table S4. Genera that distinguish MAM and LAM. Negative LDA score: enrichment in MAM. Positive LDA score: enrichment in LAM. Only statistically significant genera were included in the table (p-value and FDR < 0.05). The updated names covered by the International Code of Nomenclature for Prokaryotes are indicated in parentheses (Oren A, Garrity GM. Valid publication of the names of forty-two phyla of prokaryotes. Int J Syst Evol Microbiol. 2021 Oct;71(10). doi: 10.1099/ijsem.0.005056)

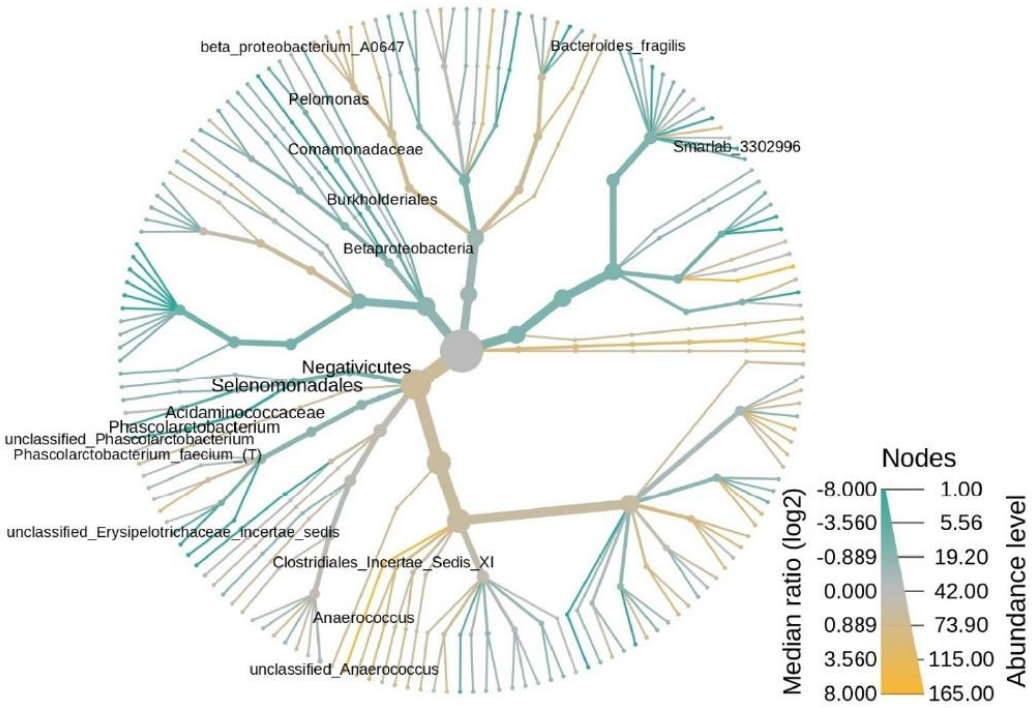


Fig. S2. Phylogenetic heat tree showing differences in the relative abundance of MAM taxa between low- (green) and high-grade dysplasia (yellow) groups. Phyla, classes, orders, families, genera, and species are represented. The nodes indicate the hierarchical structure of taxa. Only the nodes with a statistically significant difference ($p < 0.05$) between the two groups are labeled.

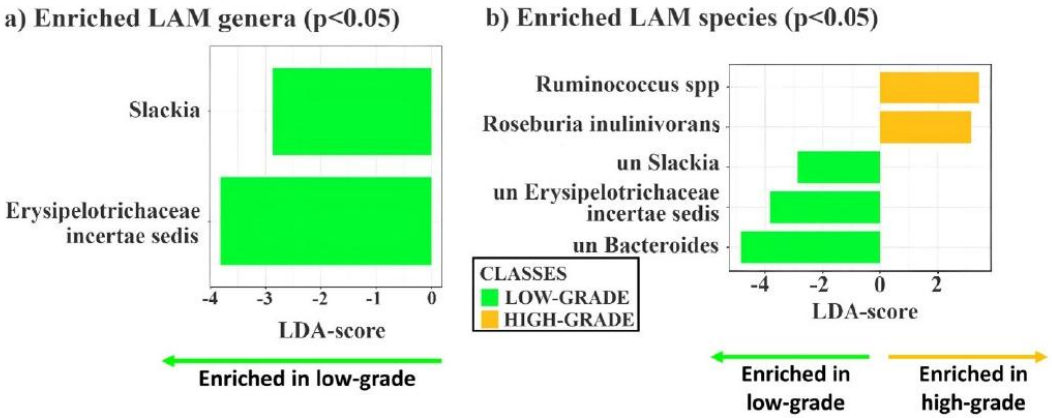


Fig. S3. Linear discriminant analysis effect size (LDA-LEfSe) showing (a) bacterial genera and (b) species enriched in LAM high-grade dysplastic (yellow; LDA score > 2) *vs* low-grade dysplastic polyps (green; LDA score < -2). This method incorporates statistical significance (Kruskal-Wallis) with biological consistency (effect size). The length of the bar represents a log10 transformed LDA score. This value is positive if the bacterial species is enriched in the first compared to the second group and negative if the second group shows enrichment compared to the first group. A significance level of $p < 0.05$ and an LDA score of 2 and -2 are used to determine the species best characterizing each phenotype.

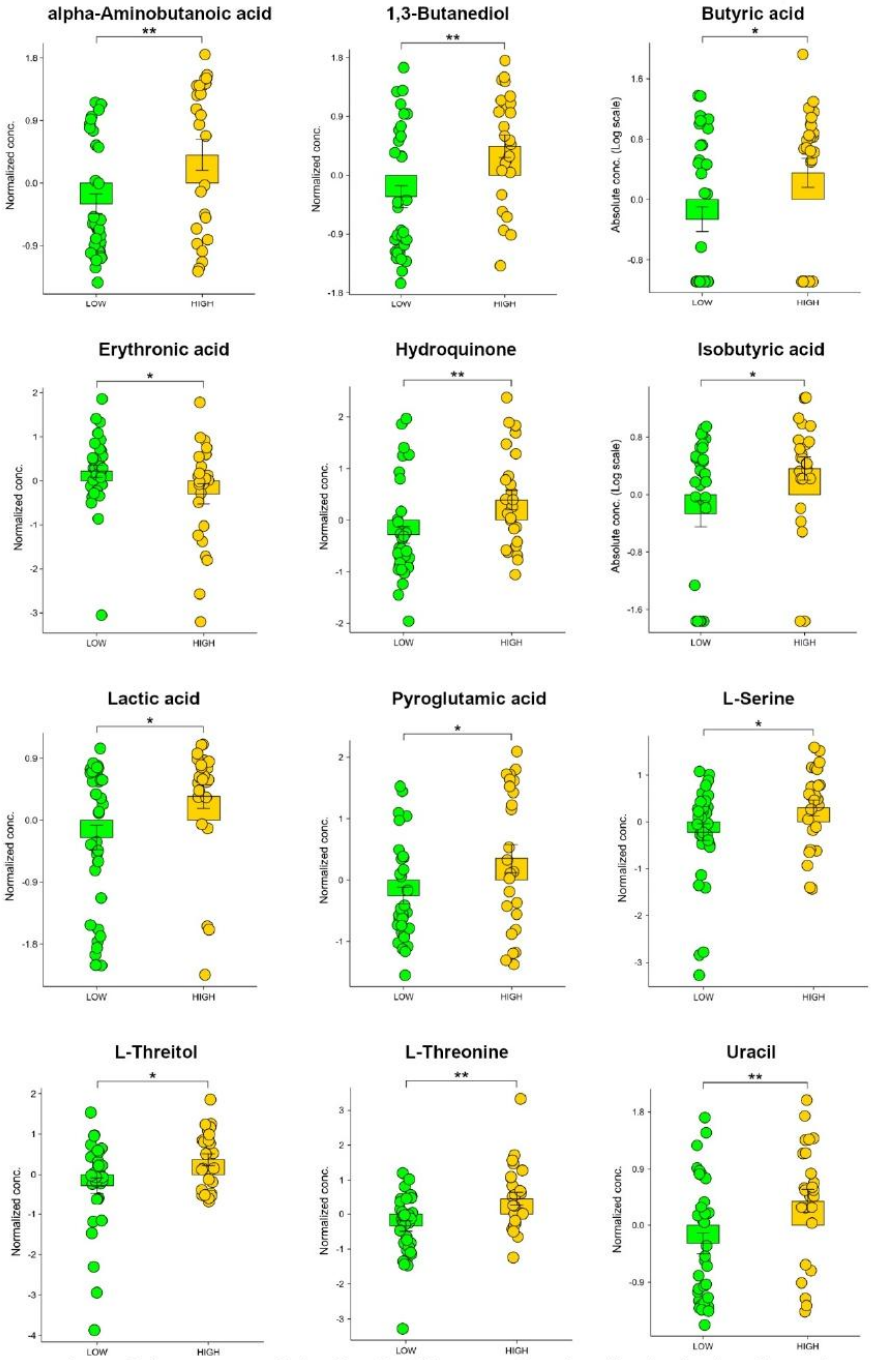


Fig. S4. Box plots of the most statistically significant molecules discriminating low- (green) from high-grade (yellow) dysplastic adenomatous polyps (*$p < 0.05$; **$p < 0.01$).

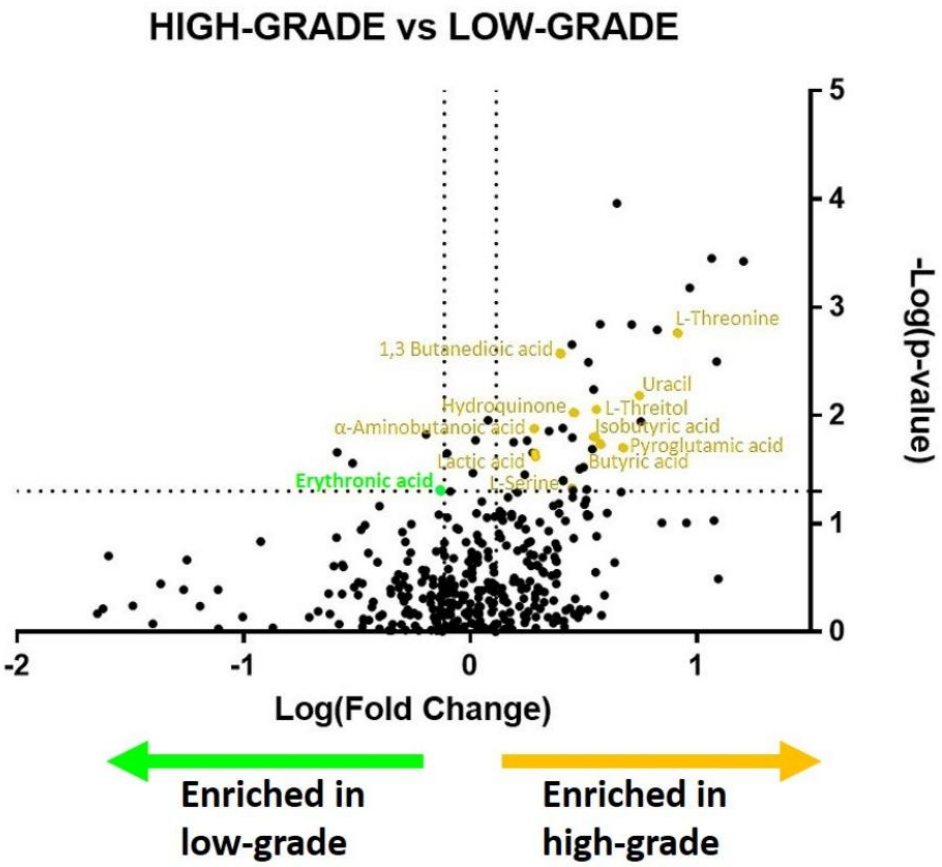


Fig. S5. Volcano plot of quantified metabolites. The figure shows the metabolites enriched in low- (green; on the left) or high- (yellow; on the right) grade dysplastic polyps that are discussed in the text.

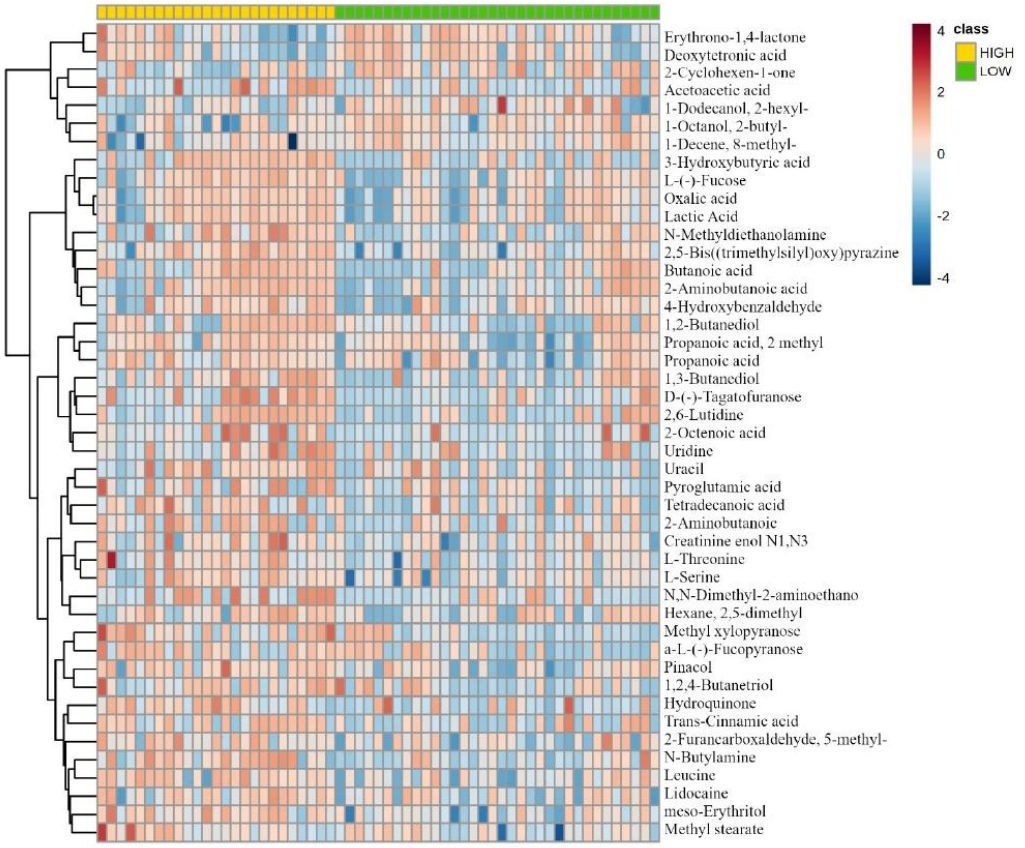
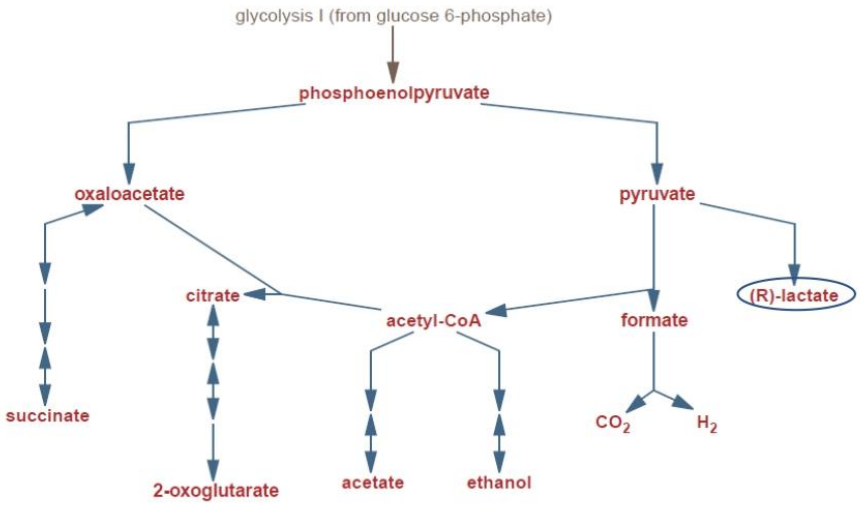


Fig. S6. Alternative representation of the hierarchical clustering heat-map reported in Figure 3a. Here the patients with low-grade dysplastic polyps (green) are grouped on the right, while the patients with high-grade dysplastic polyps (yellow) on the left.

a)



b)

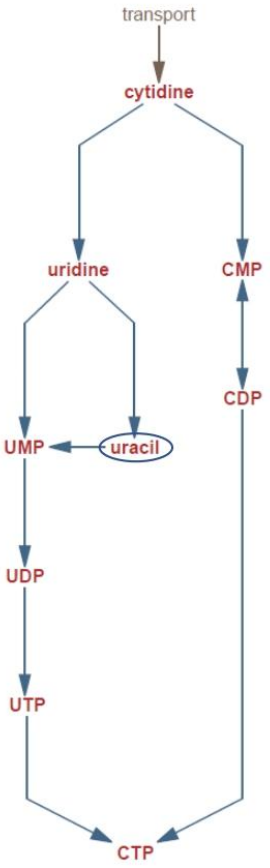


Fig. S7. Products of mixed acid fermentation pathway (a), and superpathway of pyrimidine ribonucleosides salvage (b) from the MetaCyc website (https://metacyc.org/). Circled products are discussed in the text.

Chapter 5

Dissecting the interaction among the intestinal microbiota, diet, and colorectal cancer in obese or normal-weight patients

(Manuscript in preparation, La Vecchia *et al.*)

5.1 Introduction

Trillions of microorganisms, collectively referred as the gut microbiota, colonize our intestine and strongly impact human physiology and pathology. The gut microbiota plays a pivotal role in modulating the maturation and function of immune cells and several diseases, characterized by inflammation and autoimmunity, have been recently associated with intestinal dysbiosis, i.e. compositional and functional alterations of the gut microbiota.

A role of the intestinal microbiota in colon carcinogenesis, either directly or through the production of metabolites able to increase enterocyte proliferation and to support opportunistic pathogens[124], is now generally accepted. Interestingly, also gut inflammation and obesity, well known risk factors for colorectal cancer (CRC), have been associated with gut dysbiosis[3], suggesting a biological link among these conditions.

An increased body mass index (BMI) has been associated with CRC development[3]. Several pathogenic mechanisms have been hypothesized, including the activation of insulin and IGF receptor pathways and the ensuing induction of proliferation and inhibition of apoptosis; the proinflammatory effect of secondary bile acids if the patients are on a high fat diet; systemic inflammation[3]. It is well known that the intestinal microbiota differs between lean and obese individuals. The transplant of the microbiota from fat to lean mice or humans causes a weight increase in the lean individuals and *viceversa*[157]. Several microbial species have been associated with obesity, either with a favoring or a preventative role.

Chronic low-grade inflammation that characterizes obesity[158] leads to intestinal dysbiosis that provides a favorable environment for proliferation of bacteria such as *Enterobacteriaceae*[159]. *Enterobacteriaceae* deplete oxygen in the gut and allow the colonization of *Clostridium* and *Bacteroides*, that together with *Enterobacteriaceae* may induce CRC initiation[138,160].

The diet of obese subjects is mainly characterized by a high intake of red and processed meat, sugar-sweetened beverages, refined grains, saturated fats, and low consumption of fruits, vegetables, whole grains, fish, nuts, and seeds[161]. This unhealthy dietary pattern known as Western diet leads to an excessive energy intake that causes fat accumulation and consequently chronic inflammation. Western diet is also associated with an increased risk of CRC and tumor recurrence[65]. A high intake of red and processed meat causes gut dysbiosis and promotes the proliferation of some bacterial genera including *Bacteroides* and *Erysipelotrichaceae*[162,163], candidate driver and passenger genera involved in CRC initiation and progression respectively[138].

Based on the aforementioned evidence, the aim of this study was to deepen the knowledge about different risk factors for CRC (obesity, diet, gut microbiota and metabolome) and to understand how they cooperate for CRC development in different conditions (obese or normal-weight). To this end,

we collected anthropometric parameters, dietary habits (gathered through administration of a validated EPIC questionnaire[164]), mucosa-associated and lumen-associated microbiota (MAM and LAM respectively) and metabolome of 120 patients that carried colorectal polyps.

5.2 Methodology

5.2.1 Patients

The study involved 120 patients with colorectal polyps, 11 of which included malignant portions, recruited before colonoscopy at the Gastroenterology Unit of University Hospital "Maggiore della Carità" in Novara, Italy. Patients were included in the study after obtaining their informed consent only if they had polyps larger than 1 cm and were older than 18 years of age.
Anthropometric parameters (height, weight and waist circumference) were collected 14 days after colonoscopy and were used to classify patients in two different groups: normal weight (NW) subjects with BMI < 25 Kg/m^2 and WC ≤ 88 cm for women and 102 cm for men, whereas overweight/obese (OB) subjects with BMI ≥ 25 Kg/m^2 and/or WC > 88 cm for women and 102 cm for men.

5.2.2 Sample collection

MAM was collected using an e-NAT™ (COPAN, Brescia, Italy) swab, that was gently brushed on the adenoma surface without compromising the tissue integrity. Samples were stored at -80°C until 16S rDNA sequencing. Fecal samples to analyze LAM were collected from patients 14 days after colonoscopy, aliquoted and stored as swab samples.

5.2.3 Histology

Polyps were included in neutral buffered formalin and paraffin, then cut in sections of 4 μm and stained using hematoxylin-eosin. All polyps were evaluated by a pathologist at the University Hospital in Novara, Italy. Based on their anatomical site, polyps were categorized as proximal (cecum, ascending colon, hepatic flexure, and transverse colon) or distal (splenic flexure, descending colon, sigmoid colon and rectal)[165]. Based on polyp histology, patients with hyperplastic polyps, serrated polyps without dysplasia or low-grade dysplasia adenomas were included in the “low-grade” group, while patients with high-grade dysplasia adenomas or adenocarcinomas were included in the “high-grade” group (Table 5.1). Hyperplastic polyps were included in the low-grade group because hyperplasia represents the first step in multistep carcinogenesis[166].

5.2.4 Mucosa-associated and lumen-associated microbiota analyses

MAM was obtained extracting microbial DNA from e-NAT™ swabs using QIAamp® DNA Microbiome kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.
Microbial DNA to analyze LAM was extracted with QIAamp® PowerFecal® Pro DNA kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions. The yield and quality of microbial DNA was determined using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The quantity was assessed with Invitrogen™ Qubit™ 1X dsDNA HS Assay Kit (Invitrogen Co., Thermo Fisher Scientific Inc.) using a Qubit 4 fluorometer (Invitrogen).
16S rRNA amplicon sequencing analysis was performed both for MAM and LAM using Microbiota Solution B Kit, a next generation sequencing (NGS) *in vitro* molecular test, CE-IVD marked (Arrow Diagnostics Srl, Genoa, Italy). The patented degenerate primer sets within the Arrow Microbiota Solution B kit (cod. AD-002.024) were used to amplify the V3-V4-V6 hypervariable regions of bacterial 16S rRNA by polymerase chain reaction (PCR). PCR products were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter Inc., Brea, CA, USA). The DNA concentration of the libraries was fluorometrically measured and samples were pooled in equimolar concentrations. Finally, 16S rRNA amplicon libraries were sequenced on a MiSeq Illumina® sequencing platform (Illumina, San Diego, CA, USA) using a MiSeq Reagent Nano Kit v2 cartridge for a 2x250 paired-end sequencing.

5.2.5 Raw sequence processing

MAM and LAM raw sequences were processed using the software MicrobAT Suite v1.2.1 (SmartSeq srl, Novara, Italy), based on the Ribosomal Database Project (RDP) database. It allows the user to load the FASTQ files, download the raw data of the analysis and print the personalized reports of the samples. The first step is a cleaning of the reads obtained from the FASTQ files using custom algorithms that remove the short sequences (read length < than 200 nt) and sequences with a low quality (average Phred quality score[167] < than 25). Remained sequences are then aligned using the reference database[168]. Finally, the software generates three files (OTU, taxonomy, metadata) used as input for the statistical analyses[169,170]. Statistical analysis was performed using MicrobiomeAnalyst software (Comprehensive Statistical, Visual, and Meta-Analysis of Microbiome data)[171]. The software performs a data integrity check to show the information collected and removes all the taxa having zero reads across all the samples or appearing in only one sample. Finally, it applies a low-count filter to remove taxa containing less than 4 reads in at least 20% of samples.

5.2.6 Metabolome analysis

Mucosa-associated metabolome was collected brushing a dry swab on the polyp surface and small molecules were extracted and analyzed as reported in our previous validated method[153]. Briefly, short chain fatty acid (SCFAs) extraction was performed using water and sonication and then liquid-liquid extraction with methyl tert-butyl ether, while other metabolites (i.e., amino acids, sugars, long fatty acids, and medium fatty acids) were extracted from the aqueous phase using methanol-isopropanol-acetonitrile.

For fecal metabolome analysis, approximately 30 mg of feces were placed in a tube, then 300 mL of water and 15 µL of propanoic acid d2 (20.4 ppm) and 15 µL of acetic acid d4 (0.1 mg/mL) were added as internal standards. The sample was homogenized with the tissue lyser for three cycles of 40s at 6.5 m/s. The tube was then incubated for 30 min at 4°C and 1000 rpm and then centrifuged for 30 min at 4 °C and 21100 x g. For SCFAs extraction, 140 µL of MTBE were added to 100 µL of water-extract, the sample was mixed for 15 min at 40 rpm and then centrifuged for 10 min at 4 °C and 21100x g.

The remaining water phase solution was then subjected to a second extraction. Briefly, small molecules were extracted from 100 µL of sample using a 1 mL mixture of acetonitrile (ACN), isopropanol (IPA), and water 3:3:2, with 5 µL of tridecanoic acid (0.5 mg/mL), 5 µL of palmitic acid d31 (0.5 mg/mL), 5 µL of stearic acid d35 (0.5 mg/mL), 3.5 µL of glycine d4 (10.07 mg/mL) as internal standards. The sample was vortexed for 15 s and centrifuged for 15 min at 20 °C at 14500 x g. Next, 1 mL of the supernatant was placed in a new tube, dried in a speed vacuum at 40 °C, and stored at -20 °C until derivatization. The samples were processed with a two steps derivatization, methoximation (20 µL of methoxamine at 80 °C for 20 min) and silylation (30 µL of N,O-Bis(trimethylsilyl)trifluoroacetamide at 80 °C for 20 min).

SCFAs and small molecules from both swab and fecal samples were analyzed by bidimensional gas chromatography mass spectrometry GCXGC/TOFMS (BT 4D, Leco Corp., St. Josef, MI, USA), as described in our previous works[153,154]. The samples were analyzed using both targeted and untargeted approaches. ChromaTOF version 5.31 was used for raw data processing and mass spectral assignment was performed by matching with NIST MS Search 2.3 libraries adding Fiehn Library. Identification of molecules was also performed using an in-house library built with commercial mix standards that contain hundreds of molecules.

5.2.7 Dietary habits

Dietary habits were evaluated using the validated European Prospective Investigation into Cancer and nutrition (EPIC) questionnaire on nutrition[164]. The questionnaire is composed of 16 categories and includes questions about 266 different items. The questionnaire is statistically analyzed, and the frequency intake for different nutrients is transformed in grams/day. We compared the consumption of the most important nutrients in normal-weight (NW) *vs* obese (OB) patients, i.e. red meat, processed meat, fruits and vegetables, total fiber, and total lipids. Moreover, the following nutrients consumptions were considered to calculate the adherence index to Mediterranean diet: pasta, vegetables, fruit, legumes, olive oil, alcohol, soft drinks, red meat, fish, potatoes, butter.

5.2.8 Statistical analyses

The chi-squared test was applied to compare the sex, the polyp localization, and the grade of dysplasia between normal-weight and obese patients. The unpaired t-test was used to determine differences between the age of the different groups of patients. A p-value < 0.05 was considered statistically significant.

For the microbiota comparison, heat tree analysis allowed to compare statistically significant differences between the two groups (NW *vs* OB). This analysis is based on a hierarchical structure of taxonomic classifications that allows to quantitatively (median abundance) and statistically (non-parametric Wilcoxon Rank Sum test) depict taxon differences among communities[172].

Linear discriminant analysis effect size (LDA-LEfSe) estimates both statistical significance and biological consistency (effect size), and it was used to identify statistically enriched taxa characterizing each group (NW *vs* OB). LEfSe uses the Kruskal-Wallis sum-rank test to identify taxa statistically different between groups, then applies LDA to calculate the effect size of the different abundant taxa. Taxa with $p < 0.05$ and an LDA score > 2 or $<- 2$ were considered able to discriminate between groups.

For metabolome analyses, partial least square discriminant analysis (PLS-DA) was applied to identify more relevant metabolites for each group, as described in Barberis *et al*[153]. A fold change, $FC > 1.3$ indicate an enrichment in the obese group while a $FC < 0.769$ indicate a downregulation in the obese group.

5.3 Results

120 patients (71 males; 49 females) carrying polyps larger than 1 centimeter were recruited before colonoscopy at the Gastroenterology Unit of the University Hospital Maggiore della Carità in Novara,

Italy, and their clinical features are shown in Table 5.1. Normal-weight and obese patients show no statistically significant differences according to sex, age, polyp histology and localization.

	Patients 120 (%)	Normal weight 75 (%)	Obese 45 (%)	p-value
Sex				
Male	71 (59.17%)	47 (62.67%)	24 (53.33%)	0.3139
Female	49 (40.83%)	28 (37.33%)	21 (46.67%)	
Age				
Mean (IQR)	62 (58-69)	61 (57.5-69.5)	62 (58-69)	0.8459
Polyp histology				
Low-grade group	66 (55%)	37 (49.3%)	29 (64.4%)	0.1072
High-grade group	54 (45%)	38 (50.7%)	16 (35.5%)	
Polyp localization				
Distal	75 (62.5%)	45 (60%)	30 (66.67%)	0.465209
Proximal	45 (37.5%)	30 (40%)	15 (33.33%)	

Table 5.1. *Clinical features of the 120 patients. IQR: interquartile range*

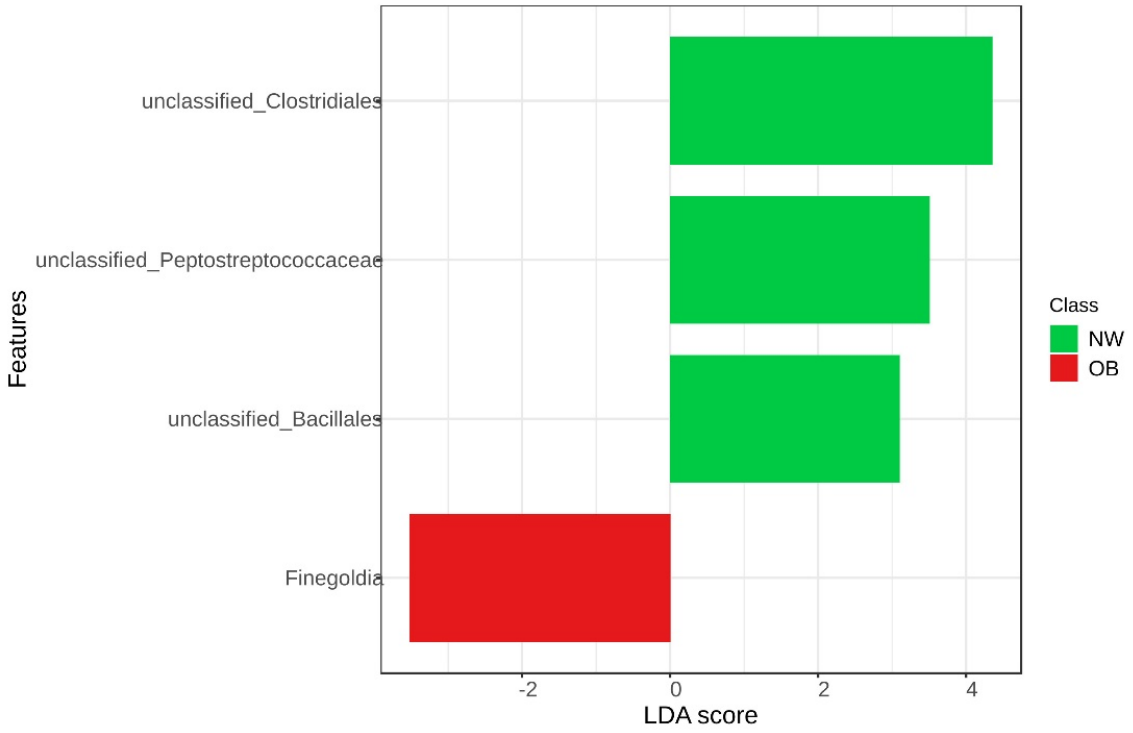
5.3.1 LAM and MAM composition

MAM was collected brushing the adenoma surface after polyp removal using an e-NAT swab, that allows to preserve the nucleic acids until extraction, while LAM was collected from fecal samples. Both MAM and LAM samples were studied using 16S rRNA sequencing and they showed an average number of reads of 34,973.53 (±25,525.67 SD) and 56,761.96 (±22,519.59 SD), respectively. Statistical analyses on these data were performed using MicrobAT and applying a low-count filter that removed taxa showing less than 4 reads. We obtained 257 and 287 taxa for MAM and LAM, respectively.

5.3.2 MAM signatures differentiate between normal-weight and obese patients

Patients were stratified according to their BMI and WC in normal-weight (NW, 75 subjects) and overweight/obese (OB, 45 subjects) (see Methods). LEfSe analyses comparing the two groups allowed to identify an enrichment of the genus *Finegoldia* in OB subjects, whereas NW patients showed an enrichment of *unclassified Clostridiales, Peptostreptococcaceae* and genera. Moreover, at species level obese patients showed an enrichment of *Finegoldia magna* and *unclassified Finegoldia,* whereas NW subjects showed an enrichment of *unclassified Clostridiales, Clostridium aff innocuum CM970, unclassified Peptostreptococcaceae, Bacteroides sp ANH 2438, Roseburia inulnivorans* and *unclassified Bacillales* (Fig. 5.1).

a)



b)

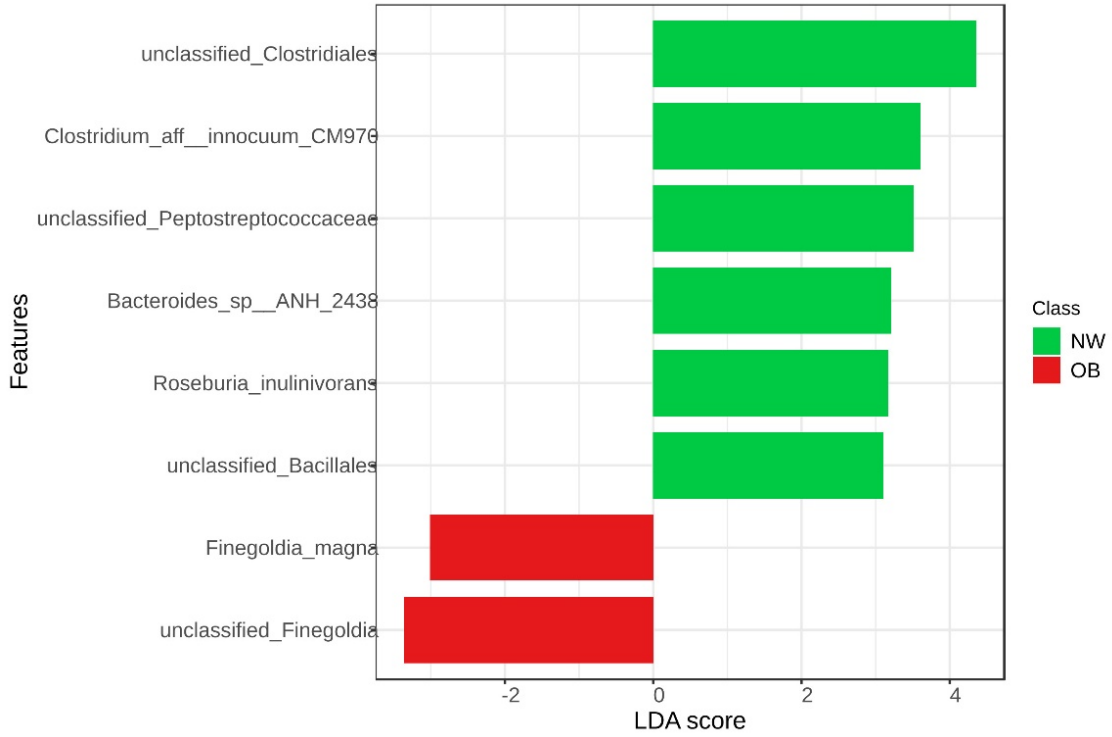


Figure 5.1. *Linear discriminant analysis effect size (LDA-LEfSe) showing (a) bacterial genera and (b) species enriched in MAM normal-weight (green, LDA score > 2) vs obese patients (red, LDA score < −2). This method incorporates statistical significance (Kruskal–Wallis) with biological consistency (effect size). The length of the bar represents a log10 transformed LDA score. This value is positive if the bacterial species is enriched in the first compared to the second group and negative if the second group shows enrichment compared to the first group. A significance level of $p < 0.05$ and an LDA score of 2 are used to determine the species best characterizing each phenotype.*

Next, we asked whether the enrichment of *F. magna* observed in obese patients could be correlated to different pathways or stages of carcinogenesis. Therefore, we repeated the previous analysis after excluding patients with hyperplastic polyps, serrated polyps and adenocarcinomas, and found that *F. magna* was associated with obesity in patients with tubular, tubulovillous or villous polyps (n=91) (Fig. 5.2). This result demonstrates that *F. magna* is not specifically correlated to adenocarcinomas, hyperplastic or serrated polyps. Moreover, we classified the subgroup of obese patients according to the grade of dysplasia of their polyps (low-grade *vs* high-grade) and observed that *F. magna* was not differently enriched (Fig. 5.3), suggesting that this bacterium is correlated with obesity regardless of the polyp stage.

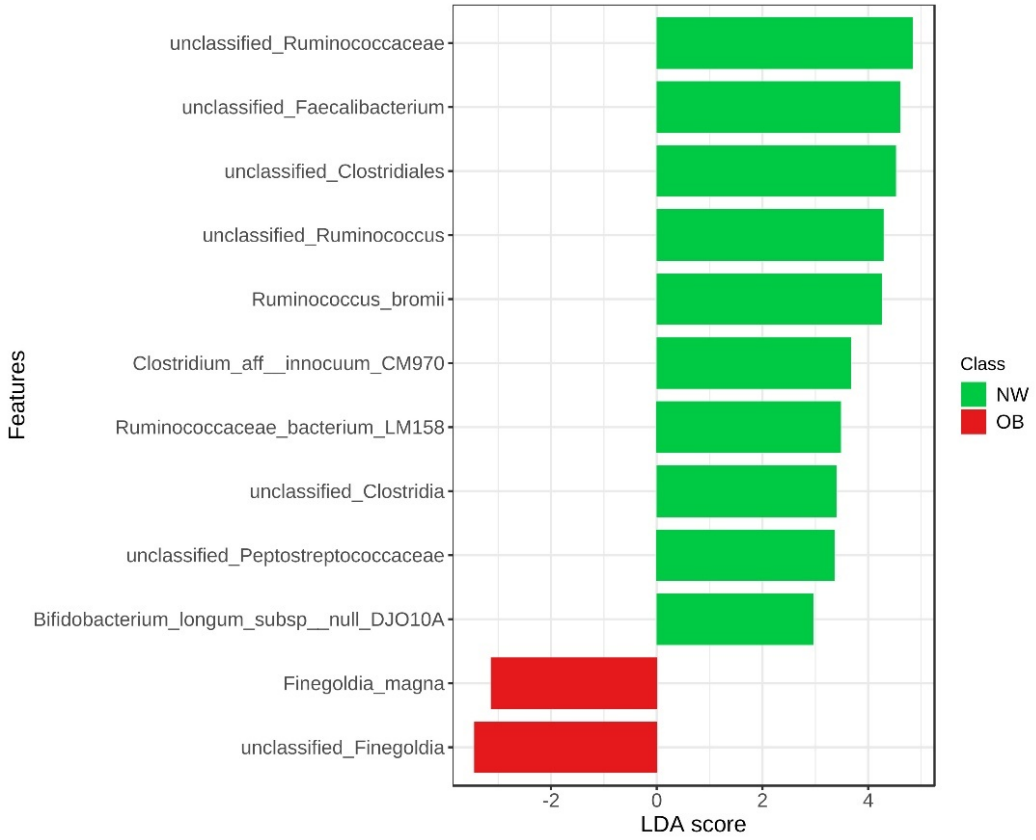


Figure 5.2. *Linear discriminant analysis effect size (LDA-LEfSe) showing bacterial species enriched in MAM of the subgroup of normal-weight (green, LDA score > 2) vs obese patients (red, LDA score < −2) with low-grade or high-grade tubular, tubulovillous or villous polyps Patients with hyperplastic polyps, serrated polyps and adenocarcinomas were excluded for this analysis.*

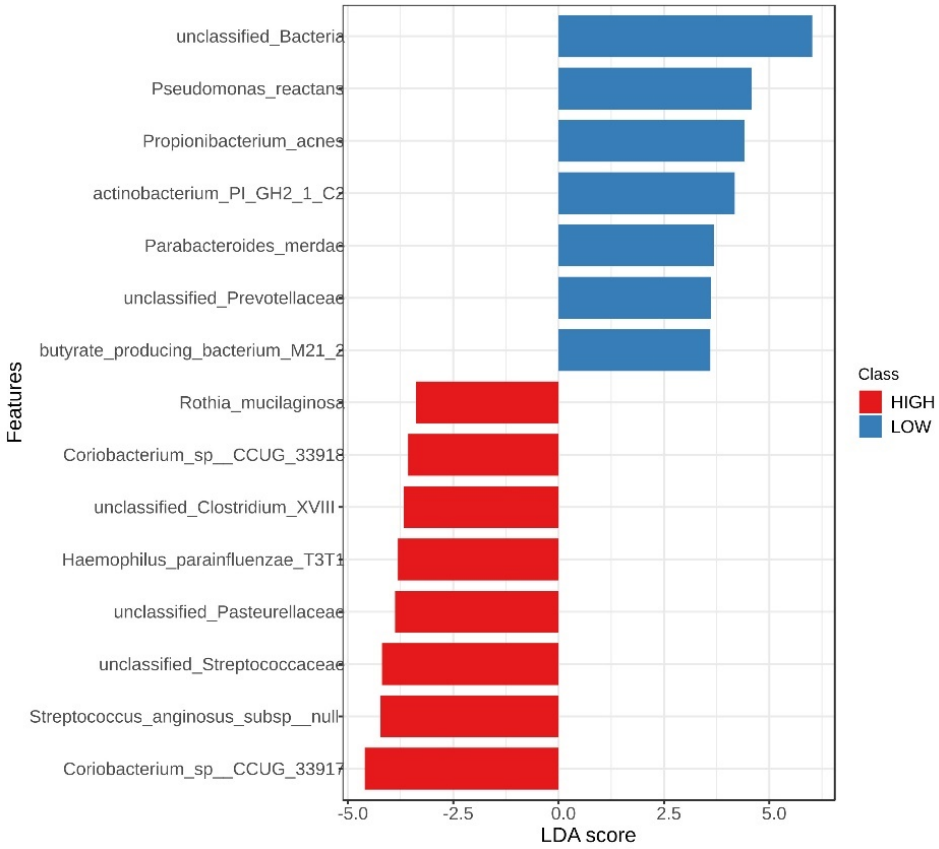


Figure 5.3. *Linear discriminant analysis effect size (LDA-LEfSe) showing bacterial species enriched in MAM of the obese patients divided according to the grade of dysplasia of their polyps (high-grade, red, LDA score > 2 vs low-grade, blue, LDA score < −2).*

5.3.3 LAM signatures differentiate between normal-weight and obese patients

LDA LEfSe analysis was carried out to identify the enriched bacteria in the LAM of normal weight or obese patients. The genera *Clostridium XIVb, Butyricicoccus, Phascolarctobacterium, Ruminococcus2, Collinsella, Erysipelotrichaceae incertae sedis* were enriched in obese subjects, whereas *unclassified Veillonellaceae* was found enriched in NW patients (Fig. 5.4).

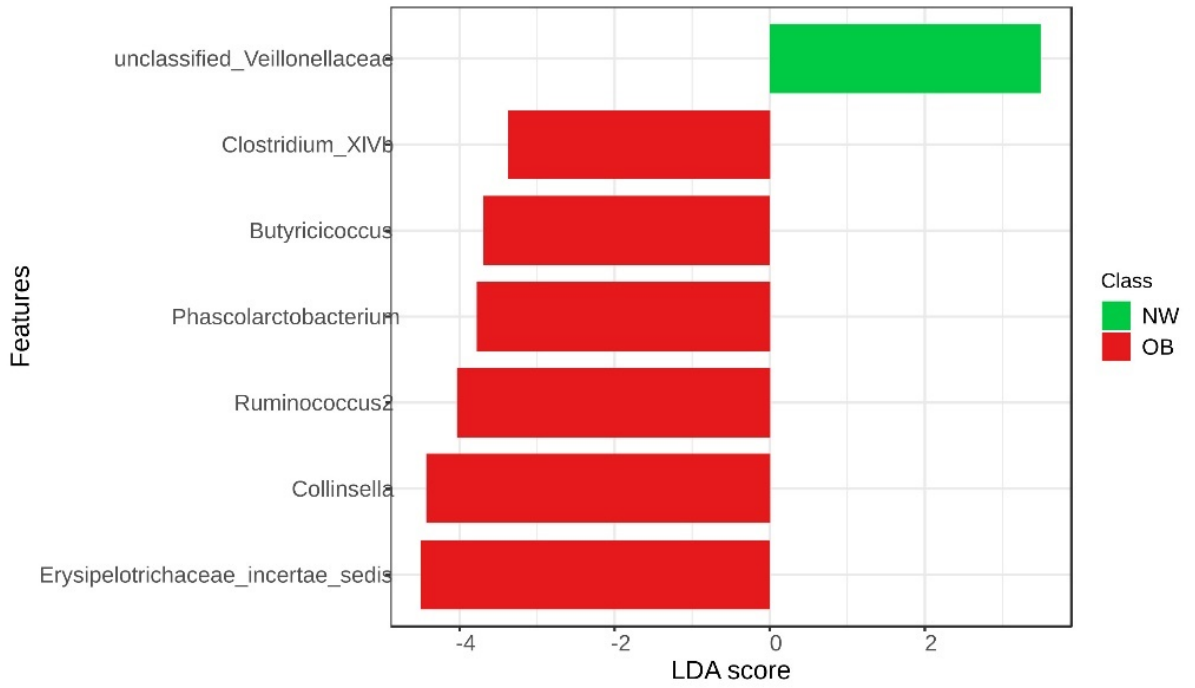


Figure 5.4. *Linear discriminant analysis effect size (LDA-LEfSe) showing bacterial genera enriched in LAM normal-weight (green, LDA score > 2) vs obese patients (red, LDA score < −2). This method incorporates statistical significance (Kruskal–Wallis) with biological consistency (effect size). The length of the bar represents a log10 transformed LDA score. This value is positive if the bacterial species is enriched in the first compared to the second group and negative if the second group shows enrichment compared to the first group. A significance level of p < 0.05 and an LDA score of 2 are used to determine the species best characterizing each phenotype.*

5.3.4 Dietary habits

Dietary habits data were obtained using a validated EPIC questionnaire[164]. Red meat, processed meat, fruits and vegetables, total fiber, and total lipids intake were evaluated, and the only statistically significant difference between the two groups of patients has been observed for processed meat (Table 5.2). The adherence index to the Mediterranean diet is not statistically different between normal-weight and obese patients (p=0.346).

Nutrients	Normal-weight (n=74)	Obese (n=45)	p-value
Dressing dips Median (IQR)	0.565 (0.1-1.9)	0. 7 (0.27-1.6)	0.52
Red meat Median (IQR)	50.995 (30.16-82.6)	72.1 (34.1-96.16)	0.1298
Processed meat Median (IQR)	11.4 (6-29.6)	22.6 (11.6-38.7)	**0.0095**
Fruit and vegetables Median (IQR)	436.25 (335.9-567.5)	441.9 (292.3-511)	0.96

Fiber Median (IQR)	19.83069 (14.85-24.93)	18.014 (12.75-20.75)	0.0588
Lipids Median (IQR)	73.84282 (58.91-96.14)	76.385 (50.85-88.88)	0.3630

Table 5.2. *Mean consumption of specific foods in "normal-weight" or "obese" patients. In bold are shown the statistically significant values (p-value < 0.05).*

Patients that carried *Finegoldia* on their polyps (27 normal-weight and 26 obese) had a statistically higher intake of processed meat compared with patients that did not (p=0.023).

5.3.5 Mucosa-associated metabolome signatures differentiate between normal-weight and obese patients

Mucosa-associated metabolome analysis allowed to identify different signatures for NW *vs* OB patients with dysplastic polyps (fold change, FC > 1.3 enriched in obese or FC < 0.769 depleted in obese; p < 0.05), as shown by the partial least square discriminant analysis (PLS-DA) plot (Fig. 5.5).

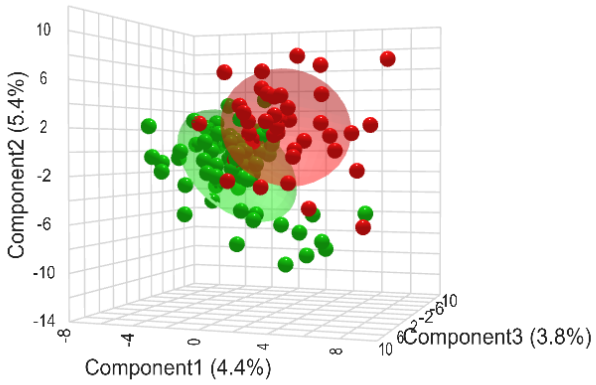


Figure 5.5. *Partial least square discriminant analysis (PLS-DA) showing different mucosa-associated metabolite distribution between normal-weight (green) and obese (red) patients.*

The hierarchical clustering heat map (Fig. 5.6) shows the distribution of the top 25 mucosa-associated metabolites that are statistically different (t-test) between NW *vs* OB patients.

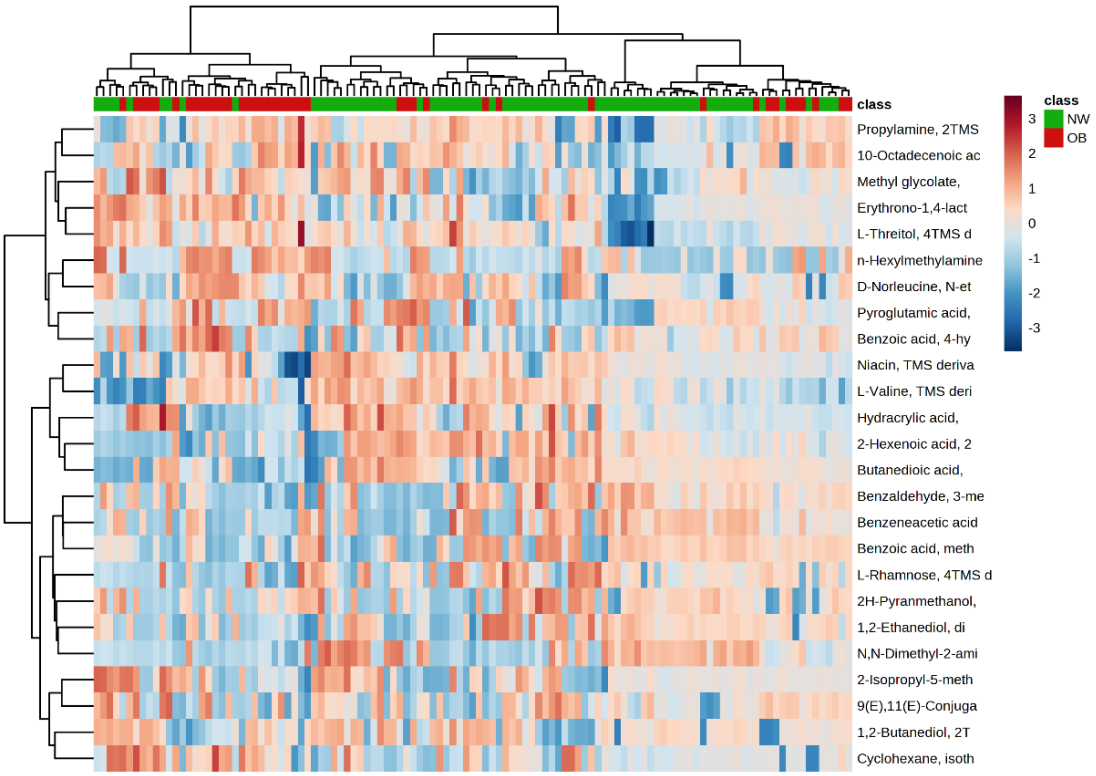


Figure 5.6. *Hierarchical clustering heat-map showing different metabolite distributions between normal-weight (green) and obese (red) patients. All the metabolites listed show a statistically significant difference between low and high-grade dysplasia groups ($p < 0.05$). Higher concentrations are reported in red, while low levels are in blue (auto-scaled data).*

In particular, in obese patients we found a higher concentration of pyroglutamic acid (FC=3.04; $p<0.01$) and propylamine (FC=15.64; $p<0.01$), whereas in normal-weight patients we observed a higher concentration of niacin (FC=0.44; $p<0.01$) and butanedioic acid (FC=0.27; $p<0.05$) (Fig. 5.7).

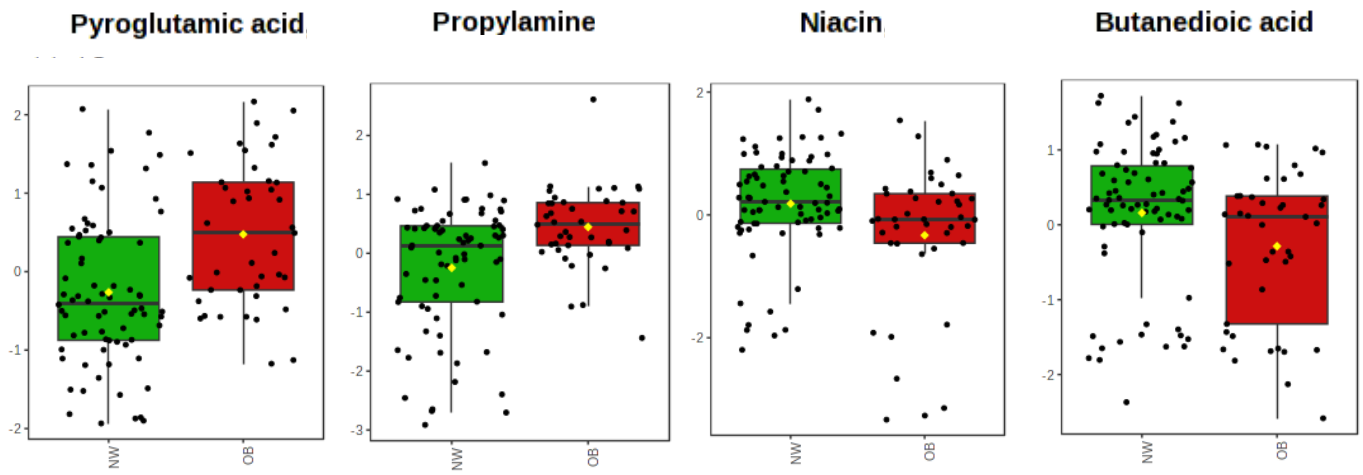


Figure 5.7. *Box plots of some of the most statistically significant mucosa-associated molecules discriminating normal-weight (green) and obese (red) patients.*

5.3.6 Luminal metabolome signatures differentiate between normal-weight and obese patients

Metabolome from fecal samples was analyzed for 79 of the 120 patients and we identified different signatures for NW *vs* OB patients with dysplastic polyps (fold change, FC > 1.3 enriched in obese or FC < 0.769 depleted in obese; $p < 0.05$), as shown by the partial least square discriminant analysis (PLS-DA) plot (Fig. 5.8).

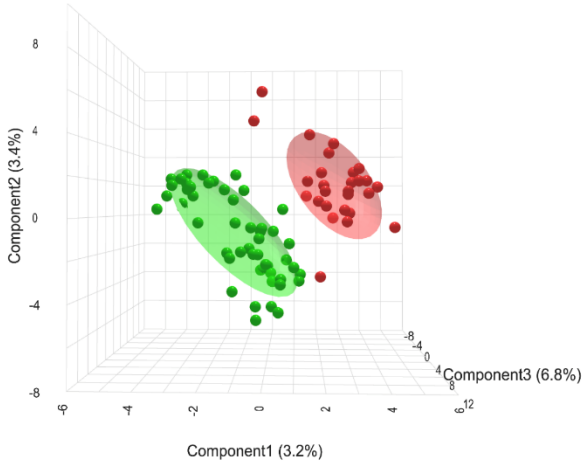


***Figure 5.8**. Partial least square discriminant analysis (PLS-DA) showing different luminal metabolite distribution between normal-weight (green) and obese (red) patients.*

In particular, in fecal samples of obese patients we found a higher concentration of indol-5-ol (FC=7.44; $p<0.01$) and hexanedioic acid (FC=1.31; $p<0.05$), whereas in normal-weight patients we observed a higher concentration of pimelic acid (FC=0.0005; $p<0.01$) (Fig. 5.9).

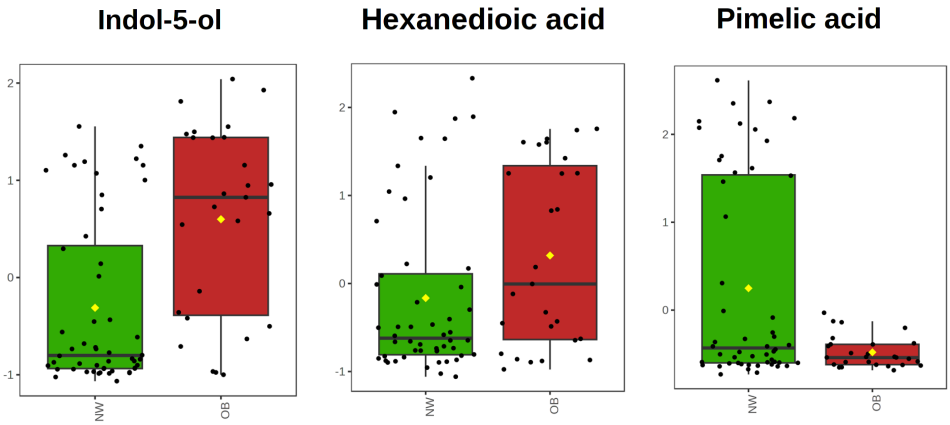


***Figure 5.9.** Box plots of some of the most statistically significant luminal metabolites discriminating normal-weight (green) and obese (red) patients.*

5.4 Discussion

In this study, we wanted to look whether obese and normal-weight patients carrying colorectal polyps were exposed to different risk factors for CRC. We analyzed 120 patients with colorectal polyps characterized for obesity, diet, gut microbiota, and CRC development. MAM and LAM composition were analyzed dividing patients in two groups (NW *vs* OB) according to their BMI and waist circumference.

Dietary habits analyses showed that obese patients with colon polyps have a higher intake of **processed meat** compared to normal-weight ones (p-value=0.0095), while fiber consumption is higher, although not statistically significant (p-value=0.0588), in normal-weight compared to obese patients (Table 5.2).
Processed meat is made mostly from pork or beef meat that undergoes treatments to increase preservation and change flavor, and includes bacon, ham and sausages[173]. It is a well-known risk factor for CRC and the risk associated with consumption of one gram of processed meat is reported to be two to ten times higher than the risk associated with one gram of fresh red meat[173]. Thus, our data confirm that CRC risk in obese patients is partly due to their nutritional habits.

Both MAM and LAM analyses revealed different signatures in normal-weight *vs* obese patients with colon polyps (Fig. 5.1 and 5.4). As we showed in a previous work, MAM, obtained with a new histology-preserving approach, and LAM differ in α/β-diversity and composition[154]. Fecal metabolome better discriminates between normal-weight and obese patients than mucosa-associated metabolome, suggesting that feces reflect diet while mucosa-associated samples are representative of polyps' microenvironment (Fig. 5.5 and 5.8).
We focused on MAM and mucosa-associated metabolites to identify key carcinogenic pathogens, since bacteria and metabolites in close contact with enterocytes are expected to play a role in colon carcinogenesis. Like fecal metabolome, LAM, although non-invasive and essential to identify CRC biomarkers, only seem to recapitulate the general gut environment[154].
MAM analysis showed an enrichment of ***Finegoldia*** genus and ***Finegoldia magna*** species in obese patients compared to normal-weight ones (Fig. 5.1). This enrichment was not found in LAM, in accordance with a study that showed a statistically significant reduction of *F. magna* in stool samples compared to rectal swabs[174].
Finegoldia magna (formerly *Peptostreptococcus magnus*) is an opportunistic Gram-positive anaerobic coccus that colonizes the skin and the mucus membranes, including the gastrointestinal one[175]. It has a high pathogenic potential due to its virulence factors[176].

FAF (*F. magna* adhesion factor) is expressed by more than 90% of *F. magna* isolates and facilitates bacterial adhesion to the host membrane[177]. FAF, together with SufA (subtilisin-like protease of *F. magna*,) and superantigen protein L, protects the bacteria from the host defenses[176]. SufA provides a proteolytic inactivation of antimicrobial peptides[178] and inhibits fibrin polymerization favoring bacterial spread to deeper tissue[179], while protein L can induce the release of pro-inflammatory mediators[180].
In Crohn's disease an increase in pro-inflammatory pathogens, including *Finegoldia*, in feces predict clinical relapse[181]. Moreover, *Finegoldia spp.* was observed in cases with colorectal adenomas, but not in controls[182].
F. magna may therefore be involved in CRC development in obese patients by neutralizing the host defenses and inducing inflammation.

We demonstrated that *F. magna* enrichment does not depend by polyp histology, because it is still enriched when we exclude from the analysis patients with adenocarcinomas, hyperplastic or serrated polyps, and its enrichment does not depend on the grade of dysplasia of the colon polyps (Fig. 5.2 and 5.3). This suggests that *F. magna* may have a role both in initiation and in progression of CRC.
Moreover, we observed a statistically significant correlation between processed meat consumption and the presence of *Finegoldia* in MAM. Interestingly, several of the patients in which *Finegoldia* has been identified are normal weight. Thus, *Finegoldia* seems to be associated with processed meat consumption. We speculate that the increased processed meat consumption favor *Finegoldia* proliferation, hence promoting polyp development, in their intestines.

Analyzing mucosa-associated metabolome of 115 patients (samples of five patients were not available), we found 24 metabolites that distinguish normal-weight and obese patients. In particular, pyroglutamic acid and propylamine had a higher concentration in obese patients compared to normal-weight ones, while niacin and butanedioic acid were enriched in normal-weight patients (Fig. 5.7).
Interestingly, orthology comparative analyses of 12 *Finegoldia* genomes inferred the presence of a sequence encoding a pyrrolidone-carboxylate peptidase in all the genomes[175]. This enzyme catalyzes the removal of **pyroglutamic acid** from peptides and proteins[183] and is thought to be involved in the degradation of exogenous polypeptides or in the detoxification of N-terminally blocked peptides[184]. These data are in accordance with our results that show a higher level of pyroglutamic acid in obese patients with colon polyps compared with normal-weight ones, and suggest that the higher abundance of *F. magna* in obese patients could be correlated with the higher amount of pyroglutamic acid. This metabolite has been found upregulated in the serum of patients with ulcerative colitis[185] and in feces

of subjects with irritable bowel disease[186]. *In vitro* experiments showed that pyroglutamic acid treatment increased IL-6 mRNA level and DNA damage, suggesting that it can promote the neoplastic transformation[185].

Niacin (known as vitamin B3, nicotinamide, or nicotinic acid) is not only acquired from diet, but may also be produced from aspartic acid or tryptophan by gut microbiota[187–189]. Its active form, nicotinamide adenine dinucleotide (NAD), is essential for many critical metabolic processes[190]. Niacin deficiency causes pellagra, a disease characterized by inflamed skin, diarrhea and/or dementia[191]. It has been reported that niacin reduces inflammation and increases adiponectin levels, an anti-inflammatory hormone, in high-fat diet-fed mice, thus attenuating obesity-induced adipose tissue inflammation[192]. Moreover, niacin is a ligand (as also butyrate) for the GPR109A receptor, which is found on the intestinal epithelium, and immune cells. GPR109A (called also HCA2) is encoded by *Niacr1*. Niacr1$^{-/-}$ mice had a decreased survival rate, bodyweight loss and are susceptible to colonic inflammation and colon cancer[101,193]. GPR109A is silenced in both mice and humans with colon cancer[194].

Interestingly, *F. magna* is predicted to be auxotrophic for vitamin B3, thus unable to synthesize it[195]. We speculate that *F. magna*, enriched in obese patients' MAM, utilizes host niacin, causing its decrement in obese patients, who are therefore susceptible to colonic inflammation and colon cancer. Other bacteria are certainly expected to utilize niacin and influence its level.

Normal-weight patients' mucosa-associated microbiota and metabolome were characterized by an enrichment of ***Roseburia inulinivorans*** and butanedioic acid (also known as succinic acid). Interestingly, mice that switched to a diet low in plant polysaccharides and fibers showed decreased luminal cecal concentrations of succinate and abundance of *R. inulinivorans*[196]. We observed a higher, but not statistically significant, fiber consumption in normal-weight patients compared to obese. These data suggest a connection among fiber consumption, higher abundance of *R. inulinivorans* and butanedioic acid in normal-weight patients.

In conclusion, this study represents the first attempt to evaluate tumor-adherent microbiota and metabolome in a large panel of obese (*vs* normal-weight) patients with colon polyps, whose diet and lifestyle habits have been collected with standardized questionnaires.

Our data support the hypothesis that different risk factors cooperate in tumorigenesis of obese and normal-weight patients. We confirm the role of dietary risk factors (e.g. processed meat in obese patients) and suggest a role of gut microbiota (e.g. *F. magna* in obese patients) and its metabolome in colorectal carcinogenesis. This is the first work to show an enrichment of *F. magna* in obese

patients carrying colorectal polyps. Functional studies are necessary to characterize the molecular bases of this association.

Chapter 6

Role of inherited predisposition and intestinal microbiota in colorectal carcinogenesis

(Preliminary data)

6.1 Introduction

Colorectal cancer (CRC) is the third most common tumor worldwide[1] and its etiology combines not modifiable – e.g. genetics – and modifiable – e.g. lifestyle and diet – risk factors[197]. Approximately 30% of the patients have family history of cancer, but a monogenic inherited cancer syndrome is identified in only 2-8%[24]. The most well-known inherited cancer syndromes causing CRC are Lynch syndrome (LS), due to germline mutations in genes of the DNA mismatch repair (MMR), and Familial Adenomatous Polyposis (FAP), caused by mutations in *APC*[1]. Among the modifiable CRC risk factors, that also include smoking, obesity, excessive fat and alcohol consumption, lack of regular physical activity[198], diet has a strong link with CRC risk and at least a proportion of its effects are probably mediated by the gut microbiota[197]. Increasing data indicate that gut dysbiosis has an important role in colon carcinogenesis and it has become clear that CRC development is accompanied by gut microbiota alterations[197,199].

The intestinal microbiota may exert a role on carcinogenesis in many ways, such as by inducing inflammation or mutagenesis in colonocytes, or through the production of metabolites[124,200], e.g. butyrate, a short fatty acid produced by bacterial fermentation of undigested fiber[109,201]. Butyrate has opposing effects on the growth of normal versus cancerous colonocytes. Butyrate, metabolized by β-oxidation and Krebs cycle, stimulates the growth of normal colonocytes by functioning as an energy source. Conversely, cancerous colonocytes inefficiently metabolize butyrate because of the Warburg effect: butyrate accumulates in CRC cells, where it acts as a histone deacetylases inhibitor, increasing the expression of pro-apoptotic molecules and reducing anti-apoptotic molecules[147]. However, contrasting effects have been reported in tumors induced by chemical carcinogens, i.e. azoxymethane (AOM) and dextran sodium sulphate (DSS), in germ free mice and $Apc^{Min/+}$ $Msh2^{-/-}$ mice, that carry germline mutations mimicking the human FAP and LS. Butyrate-producing bacteria injected in germ-free mice have a protective effect on tumors induced by AOM-DSS, through histone deacetylases inhibition[147]. Conversely, in $Apc^{Min/+}$ $Msh2^{-/-}$ mice butyrate-producing bacteria, or rectally administered butyrate, accelerate colon cell hyperproliferation. A protective effect is obtained by reducing carbohydrate intake, that decreases butyrate production[148]. It is possible that in normal conditions butyrate balances a low WNT pathway activation to induce cell proliferation. This balance is altered when β-catenin is overactivated because of a genetic defect, leading to hyperproliferation. An increase in β-catenin activation is present not only in the case of a defect of *Apc*, but also when MMR is deficient[148]. These data suggest a different effect of butyrate on carcinogenesis linked to different genetic backgrounds in mice[202] and support the hypothesis of a connection among diet, genetic predisposition and microbiota in colon carcinogenesis.

Thus, it is crucial to define the effect of specific bacteria in patients with inherited colon cancer syndromes *vs* sporadic CRC.

Several works focused on the role of gut microbiota in two hereditary CRC syndromes (LS and FAP)[203], but they analyzed almost exclusively patients CRC syndromes comparing them with apparently healthy patients, whose genome has not been characterized. Mori *et al.* found different fecal bacterial signatures in LS patients compared to healthy subjects[204], Yan and co-workers evaluated feces and biopsies from patients with LS[205], Ferrarese *et al.* identified different oral and fecal microbial signatures in LS patients *vs* normal healthy controls[206], Gonzales and colleagues compared the fecal microbiome of LS patients with cancer *vs* LS patients without cancer[207]. Two other studies focused on FAP. The first one compared the biofilm of patients with FAP and healthy controls[208], finding an enrichment of *Escherichia coli* and *Bacteroides fragilis* in the first group. The second one compared patients with highly pathogenic variants (PV) in *APC vs* those without highly PVs in *APC*, finding *Fusobacterium mortiferum* significantly increased in the stool of those with highly pathogenic *APC* variants[209].

In the present study we aimed to compare patients with colon polyps who carry germline PVs in 107 genes responsible for cancer predisposition syndromes *vs* patients with colon polyps who are not mutated. We have studied their mucosa-associated microbiota (MAM), since bacteria in close contact with enterocytes may play an important role in colon carcinogenesis. Moreover, MAM differs from fecal microbiota, that only seem to recapitulate the general gut environment[154]. For this purpose, we used a new swab-brushing technique to collect MAM without jeopardizing tissue integrity, thus preserving the polyp for histopathological analyses[154].

6.2 Methodology

6.2.1 Patients' recruitment

242 patients with colorectal polyps larger than 1 cm were recruited at colonoscopy, through a collaboration with the Gastroenterology Unit of the Ospedale Maggiore della Carità (Novara) and of the Azienda Ospedaliera Universitaria Città della Salute e della Scienza (Torino). An informed consent was signed by all participants.

6.2.2 Patients' classification

Patients were subdivided in mutated (carrying a germline pathogenic variant (PV) identified by the genetic analyses), familial (with a suspected genetic predisposition to CRC but no identified germline PVs) or sporadic (with neither PVs nor suspected genetic predisposition to CRC). In particular,

patients were classified as familial according to the following criteria, based on Syngal *et al.* [210] and Gupta *et al.* [211]:

- age of onset <50 years, or
- more than 10 synchronous or metachronous polyps, or
- diagnosis of another histologically different malignant tumor in addition to colorectal polyps, or
- presence of two 1st or 2nd grade relatives affected by cancer in the same lineage, or
- presence of at least one 1st grade relative affected by CRC.

6.2.3 Genetic analyses on germline DNA

The germline genome was evaluated to identify cancer predisposition syndromes. DNA, extracted from patient peripheral blood (QIAamp® DNA Blood Mini Kit, QIAGEN), was studied using an NGS custom panel (synthesized by Agilent Technologies) that includes 107 genes involved in inherited cancer syndromes. Variant filtering was prioritized based on sequence quality (read depth >20). Variants were filtered for MAF (minor allele frequency) <0.01 in the ExAC browser, 1000 Genomes for the European population and in the Network for Italian Genomes (http://nigdb.cineca.it/), and for impact on protein (splice-site, frameshift, stop lost/gain, missense variants). Variants were classified as pathogenic according to the American College of Medical Genetics (ACMG) guidelines. Pathogenicity of the identified variants were studied by *in silico* tools, in particular Varsome (https://varsome.com), Franklin (https://franklin.genoox.com/clinical-db/home) and Clinvar (https://www.ncbi.nlm.nih.gov/clinvar/) databases, and literature data. Pathogenic variants (PVs) were validated using Sanger sequencing. PVs were considered high or moderate-risk variants according to literature data. All the PVs that are not responsible for inherited cancer syndromes were considered moderate risk factors for CRC.

6.2.4 Genetic analyses on tumor DNA

PVs were evaluated using Sanger sequencing on DNA extracted from histological tumor sections, in collaboration with the Pathology Department of Ospedale Maggiore della Carità.

6.2.5 Microsatellite analyses

Three short tandem repeats in the region encompassing the PV of interest were investigated on patient's germline and tumor DNA. Amplified PCR products were separated with ABI 3130xl Genetic Analyzer and fragments were analyzed using GeneMapper, to look for heterozygosity or homozygosity at the microsatellite sites.

6.2.6 Microbiota analyses on tumor samples

Microbiota analyses were performed both on e-NAT™ (Copan) swabs brushed by a research nurse on the freshly removed polyps to obtain mucosa-associated microbiota (MAM) as previously described[154]. MAM was extracted using QIAamp® DNA Microbiome kit (QIAGEN) according to the manufacturer's instructions. Shotgun metagenomics sequencing was performed by Novogene (UK) Company Limited. Libraries were constructed with the NexteraXT DNA Library Preparation Kit (Illumina) and sequenced on the Illumina NovaSeq 150nt paired end platform (target depth: 7 Gb/sample). Taxonomic profiling with estimation of relative abundances at species-level was performed using MetaPhlAn version 4.0 with marker database version 201901 [212]. Downstream statistical analysis was performed through custom scripts written in the R environment. Linear discriminant analysis effect size (LEfSe) was used to identify MAM taxa that are significantly different between mutated and sporadic groups or familial/mutated and sporadic groups ($p < 0.05$).

6.2.7 Questionnaire information

Fifteen days after colonoscopy body weight, height, waist circumference and plicometry data were collected by a nutritionist. The patients were also asked whether they developed other tumor types and whether their relatives have been affected by cancer.

6.2.8 Statistical analyses

The Mann-Whitney test was used to determine differences between the age of the different groups of patients (sporadic *vs* mutated, sporadic *vs* familial, sporadic *vs* mutated+familial), while the chi-squared test was applied for other clinical characteristics, i.e., polyp histology and body mass index. A p-value < 0.05 was considered statistically significant.

6.3 Results

6.3.1 Genetic analyses

The germline genome of 242 patients was evaluated by targeted NGS to identify cancer predisposition syndromes. Since PVs are rare, we only considered variants with MAF<0.01. We also checked the variants in Clinvar, Varsome databases and literature data. We identified 29 PVs in 26 of the 242 patients (Table 6.1). All the mutated patients are heterozygous for the PVs, with the exception of s39 who is homozygous for the p.Arg443Cys variant in *BMPR1A*. We performed short tandem repeats analysis using different microsatellites in the region encompassing the *BMPR1A* p.Arg443Cys variant, but the s39 patient's germline DNA showed apparent homozygosity for all the three

microsatellites analyzed. Moreover, the NGS reads in the region around the PV did not show heterozygous sites. These results cannot exclude a deletion of one allele in the region encompassing the *BMPR1A* gene in patient s39.

Pt ID	Gender	Age	Cancer in relatives	Gene and pathogenic variants	Gene function	Hereditary cancer syndrome or risk factor
s2	F	69	father: bladder tumor, brother: liver tumor	*FANCM* c.4561_4562insTT p.Glu1521fs	HRR	Moderate risk factor, heterozygous for FA
s9	F	63	No	*GALNT12* c.907G>A p.Asp303Asn	O-glycosylation	Moderate risk factor
s21	M	64	father: pancreatic tumor, brother (s39): urethral cancer	*BMPR1A* c.1327C>T p.Arg443Cys	WNT signaling regulation Bone morphogenetic protein/TGFb signaling pathway regulation	Moderate risk factor
s31	M	56	mother: breast cancer, father: gastric and colorectal cancer	*BRIP1* c.139C>G p.Pro47Ala	HRR	Moderate risk factor, heterozygous for FA
s39	M	71	father: pancreatic tumor, brother (s21): polyps	*APC* c.6588_6589insTAA p.Ser2197* *BMPR1A* c.1327C>T p.Arg443Cys	WNT signaling regulation Bone morphogenetic protein/TGFb signaling pathway regulation	Familial adenomatous polyposis
s42	F	69	mother: leukemia	*MLH1* c.191A>G p.Asn64Ser	MMR	Moderate risk factor
s69	M	65	father: prostate and lung cancer	*RAD51C* c.93delG p.Phe32fs	HRR	Moderate risk factor, heterozygous for FA
s81	F	71	father: lung cancer	*BLM* c.2098C>T p.Gln700*	HRR	Moderate risk factor
s120	F	58	father: kidney or bladder cancer	*RAD50* c.2976_2977delAC p.His993fs	HRR	Moderate risk factor
s140	M	31	paternal grandfather: unknown tumor, father's sister: breast cancer	*CDKN2A* c.58del p.Val20*	Cell cycle regulation, apoptosis	Familial Atypical Multiple Mole Melanoma Syndrome
s141	F	56	mother: gut cancer and uterine cancer, father: lung cancer	*TSC2* c.3856C>T p.Gln1286*	Cell growth regulation	Moderate risk factor
s164	F	62	Not available	*MSH6* c.2653A>T p.Lys885*	MMR	Lynch syndrome
s176	M	71	sister: breast cancer and colon polyps, brother: colon polyps	*GALNT12* c.907G>A p.Asp303Asn	O-glycosylation	Moderate risk factor
s182	M	80	proband: prostate cancer, father: lung cancer	*FANCG* c.1183_1192del p.Glu395fs	HRR	Moderate risk factor, heterozygous for FA
s185	M	62	father: kidney and prostate cancer, maternal grandmother: throat cancer	*SBDS* c.173dup p.Asn59fs	Ribosome assembly	Moderate risk factor
s194	M	65	sister: breast cancer	*VHL* c.241C>T p.Pro81Ser	Cell cycle, regulation of hypoxia inducible factor expression, protein degradation	von Hippel-Lindau syndrome
s203	M	59	mother: breast cancer, father: prostate cancer, uncle: possible lung cancer	*NTHL1* c.244C>T p.Gln82* *BRCA1* c.181T>G p.Cys61Gly	BER HRR	Hereditary Breast and Ovarian Cancer, heterozygous for NAP
s204	F	59	father: thyroid, lung and colon cancers, maternal uncle: lung cancer	*BRCA1* c.5030_5033del p.Thr1677fs	HRR	Hereditary Breast and Ovarian Cancer
s213	M	71	proband: bladder, lung, rectal tumors, brother: hematologic cancer, sister: lung cancer, sister: colon cancer	*MUTYH* c.1103G>A p.Gly368Asp	BER	Moderate risk factor, heterozygous for MAP

s219	F	67	Not available	*PRF1* c.695G>A p.Arg232His	Immune response	Moderate risk factor, heterozygous for familial hemophagocytic lymphohistiocytosis
s242	F	69	Not available	*MUTYH* c.1103G>A p.Gly368Asp	BER	Moderate risk factor, heterozygous for MAP
s243	M	68	Not available	*FANCE* c.1111C>T p.Arg371Trp	HRR	Moderate risk factor, heterozygous for FA
s251	M	69	father: prostate cancer, maternal grandfather: possible colon cancer	*MUTYH* c.1103G>A p.Gly368Asp	BER	Moderate risk factor, heterozygous for MAP
s269	F	48	maternal grandfather: colon cancer	*FANCM* c.3286_3287insT p.Pro1096fs *MUTYH* c.1103G>A p.Gly368Asp	HRR BER	Moderate risk factor, heterozygous for FA and MAP
s293	M	58	Not available	*MSH6* c.2570A>C p.Asp857Ala	MMR	Moderate risk factor
s297	F	53	mother: colon polyps, father: myelodysplastic syndrome, maternal grandfather: lung cancer	*NTHL1* c.244C>T p.Gln82*	BER	Moderate risk factor, heterozygous for NAP

Table 6.1. *Genetic and clinical characteristics of patients carrying pathogenic variants. Abbreviations: BER: base-excision repair; HRR: homologous recombination repair; MMR: mismatch repair, FA: Fanconi anemia; MAP: MUTYH-associated polyposis; NAP: NTHL1-associated polyposis.*

6.3.1.1 Inherited cancer syndromes

Six of the mutated patients are affected by inherited cancer syndromes as they carry high penetrance PVs with well-established cancer predisposition roles[213]. In particular, patient s39 carries the nonsense PV p.Ser2197* in *APC*. This variant causes the creation of a premature stop codon and has not been previously reported in literature or databases. Variants in this domain (C-terminal of APC) are associated with attenuated FAP[214].

Patient s140 carries the variant p.Val20* in *CDKN2A*, that affects the p14 transcript, involved in cell-cycle regulation. This variant has been reported in patients with melanoma and multiple cancers[215,216]. Germline heterozygous PVs in *CDKN2A* predispose to the familial atypical multiple-mole melanoma (FAMMM) syndrome[217].

Patient s164 carries a germline high penetrance PV in a MMR gene and is therefore affected by LS. The nonsense variant p.Lys885* in exon 4 of *MSH6* leads to the creation of a premature stop codon and is expected to cause nonsense-mediated mRNA decay. This PV has been reported in a patient with Constitutional Mismatch Repair Deficiency Syndrome (CMMRD), caused by biallelic mutations in MMR genes[218], and in an individual affected by breast cancer[219].

The PV p.Pro81Ser in *VHL* has been shown to facilitate a metabolic switching, suppress apoptosis signals and reduce DNA damage response[220]. Since pVHL regulates hypoxia-inducible factors (HIF1a and HIF2a) expression, when it is mutated, HIF1a and HIF2a are not degraded and can promote cell proliferation and angiogenesis. Germline heterozygous PVs in the *VHL* gene causes the Von Hippel-Lindau disease[221], an autosomal dominant syndrome that predisposes to a wide range of

tumors such as clear cell renal cell carcinoma, pheochromocytomas that may cause hypertension, pancreatic tumors[222,223]. Two individuals affected by Von Hippel-Lindau disease have been reported to have developed CRC[224,225]. The only conditions correlated to Von Hippel-Lindau disease presented by patient s194, as far as we can tell from the questionnaires, are high blood pressure and a reported myocardial infarction[226].

Two patients (s203 and s204) are affected by Hereditary Breast and Ovarian Cancer (HBOC) syndrome, as they carry two different PVs in *BRCA1*: the frameshift mutation p.Thr1677fs, that causes the creation of a premature stop codon 2 amino acids downstream the deletion in exon 15, and the missense variant p.Cys61Gly. The first one has been described in numerous individuals with breast, ovarian[227–230] or pancreatic cancer[231]. The latter has been observed in numerous patients with breast or ovarian cancer[232–235] and functional studies showed its pathogenic effect[236]: the mutant protein reduces BRCA1/BARD1 heterodimerization and abrogates its ubiquitin ligase activity *in vitro*[237], disrupts homodimer formation[238] and is defective for double-strand break repair[239]. Germline PVs in *BRCA1* and *BRCA2* have been identified in CRC patients and a possible link between CRC risk and mutations in *BRCA1/2* has been suggested[213,240].

6.3.1.2 Moderate risk factors

We considered all the other PVs as moderate risk factors for CRC for the following reasons: 1) literature classification based on functional studies; 2) their role in cancer predisposition in the heterozygous state is reported to be lower compared to homozygous state.

Besides the PVs in *BRCA1*, we identified other five PVs in Fanconi anemia (FA) genes, whose cancer risk in the heterozygous state has not been established yet. McReynolds and colleagues found an increased risk of cancer in the heterozygous relatives of patients with FA only for *BRCA1, BRCA2, PALB2, BRIP1* or *RAD51C*[241,242]. We considered variants in *BRIP1* and *RAD51C* as moderate risk factors for CRC following Yurgelun *et al*[213].

A recent review summarizes the role of the FA pathway in CRC, reporting that germline monoallelic mutations in almost all Fanconi genes have been identified in CRC patients[243]. The variant p.Pro47Ala in *BRIP1(FANCJ)* has been reported as a loss-of-function variant because of the ATPase and helicase activity deficiency, and protein instability of the mutated protein[244,245]. Germline PVs in *BRIP1* are reported to be associated with the increased risk of developing CRC[243,246]. The variant p.Phe32Serfs*8 in *RAD51C (FANCO)* has been reported in patients with breast/ovarian cancer, and prostate cancer[247,248]. PVs in *RAD51C* have been observed in patients with early-onset CRC[249,250]. The p.Glu395fs PV in *FANCG* cause the creation of a premature stop codon in exon 10 and has been reported in FA patients[251,252]. Another variant that we found and has been observed in FA patient is

p.Arg371Trp in *FANCE*[253,254], that has been demonstrated to disrupt the interaction between FANCE and FANCD2[255]. *FANCE* has been found mutated at the germline level in two brothers with CRC[246]. In patients s2 and s269 we identified two different PVs in *FANCM*: p.Pro1096fs and p.Glu1521fs, that have never been reported in literature. Heterozygous mutations in *FANCM* have been observed in patients with CRC and breast cancer[256].

Four patients (s213, s242, s251, s269) carry a p.Gly368Asp variant in *MUTYH.* This gene encodes a DNA glycosylase that excises adenine bases inappropriately paired with guanine, cytosine, or 8-oxo-7,8-dihydroguanine, a major oxidatively damaged DNA lesion[257]. Functional studies demonstrated a reduced glycosylase and DNA binding activity for this variant compared to the wild-type[258–261]. Biallelic germline mutations in the base excision repair gene *MUTYH* are responsible for MUTYH-associated polyposis (MAP)[262]. Since our patients are heterozygous for this variant, they are not affected by MAP, but they have an increased risk for CRC[38].

The p.Gln82* variant in *NTHL1* has been found in patients s203 and s297. The *NTHL1* gene encodes a DNA glycosylase involved in the BER pathway and its biallelic mutations are responsible for NTHL1-associated polyposis (NAP). This PV has been identified in the homozygous or compound heterozygous state in multiple unrelated individuals with polyposis and/or CRC[263,264]. It has not been elucidated yet if carriers of a heterozygous PV in *NTHL1* have an increased cancer risk, like for monoallelic *MUTYH* PVs[24].

Studies on expression and MMR function for p.Asn64Ser in the *MLH1* gene[265–269] suggest a "no full mutator phenotype", and propose that the p.Asn64Ser variant alone is a low-risk variant. For this reason, we did not consider this patient as affected by LS, but as a carrier of a moderate risk factor for CRC.

The missense variant p.Asp857Ala in *MSH6*, another MMR gene, has never been reported before, so we cannot assume its pathogenic effect. Since the *in-silico* prediction tools predict a moderate pathogenic effect for this variant, we classify it as a moderate risk factor.

The variant p.Asp303Asn in *GALNT12,* carried by two of our patients (s9 and s176), has been already identified in CRC cases and functional studies have been performed on the enzymatic activity of the mutant protein, showing a partial reduction in O-glycosylation compared to the wild-type protein. GALNT12 is part of a large family of transferases involved in the initial steps of mucin-type O-glycosylation process and aberrant glycosylation is a well-described hallmark of colon cancer[51,270,271]. These studies concluded that this could be a modifier low-penetrance variant and can be involved in polyposis, most likely conferring a moderate risk of cancer[272].

One of our patients (s21) is heterozygous for the p.Arg443Cys variant in the *BMPR1A* gene, while his brother (s39), who carries also a nonsense PV in *APC*, is homozygous or has a deletion of the

wild-type allele. This *BMPR1A* variant has been reported in patients with juvenile polyposis[273,274] and colorectal cancer[213]. Howe and colleagues[275] showed that this variant leads to an impaired cellular localization – intracellular instead of on the cell membrane – compared to the wild-type protein, although no differences in protein amount were observed. *BMPR1A* is a tumor suppressor gene involved in the tumor growth factor-β signaling and its conditional inactivation in mice lead to an expansion of the stem and progenitor intestinal cells. Normal bone morphogenetic protein signaling suppresses Wnt/β-catenin signaling to ensure a balanced control of intestinal stem cell self-renewal[276].

The stopgain variant p.Gln700* in *BLM* has been reported in Italian patients with Bloom syndrome and caused a reduction in mRNA levels in cell lines derived from Bloom patients compared to normal cells[277]. Interestingly, carriers of germline PVs in *BLM* have an increased risk of developing CRC[278,279].

The variant p.His993Glnfs*6 in *RAD50* has not been previously reported in literature, but Rad50 is known to be part of the MRN complex, involved in DNA repair[280].

The variant p.Gln1286* in *TSC2* has not been previously reported in literature, but variants in this gene, a tumor suppressor that regulates cell growth, have already been associated with colon cancer[281]. We do not have clinical data on patient s141 concerning the Tuberous Sclerosis Complex disease and we considered this variant a moderate risk factors for CRC.

The p.Asn59fs variant in *SBDS* gene has never been reported. This gene encodes for a protein involved in ribosome assembly and when mutated is responsible for the Shwachman-Diamond syndrome (SDS), a rare autosomal recessive disease characterized by exocrine pancreatic insufficiency and bone marrow failure[282]. Two papers have reported an association between PVs in *SBDS* and solid tumors in two SDS patients. The first patient developed juvenile and aggressive breast cancer and did not carry *BRCA1/2* mutations[283], while the second one developed juvenile pancreatic adenocarcinoma[284].

Functional studies demonstrated that the p.Arg232His variant in *PRF1* resulted in no perforin expression in cytotoxic lymphocytes[285–287], cytotoxic activity of the 30% compared to the wild-type perforin[288] and only partial maturation of the protein[289]. Perforin is a pore-forming protein released from cytotoxic lymphocytes, allowing the granzyme-mediated death of target cells[290]. Biallelic loss-of-function mutations in *PRF1* cause familial hemophagocytic lymphohistiocytosis (FHL), but it has been hypothesized that some missense mutations causing partial loss of expression or function can result in late-onset FHL and malignancies[290,291].

6.3.1.3 Multilocus Inherited Neoplasia Allele Syndrome (MINAS)

Multi-locus Inherited Neoplasia Allele Syndrome (MINAS) refers to individuals with germline PVs in two or more cancer susceptibility genes[292]. Three patients (1.24%) carry two different PVs: s39 carries a nonsense variant in *APC* and a missense variant in *BMPR1A*, s203 a nonsense variant in *NTHL1* and a missense variant in *BRCA1*, and s269 a frameshift variant in *FANCM* and a missense variant in *MUTYH*. This percentage is in accordance with the one found by Stradella and colleagues (1.37%), who analyzed 1023 unrelated cancer patients[293]. To note, patient s39 has a worse phenotype (14 low-grade dysplastic polyps and a previous urethral cancer) compared to his brother s21 (2 low-grade dysplastic polyps). They share the moderate-risk variant in *BMPR1A*, but only patient s39 carries the nonsense variant in *APC* and is therefore affected by FAP.

6.3.2 Patients' classification

Patients were classified as follows:

- mutated, if they carry a germline PV identified by the genetic analyses reported in the previous paragraph;
- familial, if they are suspected to be genetically predisposed to CRC as described in the Methods section, but we did not identify germline PVs (they are probably carriers of multiple low-risk genetic factors[294] that go beyond the scope of this work);
- sporadic, if they do not carry any germline PV and do not meet the criteria for familiarity described in the Methods.

In Table 6.2, the clinical features of the patients and the statistical analyses among the different groups are reported. The classification in the sporadic or familial group was not feasible for 47 non-mutated patients whose information on family cancer history and previous tumors was not available.

Clinical features	**Sporadic patients (N =80)**	**Familial patients (N = 89)**	**p-value**	**Mutated patients (N = 26)**	**p-value**
Age in years Mean (±SD)	64.09 (±6.9)	60.25 (±10.36)	0.046	62.81 (±9.27)	0.86
Histology					
Adenocarcinoma	3 (3.75%)	10 (11.24%)		3 (11.54%)	
Polyps with high grade of dysplasia	34 (42.5%)	31 (34.83%)		8 (30.77%)	
Polyps with low grade of dysplasia	31 (38.75%)	33 (37.08%)	0.16	9 (34.61%)	0.38
Hyperplastic/Serrated polyps	6 (7.5%)	12 (13.48%)		4 (15.39%)	
Other	6 (7.5%)	3 (3.37%)		2 (7.69%)	

Body Mass Index					
Normal weight	40 (50%)	49 (55.06%)		15 (57.69%)	
Overweight/obese	35 (43.75%)	37 (41.57%)	0.61	7 (26.92%)	0.17
Not available	5 (6.25%)	3 (3.37%)		4 (15.38%)	

Table 6.2. *Clinical features of 195 analyzed patients (47 patients were excluded for lack of information). N, number of subjects; SD, standard deviation. The p-values are referred to the comparison with sporadic patients.*

6.3.3 Microbiota analyses

We performed two different analyses to compare mucosal-associated microbiota (MAM) of:

1) **mutated patients** – i.e. patients carrying germline PVs – *vs* **sporadic patients** – i.e. patients without germline PVs and without familiarity to CRC according to the criteria reported in the Methods. In this first analysis, we excluded familial patients, since they may carry PVs not identified by our approach or low-risk variants.
2) **mutated and familial patients** – i.e. patients with predisposition to CRC – *vs* **sporadic patients**.

We excluded 86 patients since their MAM sample was not available or not yet analyzed and 2 non-mutated patients whose information on family cancer history and previous tumors was not available (for these 2 patients the classification in the sporadic or familial group was not possible). The remaining 154 patients were categorized as sporadic, familial or mutated. Table 6.3 reports the clinical features of sporadic, familial, mutated and familial+mutated patients.

Clinical features	**Sporadic patients (N = 63)**	**Familial patients (N = 71)**	**p-value**	**Mutated patients (N = 20)**	**p-value**	**Familial+mutated patients (N = 91)**	**p-value**
Age in years Mean (±SD)	63.9 (±7.15)	60.6 (±10.58)	0.21	63.4 (±9.66)	0.65	61.2 (±10.39)	0.37
Histology							
Adenocarcinoma	3 (4.8%)	9 (12.7%)		3 (15%)		12 (13.2%)	
Polyps with high grade of dysplasia	26 (41.3%)	25 (35.2%)		8 (40%)		33 (36.3%)	
Polyps with low grade of dysplasia	23 (36.5%)	26 (36.6%)	0.47	5 (25%)	0.58	31 (34%)	0.49
Hyperplastic/Serrated polyps	6 (9.5%)	8 (11.3%)		2 (10%)		10 (11%)	
Other	5 (7.9%)	3 (4.2%)		2 (10%)		5 (5.5%)	

Body Mass Index							
Normal weight	31 (49.2%)	39 (55%)	0.57	13 (65%)	0.24	52 (57.1%)	0.38
Overweight/obese	31 (49.2%)	32 (45%)		7 (35%)		39 (42.9%)	
Not available	1 (1.6%)	0		0		0	

Table 6.3. *Clinical features of the 154 patients whose MAM has been analyzed. N, number of subjects; SD, standard deviation. The p-values are referred to the comparison with sporadic patients.*

By MAM shotgun sequencing, we identified different bacterial signatures in sporadic *vs* mutated patients (Fig. 6.1). In particular, in mutated patients we found enriched *Fusobacterium nucleatum*, one of the most studied CRC-associated bacteria[295].

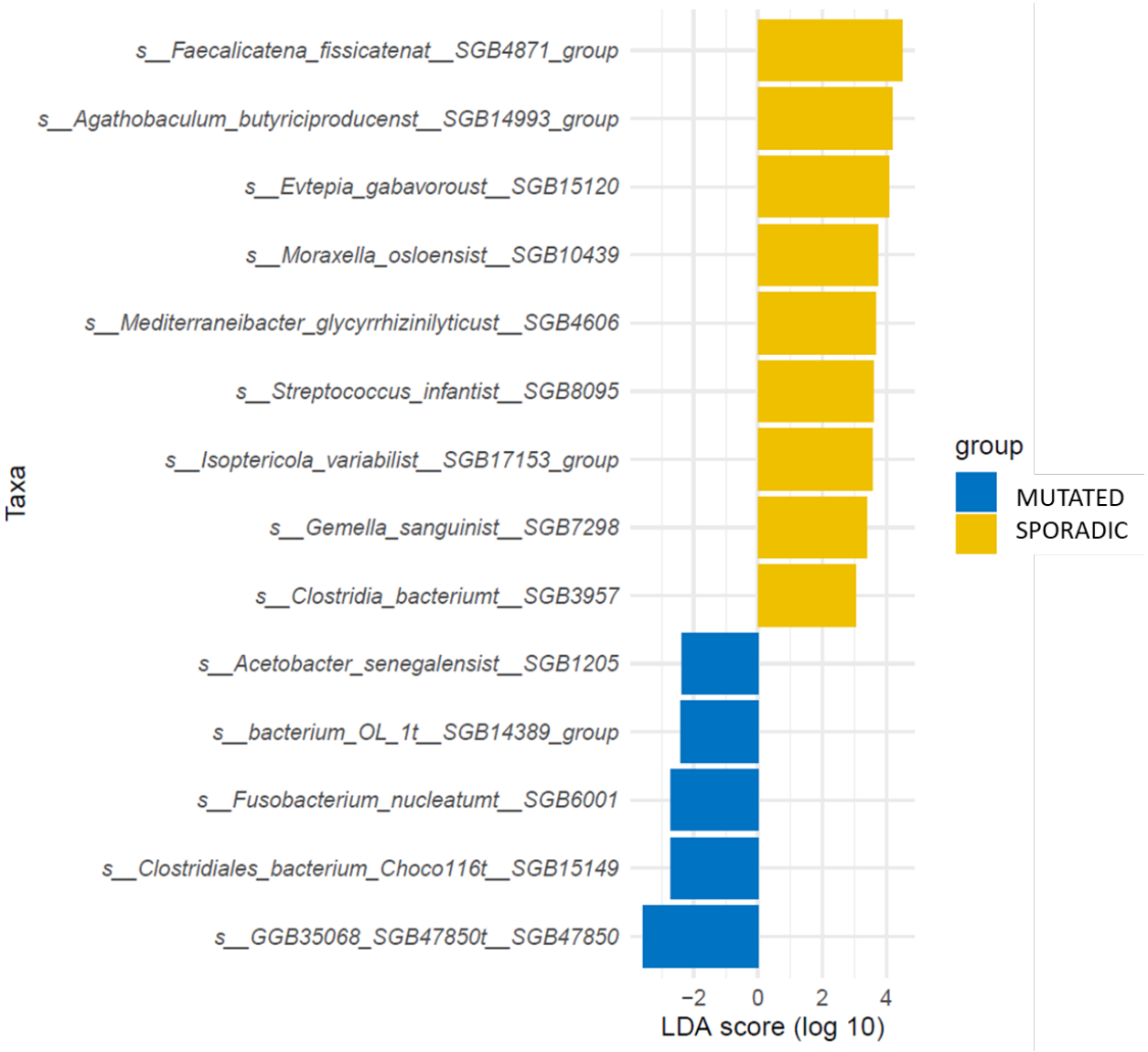


Figure 6.1. *Linear discriminant analysis effect size (LDA-LEfSe) showing bacterial species enriched in MAM of sporadic (yellow) vs mutated patients with colorectal polyps (blue).*

To note, *F. nucleatum*, which is associated to advanced stages of colorectal carcinogenesis[128], was not found enriched when we added the familial patients to the mutated ones (Fig. 6.2), though the percentage of patients with adenocarcinoma was higher in both mutated and familial groups than the sporadic one. This suggest that the enrichment of *F. nucleatum* is not due to the advanced stages of carcinogenesis, but instead to the genetic predisposition of the mutated group of patients.

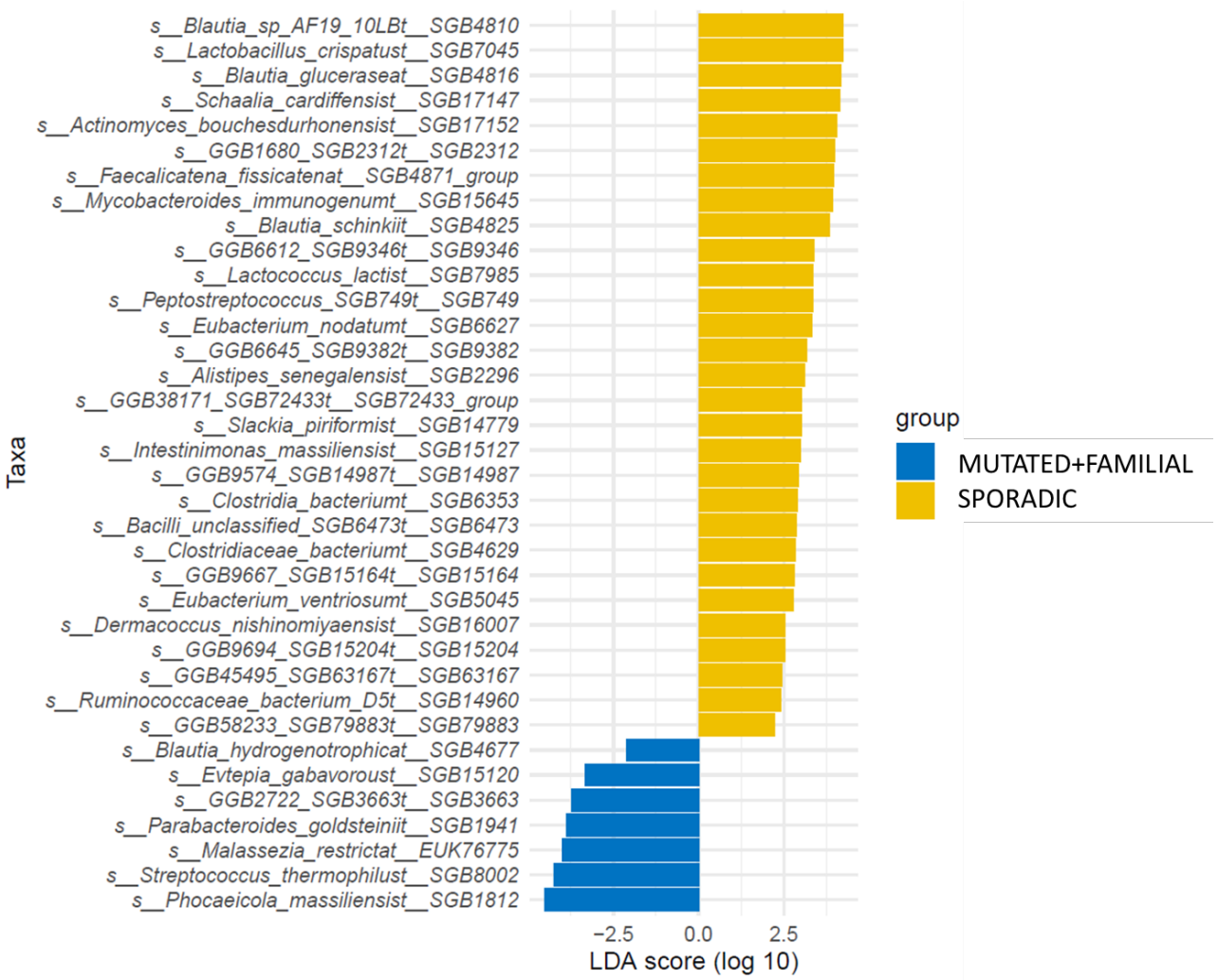


Figure 6.2. *Linear discriminant analysis effect size (LDA-LEfSe) showing bacterial species enriched in MAM of sporadic (yellow) vs mutated and familial patients with colorectal polyps (blue).*

6.4 Discussion

The aim of this work is to dissect the interaction among gut microbiota, inherited cancer predisposition and colon carcinogenesis. In particular, we wanted to evaluate whether patients with inherited risk factors showed different microbiota risk factors compared with sporadic patients. We first characterized the genetic background of 242 patients with colon polyps and then we analyzed the gut microbiota of the available patients. We found that the 10.74% of the patients (n=26) carried germline PVs conferring high or moderate risk to CRC. This percentage is higher than the one

expected for genetic predisposition to CRC (2-8%)[24], because we considered the high- and moderate-risk PVs in the incident cases without preselection, an approach similar to the one used by Yurgelun and colleagues[213] who identified high- or moderate-penetrance germline mutations in 9.9% of their patients. Conversely, previous studies focused only on high-penetrance mutations in high-risk patients with CRC (with early onset, family history, or microsatellite instability)[213].

The clinical implications of genetic testing for all the CRC patients are important in view of precision medicine-based therapies. Importantly, the Food and Drug Administration (FDA) approved immune checkpoint inhibitors (ICIs) for the treatment of microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR) tumors and as a first-line treatment for patients with MSI-H/dMMR metastatic CRC[296]. Moreover, it is known that patients with breast and ovarian cancer, but also pancreatic and prostate cancer, who carry inherited PVs in *BRCA1* or *BRCA2* show a better response to PARP inhibitors compared with patients that do not carry germline PVs[297]. Similarly, patients with CRC carrying germline PVs in *BRCA1/2* and other HRR genes may benefit from drugs that induce synthetic lethality[213].

The potential therapeutic role of gut microbiota has also been explored. The use of probiotics (i.e., live microorganisms that confer a health benefit on the host) has been shown to restore a healthy gut microbiota, improving sensitivity to chemotherapy or ICIs[203]. For this reason, the analysis of the gut microbiota and the understanding of its mechanisms and interactions are increasingly important.

The concept of a different gut microbiota in patients with sporadic or hereditary CRC is well described by Mori and Pasca[203]. However, the reported studies on microbiota analyses in hereditary CRC syndromes mostly focused on fecal samples and only on a subgroup of patients – generally LS patients – without comparing them to non-mutated individuals. Therefore, our work aims to fill this gap by comparing MAM of mutated and non-mutated patients with colorectal polyps.

By analyzing polyp-adherent microbiota, we identified different signatures of bacterial species between mutated and sporadic patients. The addition of familial patients to the mutated group changes the bacterial signatures, confirming that patients with genetic risk factors are different from those with familiarity.

In particular, we found that *Fusobacterium nucleatum* is enriched in mutated patients compared to sporadic ones. This bacterium is known to be enriched in CRC tissue compared to normal one and in the microbiome of CRC patients compared to healthy controls[197]. Moreover, it has a role in colorectal carcinogenesis with several mechanisms. *F. nucleatum* promotes tumor progression by activating the Wnt/β-catenin signaling via its FadA adhesin[141,142] and requires the expression of annexin A1, selectively expressed in cancerous colon cells, for activation of Wnt/β-catenin signaling.

Intriguingly, Rubinstein and colleagues propose the two-hit model for colorectal carcinogenesis, where the first hit is represented by the accumulation of host driver somatic mutations and the second hit by microbes activity to exacerbate cancer progression[298]. Interestingly, we found several patients carrying mutations in genes involved in the control of cell proliferation (Fig. 6.3). In these patients the germline predisposition and the higher abundance of *F. nucleatum* on their polyps may cooperate to favor cell proliferation.

Moreover, *F.nucleatum* induces DNA damage[198] through its FadA[299]. It has been proposed that *F. nucleatum* induced DNA damage in CRC through FadA-dependent activation of the E-cadherin/β-catenin pathway, leading to up-regulation of chk2 in $Apc^{Min/+}$ mice. However no causal relationship between FadA and chk2 has been established[299]. In another study on head and neck squamous cell carcinoma (HNSCC), the authors demonstrated that *F. nucleatum* promotes HNSCC proliferation by suppressing *MLH1, MSH2,* and *MSH6* expression through the upregulation of miR-205-5p[300]. It is worth noting that most of the genes that we found mutated are involved in DNA repair pathways (HRR, BER, MMR) (Fig. 6.3). Interestingly, it has recently been shown that *F. nucleatum* load in MSI CRC (including patients with LS and sporadic MSI CRC), was higher than in sporadic microsatellite stable (MSS) CRC[301]. Thus, *F. nucleatum* and genetic predisposition may cooperate to favor mutagenesis.

Another virulence factor of *F.nucleatum* is Fap2, a lectin that attaches to D-galactose-(1-3)-N-acetyl-D-galactosamine (Gal-GalNAc), the expression of which is increased in CRCs[302]. GalNAc and GalGalNAc are O-GalNAc glycans[303], which are significant components of mucins that compose the barrier between the intestinal lumen and epithelium[304].

The importance of the mucus layer abnormalities in colon carcinogenesis is underlined by the fact that defective O-glycosylation leads to impaired mucus barrier function and colitis-associated CRC in mice in which the microbiota promotes inflammation and cancer[305]. *GALNT12*, one of the gene that we found mutated (Fig. 6.3), encodes a key enzyme catalyzing the initiation of O-glycosylation[51] and the incomplete protein glycosylation process, which leads to the biosynthesis of truncated structures, occurs more often in early stages of cancer[306].

Moreover, the Fap2 protein of *F. nucleatum* can bind to the human inhibitory receptor TIGIT, reducing the killing of tumor cells by natural killer cells and suppressing host immunity[198].

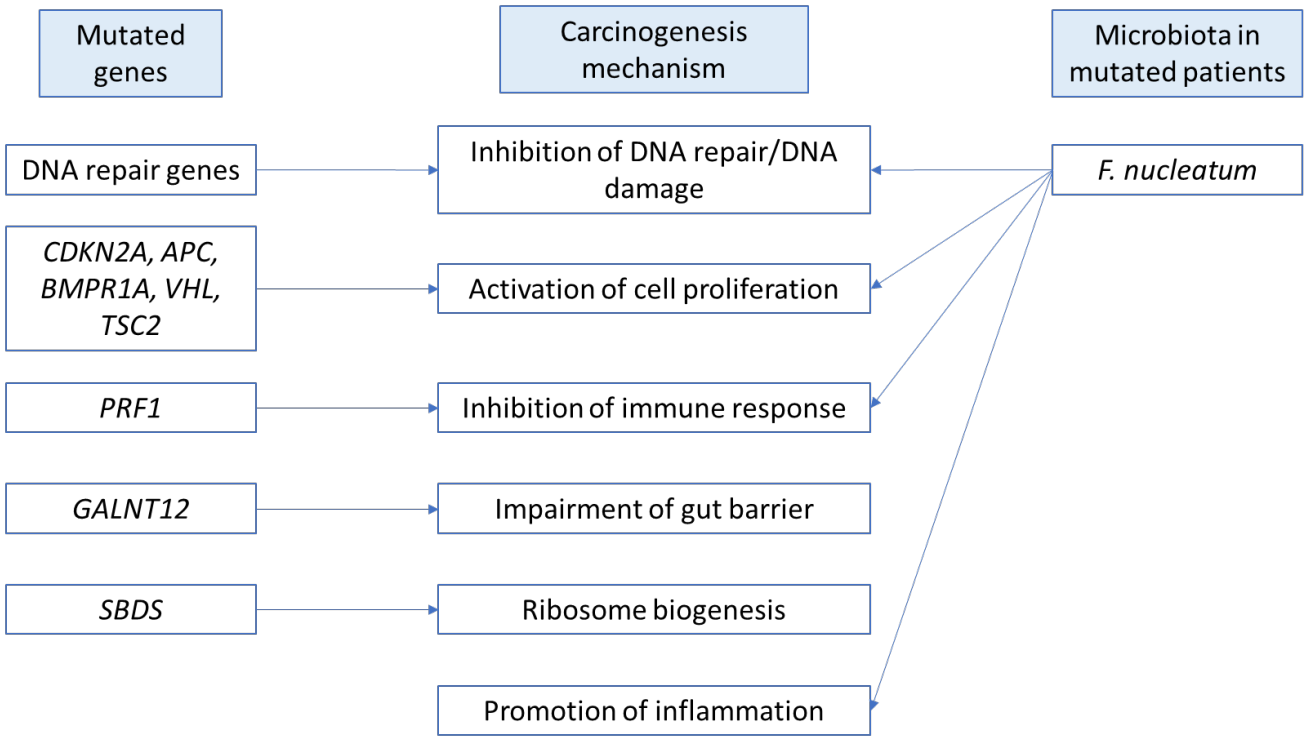


Figure 6.3. *Schematic representation of the interactions between genes the inherited risk factors and gut microbiota in colorectal carcinogenesis.*

Overall, our data support the hypothesis that tumorigenesis in patients with different germline genetic background is favored by different bacterial risk factors.

As an extension of the two-hit model proposed by Rubinstein and colleagues[298], we speculate that bacteria may cooperate with the genetic defect to promote cancer in hosts predisposed at the germline level. For example, when β-catenin is hyperactivated, as in FAP heterozygous cells, the presence of *F. nucleatum* may further increase this pathway and hence cell proliferation, resulting ultimately in loss of heterozygosity and cancer.

The perception of this interaction in the scientific community is supported by a study of the Memorial Sloan Kettering Cancer Center that is evaluating fecal microbiota and diet in patients with either LS or hereditary colonic polyposis syndromes, to evaluate the role of gut bacteria in colorectal cancer risk in addition to the hereditary risk (Clinicaltrials.gov ID: NCT02371135).

In conclusion, we identified a subgroup of patients with colon polyps that carry germline PVs in cancer predisposing genes. Moreover, by analyzing polyp-adherent microbiota, we found different bacterial signatures in mutated *vs* sporadic patients.

Our results stress the importance of characterizing the different influence of gut microbiota in colorectal carcinogenesis according to the genetic background.

Chapter 7

Conclusions and future perspectives

We have devised a novel sampling strategy that allows the identification of key carcinogenic pathogens of the mucosa-associate microbiota (MAM) and their metabolites that are in close contact with enterocytes[153,154]. With this strategy, the polyp integrity is preserved and can be used for the necessary histological analyses and diagnostic procedures.

Our data show that different mechanisms and risk factors are involved in colorectal carcinogenesis in specific groups of patients.

In obese individuals, the alteration of the mucosa-associated microbiota (e.g. the enrichment of the pathogenic species *F. magna*) and metabolome (e.g. pyroglutamic acid increase and niacin depletion) and the dietary habits (e.g. a higher consumption of processed meat) are all factors that may cooperate in tumorigenesis. The integration analysis of microbiota and metabolome data will be performed to complete this work.

In individuals predisposed at the germline level, specific gut bacteria may cooperate with the genetic defect to promote CRC development and progression. For example, when β-catenin is hyperactivated because of a mutation in *APC*, the presence of *F. nucleatum* may further activate this pathway and favor cell proliferation and mutagenesis.

To conclude the evaluation of the role of inherited predisposition and gut microbiota (Chapter 6) in colorectal carcinogenesis, we are planning to also analyze the mucosa-associated metabolome in mutated *vs* non-mutated patients and to integrate the metabolome results with the shotgun microbiota data. Moreover, we will analyze the dietary habits of these patients to determine whether mutated and non-mutated patients have different risk factors.

Shotgun analyses are ongoing on the most recent samples, as well as on the samples previously analyzed by 16S rDNA sequencing, to reach the bacterial strain resolution.

Future studies will test the short-term effects of the bacteria of interest in an innovative gut *ex-vivo* mouse model, which models the interaction among the entire colon, the immune system and the microbiota[307,308].

An *in vivo* study using mice models of CRC carcinogenesis will investigate the long-term effects of the bacteria in colorectal development and progression.

BIBLIOGRAPHY

1. Sung, H. *et al.* Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **71**, 209–249 (2021).

2. Wang, X. *et al.* Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. *BMJ Open Gastroenterol.* **6**, e000339 (2019).

3. Song, M., Chan, A. T. & Sun, J. Influence of the Gut Microbiome, Diet, and Environment on Risk of Colorectal Cancer. *Gastroenterology* **158**, 322–340 (2020).

4. Morgan, E. *et al.* Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. *Gut* **72**, 338–344 (2023).

5. Rawla, P., Sunkara, T. & Barsouk, A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Przeglad Gastroenterol.* **14**, 89–103 (2019).

6. Keum, N. & Giovannucci, E. Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **16**, 713–732 (2019).

7. Conteduca, V., Sansonno, D., Russi, S. & Dammacco, F. Precancerous colorectal lesions (Review). *Int. J. Oncol.* **43**, 973–984 (2013).

8. Colucci, P. M., Yale, S. H. & Rall, C. J. Colorectal polyps. *Clin. Med. Res.* **1**, 261–262 (2003).

9. Simon, K. Colorectal cancer development and advances in screening. *Clin. Interv. Aging* **11**, 967–976 (2016).

10. East, J. E. *et al.* British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. *Gut* **66**, 1181–1196 (2017).

11. Testa, U., Pelosi, E. & Castelli, G. Colorectal cancer: genetic abnormalities, tumor progression, tumor heterogeneity, clonal evolution and tumor-initiating cells. *Med. Sci. Basel Switz.* **6**, 31 (2018).

12. Li, F. & Lai, M. Colorectal cancer, one entity or three. *J. Zhejiang Univ. Sci. B* **10**, 219–229 (2009).

13. Lee, G. H. *et al.* Is right-sided colon cancer different to left-sided colorectal cancer? - a systematic review. *Eur. J. Surg. Oncol. J. Eur. Soc. Surg. Oncol. Br. Assoc. Surg. Oncol.* **41**, 300–308 (2015).

14. Grady, W. M. & Carethers, J. M. Genomic and epigenetic instability in colorectal cancer pathogenesis. *Gastroenterology* **135**, 1079–1099 (2008).

15. Bian, J., Dannappel, M., Wan, C. & Firestein, R. Transcriptional Regulation of Wnt/β-Catenin Pathway in Colorectal Cancer. *Cells* **9**, 2125 (2020).

16. Pawlik, T. M., Raut, C. P. & Rodriguez-Bigas, M. A. Colorectal carcinogenesis: MSI-H versus MSI-L. *Dis. Markers* **20**, 199–206 (2004).

17. Popat, S., Hubner, R. & Houlston, R. S. Systematic review of microsatellite instability and colorectal cancer prognosis. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **23**, 609–618 (2005).

18. Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8681–8686 (1999).

19. Shen, L. *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18654–18659 (2007).

20. Mauri, G. *et al.* Early-onset colorectal cancer in young individuals. *Mol. Oncol.* **13**, 109–131 (2019).

21. Stoffel, E. M. & Kastrinos, F. Familial colorectal cancer, beyond Lynch syndrome. *Clin.*

Gastroenterol. Hepatol. Off. Clin. Pract. J. Am. Gastroenterol. Assoc. **12**, 1059–1068 (2014).

22. Olkinuora, A. P., Peltomäki, P. T., Aaltonen, L. A. & Rajamäki, K. From APC to the genetics of hereditary and familial colon cancer syndromes. *Hum. Mol. Genet.* **30**, R206–R224 (2021).

23. Tian, Y. *et al.* Calculating the Starting Age for Screening in Relatives of Patients With Colorectal Cancer Based on Data From Large Nationwide Data Sets. *Gastroenterology* **159**, 159-168.e3 (2020).

24. Valle, L. *et al.* Update on genetic predisposition to colorectal cancer and polyposis. *Mol. Aspects Med.* **69**, 10–26 (2019).

25. Lynch, H. T. *et al.* Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin. Genet.* **76**, 1–18 (2009).

26. Tutlewska, K., Lubinski, J. & Kurzawski, G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. *Hered. Cancer Clin. Pract.* **11**, 9 (2013).

27. Niessen, R. C. *et al.* Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. *Genes. Chromosomes Cancer* **48**, 737–744 (2009).

28. Idos, G. & Valle, L. Lynch Syndrome. in *GeneReviews® [Internet]* (University of Washington, Seattle, 2021).

29. Williams, M. H., Hadjinicolaou, A. V., Norton, B. C., Kader, R. & Lovat, L. B. Lynch syndrome: from detection to treatment. *Front. Oncol.* **13**, 1166238 (2023).

30. Roudko, V. *et al.* Lynch Syndrome and MSI-H Cancers: From Mechanisms to 'Off-The-Shelf' Cancer Vaccines. *Front. Immunol.* **12**, 757804 (2021).

31. Yen, T., Stanich, P. P., Axell, L. & Patel, S. G. APC-Associated Polyposis Conditions. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

32. Lung, M. S., Trainer, A. H., Campbell, I. & Lipton, L. Familial colorectal cancer. *Intern. Med. J.* **45**, 482–491 (2015).

33. Zbuk, K. M. & Eng, C. Hamartomatous polyposis syndromes. *Nat. Clin. Pract. Gastroenterol. Hepatol.* **4**, 492–502 (2007).

34. Jelsig, A. M. *et al.* Hamartomatous polyposis syndromes: a review. *Orphanet J. Rare Dis.* **9**, 101 (2014).

35. Larsen Haidle, J., MacFarland, S. P. & Howe, J. R. Juvenile Polyposis Syndrome. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

36. Yehia, L. & Eng, C. PTEN Hamartoma Tumor Syndrome. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

37. Valle, L. Genetic predisposition to colorectal cancer: where we stand and future perspectives. *World J. Gastroenterol.* **20**, 9828–9849 (2014).

38. Win, A. K. *et al.* Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. *Int. J. Cancer* **129**, 2256–2262 (2011).

39. Grolleman, J. E. *et al.* Mutational Signature Analysis Reveals NTHL1 Deficiency to Cause a Multi-tumor Phenotype. *Cancer Cell* **35**, 256-266.e5 (2019).

40. Kuiper, R. P., Nielsen, M., De Voer, R. M. & Hoogerbrugge, N. NTHL1 Tumor Syndrome. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

41. Palles, C. *et al.* Germline mutations affecting the proofreading domains of POLE and POLD1

predispose to colorectal adenomas and carcinomas. *Nat. Genet.* **45**, 136–144 (2013).

42. Xu, Y. *et al.* Comparison Between Familial Colorectal Cancer Type X and Lynch Syndrome: Molecular, Clinical, and Pathological Characteristics and Pedigrees. *Front. Oncol.* **10**, 1603 (2020).

43. Garcia, F. A. de O. *et al.* New insights on familial colorectal cancer type X syndrome. *Sci. Rep.* **12**, 2846 (2022).

44. Deshmukh, A. L. *et al.* FAN1, a DNA Repair Nuclease, as a Modifier of Repeat Expansion Disorders. *J. Huntingt. Dis.* **10**, 95–122 (2021).

45. Kaddar, T. & Carreau, M. Fanconi anemia proteins and their interacting partners: a molecular puzzle. *Anemia* **2012**, 425814 (2012).

46. Nieminen, T. T. *et al.* Germline mutation of RPS20, encoding a ribosomal protein, causes predisposition to hereditary nonpolyposis colorectal carcinoma without DNA mismatch repair deficiency. *Gastroenterology* **147**, 595-598.e5 (2014).

47. Broderick, P. *et al.* Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients-a Systematic Review. *Gastroenterology* **152**, 75-77.e4 (2017).

48. Ulirsch, J. C. *et al.* The Genetic Landscape of Diamond-Blackfan Anemia. *Am. J. Hum. Genet.* **103**, 930–947 (2018).

49. Vlachos, A. *et al.* Increased risk of colon cancer and osteogenic sarcoma in Diamond-Blackfan anemia. *Blood* **132**, 2205–2208 (2018).

50. Lipton, J. M. *et al.* Early Onset Colorectal Cancer: An Emerging Cancer Risk in Patients with Diamond Blackfan Anemia. *Genes* **13**, 56 (2021).

51. Guda, K. *et al.* Inactivating germ-line and somatic mutations in polypeptide N-acetylgalactosaminyltransferase 12 in human colon cancers. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12921–12925 (2009).

52. Clarke, E. *et al.* Inherited deleterious variants in GALNT12 are associated with CRC susceptibility. *Hum. Mutat.* **33**, 1056–1058 (2012).

53. Evans, D. R. *et al.* Evidence for GALNT12 as a moderate penetrance gene for colorectal cancer. *Hum. Mutat.* **39**, 1092–1101 (2018).

54. Schulz, E. *et al.* Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. *Nat. Commun.* **5**, 5191 (2014).

55. Arem, H. *et al.* Pre- and postdiagnosis physical activity, television viewing, and mortality among patients with colorectal cancer in the National Institutes of Health-AARP Diet and Health Study. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **33**, 180–188 (2015).

56. Nguyen, L. H. *et al.* Sedentary Behaviors, TV Viewing Time, and Risk of Young-Onset Colorectal Cancer. *JNCI Cancer Spectr.* **2**, pky073 (2018).

57. Cong, Y. J. *et al.* Association of sedentary behaviour with colon and rectal cancer: a meta-analysis of observational studies. *Br. J. Cancer* **110**, 817–826 (2014).

58. Amirsasan, R., Akbarzadeh, M. & Akbarzadeh, S. Exercise and colorectal cancer: prevention and molecular mechanisms. *Cancer Cell Int.* **22**, 247 (2022).

59. Gong, J. *et al.* A pooled analysis of smoking and colorectal cancer: timing of exposure and interactions with environmental factors. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **21**, 1974–1985 (2012).

60. Giovannucci, E. & Martínez, M. Tobacco, colorectal cancer, and adenomas: a review of the evidence. *J. Natl. Cancer Inst.* **88**, (1996).

61. Deng, T., Lyon, C. J., Bergin, S., Caligiuri, M. A. & Hsueh, W. A. Obesity, Inflammation, and Cancer. *Annu. Rev. Pathol.* **11**, 421–449 (2016).

62. Song, M. *et al.* Long-term status and change of body fat distribution, and risk of colorectal cancer: a prospective cohort study. *Int. J. Epidemiol.* **45**, 871–883 (2016).

63. Calle, E. E. & Kaaks, R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat. Rev. Cancer* **4**, 579–591 (2004).

64. Yu, G.-H., Li, S.-F., Wei, R. & Jiang, Z. Diabetes and Colorectal Cancer Risk: Clinical and Therapeutic Implications. *J. Diabetes Res.* **2022**, 1747326 (2022).

65. Zhu, Y. *et al.* Dietary patterns and colorectal cancer recurrence and survival: a cohort study. *BMJ Open* **3**, e002270 (2013).

66. Schwedhelm, C., Boeing, H., Hoffmann, G., Aleksandrova, K. & Schwingshackl, L. Effect of diet on mortality and cancer recurrence among cancer survivors: a systematic review and meta-analysis of cohort studies. *Nutr. Rev.* **74**, 737–748 (2016).

67. Mehta, R. S. *et al.* Dietary Patterns and Risk of Colorectal Cancer: Analysis by Tumor Location and Molecular Subtypes. *Gastroenterology* **152**, 1944-1953.e1 (2017).

68. Chan, D. S. M. *et al.* Red and processed meat and colorectal cancer incidence: meta-analysis of prospective studies. *PloS One* **6**, e20456 (2011).

69. Baena, R. & Salinas, P. Diet and colorectal cancer. *Maturitas* **80**, 258–264 (2015).

70. Helmus, D. S., Thompson, C. L., Zelenskiy, S., Tucker, T. C. & Li, L. Red meat-derived heterocyclic amines increase risk of colon cancer: a population-based case-control study. *Nutr. Cancer* **65**, 1141–1150 (2013).

71. Bingham, S. A., Hughes, R. & Cross, A. J. Effect of white versus red meat on endogenous N-nitrosation in the human colon and further evidence of a dose response. *J. Nutr.* **132**, 3522S-3525S (2002).

72. Hughes, R., Cross, A. J., Pollock, J. R. & Bingham, S. Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. *Carcinogenesis* **22**, 199–202 (2001).

73. Stuff, J. E., Goh, E. T., Barrera, S. L., Bondy, M. L. & Forman, M. R. Construction of an N-nitroso database for assessing dietary intake. *J. Food Compos. Anal. Off. Publ. U. N. Univ. Int. Netw. Food Data Syst.* **22**, S42–S47 (2009).

74. Cross, A. J., Pollock, J. R. A. & Bingham, S. A. Haem, not protein or inorganic iron, is responsible for endogenous intestinal N-nitrosation arising from red meat. *Cancer Res.* **63**, 2358–2360 (2003).

75. Lei, L., Zhang, J., Decker, E. A. & Zhang, G. Roles of Lipid Peroxidation-Derived Electrophiles in Pathogenesis of Colonic Inflammation and Colon Cancer. *Front. Cell Dev. Biol.* **9**, 665591 (2021).

76. Seiwert, N. *et al.* Heme oxygenase 1 protects human colonocytes against ROS formation, oxidative DNA damage and cytotoxicity induced by heme iron, but not inorganic iron. *Cell Death Dis.* **11**, 787 (2020).

77. Yang, S. *et al.* Trimethylamine N-Oxide Promotes Cell Proliferation and Angiogenesis in Colorectal Cancer. *J. Immunol. Res.* **2022**, 7043856 (2022).

78. Velasquez, M. T., Ramezani, A., Manal, A. & Raj, D. S. Trimethylamine N-Oxide: The Good, the Bad and the Unknown. *Toxins* **8**, 326 (2016).

79. Jalandra, R. *et al.* Emerging role of trimethylamine-N-oxide (TMAO) in colorectal cancer. *Appl. Microbiol. Biotechnol.* **105**, 7651–7660 (2021).

80. Aykan, N. F. Red Meat and Colorectal Cancer. *Oncol. Rev.* **9**, 288 (2015).

81. Bray, G. A., Nielsen, S. J. & Popkin, B. M. Consumption of high-fructose corn syrup in beverages may play a role in the epidemic of obesity. *Am. J. Clin. Nutr.* **79**, 537–543 (2004).

82. Joh, H.-K. *et al.* Simple Sugar and Sugar-Sweetened Beverage Intake During Adolescence and Risk of Colorectal Cancer Precursors. *Gastroenterology* **161**, 128-142.e20 (2021).

83. Zoltick, E. S. *et al.* Sugar-sweetened beverage, artificially sweetened beverage and sugar intake and colorectal cancer survival. *Br. J. Cancer* **125**, 1016–1024 (2021).

84. Goncalves, M. D. *et al.* High-fructose corn syrup enhances intestinal tumor growth in mice. *Science* **363**, 1345–1349 (2019).

85. Taylor, S. R. *et al.* Dietary fructose improves intestinal cell survival and nutrient absorption. *Nature* **597**, 263–267 (2021).

86. Giovannucci, E. Alcohol, one-carbon metabolism, and colorectal cancer: recent insights from molecular studies. *J. Nutr.* **134**, 2475S-2481S (2004).

87. Amin, P. B., Diebel, L. N. & Liberati, D. M. Dose-dependent effects of ethanol and E. coli on gut permeability and cytokine production. *J. Surg. Res.* **157**, 187–192 (2009).

88. Na, H.-K. & Lee, J. Y. Molecular Basis of Alcohol-Related Gastric and Colon Cancer. *Int. J. Mol. Sci.* **18**, 1116 (2017).

89. Shukla, P. K. *et al.* Chronic ethanol feeding promotes azoxymethane and dextran sulfate sodium-induced colonic tumorigenesis potentially by enhancing mucosal inflammation. *BMC Cancer* **16**, 189 (2016).

90. Heit, C. *et al.* Transgenic mouse models for alcohol metabolism, toxicity, and cancer. *Adv. Exp. Med. Biol.* **815**, 375–387 (2015).

91. Varela-Rey, M., Woodhoo, A., Martinez-Chantar, M.-L., Mato, J. M. & Lu, S. C. Alcohol, DNA methylation, and cancer. *Alcohol Res. Curr. Rev.* **35**, 25–35 (2013).

92. Kim, Y.-I. Nutritional epigenetics: impact of folate deficiency on DNA methylation and colon cancer susceptibility. *J. Nutr.* **135**, 2703–2709 (2005).

93. Carr, P. R. *et al.* Healthy Lifestyle Factors Associated With Lower Risk of Colorectal Cancer Irrespective of Genetic Risk. *Gastroenterology* **155**, 1805-1815.e5 (2018).

94. Chang, H., Lei, L., Zhou, Y., Ye, F. & Zhao, G. Dietary Flavonoids and the Risk of Colorectal Cancer: An Updated Meta-Analysis of Epidemiological Studies. *Nutrients* **10**, 950 (2018).

95. Li, Y., Zhang, T. & Chen, G. Y. Flavonoids and Colorectal Cancer Prevention. *Antioxid. Basel Switz.* **7**, 187 (2018).

96. González-Paramás, A. M., Ayuda-Durán, B., Martínez, S., González-Manzano, S. & Santos-Buelga, C. The Mechanisms Behind the Biological Activity of Flavonoids. *Curr. Med. Chem.* **26**, 6976–6990 (2019).

97. Powers, H. J. Interaction among folate, riboflavin, genotype, and cancer, with reference to colorectal and cervical cancer. *J. Nutr.* **135**, 2960S-2966S (2005).

98. Komatsu, S.-I., Watanabe, H., Oka, T., Tsuge, H. & Kat, N. Dietary vitamin B6 suppresses colon tumorigenesis, 8-hydroxyguanosine, 4-hydroxynonenal, and inducible nitric oxide synthase protein in azoxymethane-treated mice. *J. Nutr. Sci. Vitaminol. (Tokyo)* **48**, 65–68 (2002).

99. Matsubara, K., Matsumoto, H., Mizushina, Y., Lee, J. S. & Kato, N. Inhibitory effect of pyridoxal 5'-phosphate on endothelial cell proliferation, replicative DNA polymerase and DNA topoisomerase. *Int. J. Mol. Med.* **12**, 51–55 (2003).

100. Giovannucci, E. Epidemiologic studies of folate and colorectal neoplasia: a review. *J. Nutr.* **132**, 2350S-2355S (2002).

101. Singh, N. *et al.* Activation of Gpr109a, receptor for niacin and the commensal metabolite butyrate, suppresses colonic inflammation and carcinogenesis. *Immunity* **40**, 128–139 (2014).

102. Zeman, M. *et al.* Pleiotropic effects of niacin: Current possibilities for its clinical use. *Acta Pharm. Zagreb Croat.* **66**, 449–469 (2016).

103. D'Angelo, S., Motti, M. L. & Meccariello, R. ω-3 and ω-6 Polyunsaturated Fatty Acids, Obesity and Cancer. *Nutrients* **12**, 2751 (2020).

104. Wu, S. *et al.* Fish consumption and colorectal cancer risk in humans: a systematic review and meta-analysis. *Am. J. Med.* **125**, 551-559.e5 (2012).

105. Aune, D. *et al.* Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. *BMJ* **343**, d6617 (2011).

106. Bingham, S. A. Mechanisms and experimental and epidemiological evidence relating dietary fibre (non-starch polysaccharides) and starch to protection against large bowel cancer. *Proc. Nutr. Soc.* **49**, 153–171 (1990).

107. Bazzano, L. A. *et al.* Dietary intake of whole and refined grain breakfast cereals and weight gain in men. *Obes. Res.* **13**, 1952–1960 (2005).

108. de Munter, J. S. L., Hu, F. B., Spiegelman, D., Franz, M. & van Dam, R. M. Whole grain, bran, and germ intake and risk of type 2 diabetes: a prospective cohort study and systematic review. *PLoS Med.* **4**, e261 (2007).

109. Hamer, H. M. *et al.* Review article: the role of butyrate on colonic function. *Aliment. Pharmacol. Ther.* **27**, 104–119 (2008).

110. Tullio, V., Gasperi, V., Catani, M. V. & Savini, I. The Impact of Whole Grain Intake on Gastrointestinal Tumors: A Focus on Colorectal, Gastric, and Esophageal Cancers. *Nutrients* **13**, 81 (2020).

111. Tayyem, R. F. *et al.* Consumption of Whole Grains, Refined Cereals, and Legumes and Its Association With Colorectal Cancer Among Jordanians. *Integr. Cancer Ther.* **15**, 318–325 (2016).

112. Patrignani, P. & Patrono, C. Aspirin and Cancer. *J. Am. Coll. Cardiol.* **68**, 967–976 (2016).

113. Rashid, G. *et al.* Non-steroidal anti-inflammatory drugs and biomarkers: A new paradigm in colorectal cancer. *Front. Med.* **10**, 1130710 (2023).

114. Thursby, E. & Juge, N. Introduction to the human gut microbiota. *Biochem. J.* **474**, 1823–1836 (2017).

115. Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635–1638 (2005).

116. Singhal, R. & Shah, Y. M. Oxygen battle in the gut: Hypoxia and hypoxia-inducible factors in metabolic and inflammatory responses in the intestine. *J. Biol. Chem.* **295**, 10493–10505 (2020).

117. Belizário, J. E. & Faintuch, J. Microbiome and Gut Dysbiosis. *Exp. Suppl. 2012* **109**, 459–476 (2018).

118. Corrêa-Oliveira, R., Fachi, J. L., Vieira, A., Sato, F. T. & Vinolo, M. A. R. Regulation of immune cell function by short-chain fatty acids. *Clin. Transl. Immunol.* **5**, e73 (2016).

119. LeBlanc, J. G. *et al.* Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Curr.*

Opin. Biotechnol. **24**, 160–168 (2013).

120. Bäumler, A. J. & Sperandio, V. Interactions between the microbiota and pathogenic bacteria in the gut. *Nature* **535**, 85–93 (2016).

121. Genua, F., Raghunathan, V., Jenab, M., Gallagher, W. M. & Hughes, D. J. The Role of Gut Barrier Dysfunction and Microbiome Dysbiosis in Colorectal Cancer Development. *Front. Oncol.* **11**, 626349 (2021).

122. Chen, W., Liu, F., Ling, Z., Tong, X. & Xiang, C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PloS One* **7**, e39743 (2012).

123. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).

124. Kosumi, K., Mima, K., Baba, H. & Ogino, S. Dysbiosis of the gut microbiota and colorectal cancer: the key target of molecular pathological epidemiology. *J. Lab. Precis. Med.* **3**, 76 (2018).

125. Yu, J. *et al.* Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).

126. Kostic, A. D. *et al.* Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* **14**, 207–215 (2013).

127. Mima, K. *et al.* Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut* **65**, 1973–1980 (2016).

128. Flanagan, L. *et al.* Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* **33**, 1381–1390 (2014).

129. Mima, K. *et al.* Fusobacterium nucleatum and T Cells in Colorectal Carcinoma. *JAMA Oncol.* **1**, 653–661 (2015).

130. Wu, S., Lim, K. C., Huang, J., Saidi, R. F. & Sears, C. L. Bacteroides fragilis enterotoxin cleaves the zonula adherens protein, E-cadherin. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 14979–14984 (1998).

131. Sanfilippo, L. *et al.* Bacteroides fragilis enterotoxin induces the expression of IL-8 and transforming growth factor-beta (TGF-beta) by human colonic epithelial cells. *Clin. Exp. Immunol.* **119**, 456–463 (2000).

132. Rabizadeh, S. *et al.* Enterotoxigenic bacteroides fragilis: a potential instigator of colitis. *Inflamm. Bowel Dis.* **13**, 1475–1483 (2007).

133. Goodwin, A. C. *et al.* Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 15354–15359 (2011).

134. Hwang, S., Gwon, S.-Y., Kim, M. S., Lee, S. & Rhee, K.-J. Bacteroides fragilis Toxin Induces IL-8 Secretion in HT29/C1 Cells through Disruption of E-cadherin Junctions. *Immune Netw.* **13**, 213–217 (2013).

135. Bonnet, M. *et al.* Colonization of the human gut by E. coli and colorectal cancer risk. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **20**, 859–867 (2014).

136. Nougayrède, J.-P. *et al.* Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science* **313**, 848–851 (2006).

137. Cuevas-Ramos, G. *et al.* Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11537–11542 (2010).

138. Tjalsma, H., Boleij, A., Marchesi, J. R. & Dutilh, B. E. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. *Nat. Rev. Microbiol.* **10**, 575–582 (2012).

139. Avril, M. & DePaolo, R. W. 'Driver-passenger' bacteria and their metabolites in the pathogenesis of colorectal cancer. *Gut Microbes* **13**, 1941710 (2021).

140. Wang, N. & Fang, J.-Y. Fusobacterium nucleatum, a key pathogenic factor and microbial biomarker for colorectal cancer. *Trends Microbiol.* **31**, 159–172 (2023).

141. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* **14**, 195–206 (2013).

142. Li, X. *et al.* Fusobacterium nucleatum Promotes the Progression of Colorectal Cancer Through Cdk5-Activated Wnt/β-Catenin Signaling. *Front. Microbiol.* **11**, 545251 (2020).

143. Gur, C. *et al.* Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* **42**, 344–355 (2015).

144. Wang, Z. & Zhao, Y. Gut microbiota derived metabolites in cardiovascular health and disease. *Protein Cell* **9**, 416–431 (2018).

145. Dalal, N. *et al.* Omics technologies for improved diagnosis and treatment of colorectal cancer: Technical advancement and major perspectives. *Biomed. Pharmacother. Biomedecine Pharmacother.* **131**, 110648 (2020).

146. Dalal, N. *et al.* Gut microbiota-derived metabolites in CRC progression and causation. *J. Cancer Res. Clin. Oncol.* **147**, 3141–3155 (2021).

147. Donohoe, D. R. *et al.* A gnotobiotic mouse model demonstrates that dietary fiber protects against colorectal tumorigenesis in a microbiota- and butyrate-dependent manner. *Cancer Discov.* **4**, 1387–1397 (2014).

148. Belcheva, A. *et al.* Gut microbial metabolism drives transformation of MSH2-deficient colon epithelial cells. *Cell* **158**, 288–299 (2014).

149. Wang, Z. *et al.* Impact of chronic dietary red meat, white meat, or non-meat protein on trimethylamine N-oxide metabolism and renal excretion in healthy men and women. *Eur. Heart J.* **40**, 583–594 (2019).

150. Powolny, A., Xu, J. & Loo, G. Deoxycholate induces DNA damage and apoptosis in human colon epithelial cells expressing either mutant or wild-type p53. *Int. J. Biochem. Cell Biol.* **33**, 193–203 (2001).

151. Payne, C. M. *et al.* Deoxycholate induces mitochondrial oxidative stress and activates NF-kappaB through multiple mechanisms in HCT-116 colon epithelial cells. *Carcinogenesis* **28**, 215–222 (2007).

152. Peng, Y., Nie, Y., Yu, J. & Wong, C. C. Microbial Metabolites in Colorectal Cancer: Basic and Clinical Implications. *Metabolites* **11**, 159 (2021).

153. Barberis, E. *et al.* A new method for investigating microbiota-produced small molecules in adenomatous polyps. *Anal. Chim. Acta* **1179**, 338841 (2021).

154. Clavenna, M. G. *et al.* Distinct Signatures of Tumor-Associated Microbiota and Metabolome in Low-Grade vs. High-Grade Dysplastic Colon Polyps: Inference of Their Role in Tumor Initiation and Progression. *Cancers* **15**, 3065 (2023).

155. Yachida, S. *et al.* Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat. Med.* **25**, 968–976 (2019).

156. Chen, F. *et al.* Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. *Gut* **71**, 1315–1325 (2022).

157. Lee, P., Yacyshyn, B. R. & Yacyshyn, M. B. Gut microbiota and obesity: An opportunity to alter

obesity through faecal microbiota transplant (FMT). *Diabetes Obes. Metab.* **21**, 479–490 (2019).

158. Elinav, E. *et al.* Inflammation-induced cancer: crosstalk between tumours, immune cells and microorganisms. *Nat. Rev. Cancer* **13**, 759–771 (2013).

159. Zeng, M. Y., Inohara, N. & Nuñez, G. Mechanisms of inflammation-driven bacterial dysbiosis in the gut. *Mucosal Immunol.* **10**, 18–26 (2017).

160. Wang, Y. *et al.* Analyses of Potential Driver and Passenger Bacteria in Human Colorectal Cancer. *Cancer Manag. Res.* **12**, 11553–11561 (2020).

161. Kim, J. Y., He, F. & Karin, M. From Liver Fat to Cancer: Perils of the Western Diet. *Cancers* **13**, 1095 (2021).

162. Bolte, L. A. *et al.* Long-term dietary patterns are associated with pro-inflammatory and anti-inflammatory features of the gut microbiome. *Gut* **70**, 1287–1298 (2021).

163. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.* **6**, 6528 (2015).

164. Riboli, E. *et al.* European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr.* **5**, 1113–1124 (2002).

165. Yamauchi, M. *et al.* Assessment of colorectal cancer molecular features along bowel subsites challenges the conception of distinct dichotomy of proximal versus distal colorectum. *Gut* **61**, 847–854 (2012).

166. Do, C. *et al.* A new biomarker that predicts colonic neoplasia outcome in patients with hyperplastic colonic polyps. *Cancer Prev. Res. Phila. Pa* **5**, 675–684 (2012).

167. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).

168. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42**, D633-642 (2014).

169. Bona, E. *et al.* Climatic Zone and Soil Properties Determine the Biodiversity of the Soil Bacterial Communities Associated to Native Plants from Desert Areas of North-Central Algeria. *Microorganisms* **9**, 1359 (2021).

170. Torre, E. *et al.* A Pilot Study on Clinical Scores, Immune Cell Modulation, and Microbiota Composition in Allergic Patients with Rhinitis and Asthma Treated with a Probiotic Preparation. *Int. Arch. Allergy Immunol.* **183**, 186–200 (2022).

171. Dhariwal, A. *et al.* MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* **45**, W180–W188 (2017).

172. Foster, Z. S. L., Sharpton, T. J. & Grünwald, N. J. Metacoder: An R package for visualization and manipulation of community taxonomic diversity data. *PLoS Comput. Biol.* **13**, e1005404 (2017).

173. Santarelli, R. L., Pierre, F. & Corpet, D. E. Processed meat and colorectal cancer: a review of epidemiologic and experimental evidence. *Nutr. Cancer* **60**, 131–144 (2008).

174. Shen, T.-C. D. *et al.* The Mucosally-Adherent Rectal Microbiota Contains Features Unique to Alcohol-Related Cirrhosis. *Gut Microbes* **13**, 1987781 (2021).

175. Brüggemann, H. *et al.* Pan-genome analysis of the genus Finegoldia identifies two distinct clades, strain-specific heterogeneity, and putative virulence factors. *Sci. Rep.* **8**, 266 (2018).

176. Boyanova, L., Markovska, R. & Mitov, I. Virulence arsenal of the most pathogenic species among the

Gram-positive anaerobic cocci, Finegoldia magna. *Anaerobe* **42**, 145–151 (2016).

177. Frick, I.-M. *et al.* Identification of a novel protein promoting the colonization and survival of Finegoldia magna, a bacterial commensal and opportunistic pathogen. *Mol. Microbiol.* **70**, 695–708 (2008).

178. Karlsson, C. *et al.* SufA--a novel subtilisin-like serine proteinase of Finegoldia magna. *Microbiol. Read. Engl.* **153**, 4208–4218 (2007).

179. Karlsson, C. *et al.* SufA - a bacterial enzyme that cleaves fibrinogen and blocks fibrin network formation. *Microbiol. Read. Engl.* **155**, 238–248 (2009).

180. Genovese, A. *et al.* Immunoglobulin superantigen protein L induces IL-4 and IL-13 secretion from human Fc epsilon RI+ cells through interaction with the kappa light chains of IgE. *J. Immunol. Baltim. Md 1950* **170**, 1854–1861 (2003).

181. Buffet-Bataillon, S., Bouguen, G., Fleury, F., Cattoir, V. & Le Cunff, Y. Gut microbiota analysis for prediction of clinical relapse in Crohn's disease. *Sci. Rep.* **12**, 19929 (2022).

182. Shen, X. J. *et al.* Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes* **1**, 138–147 (2010).

183. Awadé, A. C., Cleuziat, P., Gonzalès, T. & Robert-Baudouy, J. Pyrrolidone carboxyl peptidase (Pcp): an enzyme that removes pyroglutamic acid (pGlu) from pGlu-peptides and pGlu-proteins. *Proteins* **20**, 34–51 (1994).

184. Singleton, M., Isupov, M. & Littlechild, J. X-ray structure of pyrrolidone carboxyl peptidase from the hyperthermophilic archaeon Thermococcus litoralis. *Struct. Lond. Engl. 1993* **7**, 237–244 (1999).

185. Li, M. *et al.* Discovery and Validation of Potential Serum Biomarkers with Pro-Inflammatory and DNA Damage Activities in Ulcerative Colitis: A Comprehensive Untargeted Metabolomic Study. *Metabolites* **12**, 997 (2022).

186. Ponnusamy, K., Choi, J. N., Kim, J., Lee, S.-Y. & Lee, C. H. Microbial community and metabolomic comparison of irritable bowel syndrome faeces. *J. Med. Microbiol.* **60**, 817–827 (2011).

187. Peterson, C. T., Rodionov, D. A., Osterman, A. L. & Peterson, S. N. B Vitamins and Their Role in Immune Regulation and Cancer. *Nutrients* **12**, 3380 (2020).

188. Bhatt, B. *et al.* Gpr109a Limits Microbiota-Induced IL-23 Production To Constrain ILC3-Mediated Colonic Inflammation. *J. Immunol. Baltim. Md 1950* **200**, 2905–2914 (2018).

189. Gazzaniga, F., Stebbins, R., Chang, S. Z., McPeek, M. A. & Brenner, C. Microbial NAD metabolism: lessons from comparative genomics. *Microbiol. Mol. Biol. Rev. MMBR* **73**, 529–541, Table of Contents (2009).

190. Hossain, K. S., Amarasena, S. & Mayengbam, S. B Vitamins and Their Roles in Gut Health. *Microorganisms* **10**, 1168 (2022).

191. Hegyi, J., Schwartz, R. A. & Hegyi, V. Pellagra: dermatitis, dementia, and diarrhea. *Int. J. Dermatol.* **43**, 1–5 (2004).

192. Wanders, D., Graff, E. C., White, B. D. & Judd, R. L. Niacin increases adiponectin and decreases adipose tissue inflammation in high fat diet-fed mice. *PloS One* **8**, e71285 (2013).

193. Li, Z., McCafferty, K. J. & Judd, R. L. Role of HCA2 in Regulating Intestinal Homeostasis and Suppressing Colon Carcinogenesis. *Front. Immunol.* **12**, 606384 (2021).

194. Thangaraju, M. *et al.* GPR109A is a G-protein-coupled receptor for the bacterial fermentation product butyrate and functions as a tumor suppressor in colon. *Cancer Res.* **69**, 2826–2832 (2009).

195. Sharma, V. *et al.* B-Vitamin Sharing Promotes Stability of Gut Microbial Communities. *Front. Microbiol.* **10**, 1485 (2019).

196. Kovatcheva-Datchary, P. *et al.* Simplified Intestinal Microbiota to Study Microbe-Diet-Host Interactions in a Mouse Model. *Cell Rep.* **26**, 3772-3783.e6 (2019).

197. White, M. T. & Sears, C. L. The microbial landscape of colorectal cancer. *Nat. Rev. Microbiol.* (2023) doi:10.1038/s41579-023-00973-4.

198. Karpiński, T. M., Ożarowski, M. & Stasiewicz, M. Carcinogenic microbiota and its role in colorectal cancer development. *Semin. Cancer Biol.* **86**, 420–430 (2022).

199. Scott, A. J. *et al.* International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. *Gut* **68**, 1624–1632 (2019).

200. Louis, P., Hold, G. L. & Flint, H. J. The gut microbiota, bacterial metabolites and colorectal cancer. *Nat. Rev. Microbiol.* **12**, 661–672 (2014).

201. Scheppach, W. & Weiler, F. The butyrate story: old wine in new bottles? *Curr. Opin. Clin. Nutr. Metab. Care* **7**, 563–567 (2004).

202. Li, S. K. H. & Martin, A. Mismatch Repair and Colon Cancer: Mechanisms and Therapies Explored. *Trends Mol. Med.* **22**, 274–289 (2016).

203. Mori, G. & Pasca, M. R. Gut Microbial Signatures in Sporadic and Hereditary Colorectal Cancer. *Int. J. Mol. Sci.* **22**, 1312 (2021).

204. Mori, G. *et al.* Gut Microbiota Analysis in Postoperative Lynch Syndrome Patients. *Front. Microbiol.* **10**, 1746 (2019).

205. Yan, Y. *et al.* Structure of the Mucosal and Stool Microbiome in Lynch Syndrome. *Cell Host Microbe* **27**, 585-600.e4 (2020).

206. Ferrarese, R. *et al.* Oral and Fecal Microbiota in Lynch Syndrome. *J. Clin. Med.* **9**, 2735 (2020).

207. Gonzalez, A., Kapila, N., Melendez-Rosado, J., Liang, H. & Castro-Pavia, F. An Evaluation of the Fecal Microbiome in Lynch Syndrome. *J. Gastrointest. Cancer* **52**, 365–368 (2021).

208. Dejea, C. M. *et al.* Patients with familial adenomatous polyposis harbor colonic biofilms containing tumorigenic bacteria. *Science* **359**, 592–597 (2018).

209. Liang, S. *et al.* Gut microbiome associated with APC gene mutation in patients with intestinal adenomatous polyps. *Int. J. Biol. Sci.* **16**, 135–146 (2020).

210. Syngal, S. *et al.* ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am. J. Gastroenterol.* **110**, 223–262; quiz 263 (2015).

211. Gupta, S. *et al.* Potential impact of family history-based screening guidelines on the detection of early-onset colorectal cancer. *Cancer* **126**, 3013–3020 (2020).

212. Blanco-Míguez, A. *et al.* Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).

213. Yurgelun, M. B. *et al.* Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **35**, 1086–1095 (2017).

214. Knudsen, A. L., Bisgaard, M. L. & Bülow, S. Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam. Cancer* **2**, 43–55 (2003).

215. Bruno, W. *et al.* Multiple primary melanomas (MPMs) and criteria for genetic assessment: MultiMEL,

a multicenter study of the Italian Melanoma Intergroup. *J. Am. Acad. Dermatol.* **74**, 325–332 (2016).

216. Cassone, M. *et al.* Coexistence of Two Novel Mutations in CDKN2A and PMS1 Genes in a Single Patient Identifies a New and Severe Cancer Predisposition Syndrome. *Oncomedicine* **2**, 88–92 (2017).

217. Eckerle Mize, D., Bishop, M., Resse, E. & Sluzevich, J. Familial Atypical Multiple Mole Melanoma Syndrome. in *Cancer Syndromes* (eds. Riegert-Johnson, D. L., Boardman, L. A., Hefferon, T. & Roberts, M.) (National Center for Biotechnology Information (US), Bethesda (MD), 2009).

218. Tesch, V. K. *et al.* No Overt Clinical Immunodeficiency Despite Immune Biological Abnormalities in Patients With Constitutional Mismatch Repair Deficiency. *Front. Immunol.* **9**, 1506 (2018).

219. Tedaldi, G. *et al.* Multiple-gene panel analysis in a case series of 255 women with hereditary breast and ovarian cancer. *Oncotarget* **8**, 47064–47075 (2017).

220. Desimone, M. C., Rathmell, W. K. & Threadgill, D. W. Pleiotropic effects of the trichloroethylene-associated P81S VHL mutation on metabolism, apoptosis, and ATM-mediated DNA damage response. *J. Natl. Cancer Inst.* **105**, 1355–1364 (2013).

221. Nordstrom-O'Brien, M. *et al.* Genetic analysis of von Hippel-Lindau disease. *Hum. Mutat.* **31**, 521–537 (2010).

222. Maher, E. R., Neumann, H. P. & Richard, S. von Hippel-Lindau disease: a clinical and scientific review. *Eur. J. Hum. Genet. EJHG* **19**, 617–623 (2011).

223. van Leeuwaarde, R. S., Ahmad, S., van Nesselrooij, B., Zandee, W. & Giles, R. H. Von Hippel-Lindau Syndrome. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

224. Zinnamosca, L. *et al.* von Hippel Lindau disease with colon adenocarcinoma, renal cell carcinoma and adrenal pheochromocytoma. *Intern. Med. Tokyo Jpn.* **52**, 1599–1603 (2013).

225. Heo, S. J. *et al.* A Case of von Hippel-Lindau Disease with Colorectal Adenocarcinoma, Renal Cell Carcinoma and Hemangioblastomas. *Cancer Res. Treat.* **48**, 409–414 (2016).

226. Valero, E., Rumiz, E. & Pellicer, M. Cardiac Involvement in Von Hippel-Lindau Disease. *Med. Princ. Pract. Int. J. Kuwait Univ. Health Sci. Cent.* **25**, 196–198 (2016).

227. Stoppa-Lyonnet, D. *et al.* BRCA1 sequence variations in 160 individuals referred to a breast/ovarian family cancer clinic. Institut Curie Breast Cancer Group. *Am. J. Hum. Genet.* **60**, 1021–1030 (1997).

228. Schneegans, S. M. *et al.* Validation of three BRCA1/2 mutation-carrier probability models Myriad, BRCAPRO and BOADICEA in a population-based series of 183 German families. *Fam. Cancer* **11**, 181–188 (2012).

229. Kim, H. *et al.* Characteristics and spectrum of BRCA1 and BRCA2 mutations in 3,922 Korean patients with breast and ovarian cancer. *Breast Cancer Res. Treat.* **134**, 1315–1326 (2012).

230. Abdel-Razeq, H., Al-Omari, A., Zahran, F. & Arun, B. Germline BRCA1/BRCA2 mutations among high risk breast cancer patients in Jordan. *BMC Cancer* **18**, 152 (2018).

231. Ghiorzo, P. *et al.* Contribution of germline mutations in the BRCA and PALB2 genes to pancreatic cancer in Italy. *Fam. Cancer* **11**, 41–47 (2012).

232. Górski, B. *et al.* Founder mutations in the BRCA1 gene in Polish families with breast-ovarian cancer. *Am. J. Hum. Genet.* **66**, 1963–1968 (2000).

233. Cherbal, F. *et al.* BRCA1 and BRCA2 germline mutations screening in Algerian breast/ovarian cancer families. *Dis. Markers* **28**, 377–384 (2010).

234. Uglanitsa, N. *et al.* The contribution of founder mutations in BRCA1 to breast cancer in Belarus. *Clin. Genet.* **78**, 377–380 (2010).

235. Zhang, S. *et al.* Frequencies of BRCA1 and BRCA2 mutations among 1,342 unselected patients with invasive ovarian cancer. *Gynecol. Oncol.* **121**, 353–357 (2011).

236. Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).

237. Hashizume, R. *et al.* The RING heterodimer BRCA1-BARD1 is a ubiquitin ligase inactivated by a breast cancer-derived mutation. *J. Biol. Chem.* **276**, 14537–14540 (2001).

238. Brzovic, P. S., Meza, J., King, M. C. & Klevit, R. E. The cancer-predisposing mutation C61G disrupts homodimer formation in the NH2-terminal BRCA1 RING finger domain. *J. Biol. Chem.* **273**, 7795–7799 (1998).

239. Towler, W. I. *et al.* Analysis of BRCA1 variants in double-strand break repair by homologous recombination and single-strand annealing. *Hum. Mutat.* **34**, 439–445 (2013).

240. Pearlman, R. *et al.* Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer. *JAMA Oncol.* **3**, 464–471 (2017).

241. McReynolds, L. J. *et al.* Risk of cancer in heterozygous relatives of patients with Fanconi anemia. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **24**, 245–250 (2022).

242. Mehta, P. A. & Ebens, C. Fanconi Anemia. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

243. Parsa, F. G. *et al.* Fanconi Anemia Pathway in Colorectal Cancer: A Novel Opportunity for Diagnosis, Prognosis and Therapy. *J. Pers. Med.* **12**, 396 (2022).

244. Cantor, S. *et al.* The BRCA1-associated protein BACH1 is a DNA helicase targeted by clinically relevant inactivating mutations. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2357–2362 (2004).

245. Moyer, C. L. *et al.* Rare BRIP1 Missense Alleles Confer Risk for Ovarian and Breast Cancer. *Cancer Res.* **80**, 857–867 (2020).

246. Esteban-Jurado, C. *et al.* The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. *Eur. J. Hum. Genet. EJHG* **24**, 1501–1505 (2016).

247. Pelttari, L. M. *et al.* RAD51C is a susceptibility gene for ovarian cancer. *Hum. Mol. Genet.* **20**, 3278–3288 (2011).

248. Pritchard, C. C. *et al.* Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. *N. Engl. J. Med.* **375**, 443–453 (2016).

249. Xu, T. *et al.* Germline Profiling and Molecular Characterization of Early Onset Metastatic Colorectal Cancer. *Front. Oncol.* **10**, 568911 (2020).

250. Mikaeel, R. R. *et al.* Survey of germline variants in cancer-associated genes in young adults with colorectal cancer. *Genes. Chromosomes Cancer* **61**, 105–113 (2022).

251. Demuth, I. *et al.* Spectrum of mutations in the Fanconi anaemia group G gene, FANCG/XRCC9. *Eur. J. Hum. Genet. EJHG* **8**, 861–868 (2000).

252. Auerbach, A. D. *et al.* Spectrum of sequence variation in the FANCG gene: an International Fanconi Anemia Registry (IFAR) study. *Hum. Mutat.* **21**, 158–168 (2003).

253. Ameziane, N. *et al.* Genetic subtyping of Fanconi anemia by comprehensive mutation screening. *Hum.*

Mutat. **29**, 159–166 (2008).

254. Gille, J. J. P. *et al.* Diagnosis of Fanconi Anemia: Mutation Analysis by Multiplex Ligation-Dependent Probe Amplification and PCR-Based Sanger Sequencing. *Anemia* **2012**, 603253 (2012).

255. Nookala, R. K., Hussain, S. & Pellegrini, L. Insights into Fanconi Anaemia from the structure of human FANCE. *Nucleic Acids Res.* **35**, 1638–1648 (2007).

256. Cannon-Albright, L. A. *et al.* FANCM c5791C>T stopgain mutation (rs144567652) is a familial colorectal cancer risk factor. *Mol. Genet. Genomic Med.* **8**, e1532 (2020).

257. Brinkmeyer, M. K. & David, S. S. Distinct functional consequences of MUTYH variants associated with colorectal cancer: Damaged DNA affinity, glycosylase activity and interaction with PCNA and Hus1. *DNA Repair* **34**, 39–51 (2015).

258. Ali, M. *et al.* Characterization of mutant MUTYH proteins associated with familial colorectal cancer. *Gastroenterology* **135**, 499–507 (2008).

259. Goto, M. *et al.* Adenine DNA glycosylase activity of 14 human MutY homolog (MUTYH) variant proteins found in patients with colorectal polyposis and cancer. *Hum. Mutat.* **31**, E1861-1874 (2010).

260. Ruggieri, V. *et al.* Loss of MUTYH function in human cells leads to accumulation of oxidative damage and genetic instability. *Oncogene* **32**, 4500–4508 (2013).

261. Komine, K. *et al.* Functional Complementation Assay for 47 MUTYH Variants in a MutY-Disrupted Escherichia coli Strain. *Hum. Mutat.* **36**, 704–711 (2015).

262. Nielsen, M., Infante, E. & Brand, R. MUTYH Polyposis. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

263. Kuiper, R. P., Nielsen, M., De Voer, R. M. & Hoogerbrugge, N. NTHL1 Tumor Syndrome. in *GeneReviews®* (eds. Adam, M. P. et al.) (University of Washington, Seattle, Seattle (WA), 1993).

264. Weren, R. D. A. *et al.* A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. *Nat. Genet.* **47**, 668–671 (2015).

265. Ellison, A. R., Lofing, J. & Bitter, G. A. Human MutL homolog (MLH1) function in DNA mismatch repair: a prospective screen for missense mutations in the ATPase domain. *Nucleic Acids Res.* **32**, 5321–5338 (2004).

266. Wanat, J. J., Singh, N. & Alani, E. The effect of genetic background on the function of Saccharomyces cerevisiae mlh1 alleles that correspond to HNPCC missense mutations. *Hum. Mol. Genet.* **16**, 445–452 (2007).

267. Takahashi, M. *et al.* Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. *Cancer Res.* **67**, 4595–4604 (2007).

268. Hardt, K. *et al.* Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. *Fam. Cancer* **10**, 273–284 (2011).

269. Drost, M. *et al.* A functional assay-based procedure to classify mismatch repair gene variants in Lynch syndrome. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **21**, 1486–1496 (2019).

270. Lorca, V. *et al.* Role of GALNT12 in the genetic predisposition to attenuated adenomatous polyposis syndrome. *PloS One* **12**, e0187312 (2017).

271. Fernandez, A. J. *et al.* The structure of the colorectal cancer-associated enzyme GalNAc-T12 reveals how nonconserved residues dictate its function. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 20404–20410 (2019).

272. Djursby, M. *et al.* New Pathogenic Germline Variants in Very Early Onset and Familial Colorectal

Cancer Patients. *Front. Genet.* **11**, 566266 (2020).

273. Howe, J. R. *et al.* The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations. *J. Med. Genet.* **41**, 484–491 (2004).

274. Calva-Cerqueira, D. *et al.* The rate of germline mutations and large deletions of SMAD4 and BMPR1A in juvenile polyposis. *Clin. Genet.* **75**, 79–85 (2009).

275. Howe, J. R. *et al.* BMPR1A mutations in juvenile polyposis affect cellular localization. *J. Surg. Res.* **184**, 739–745 (2013).

276. He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat. Genet.* **36**, 1117–1121 (2004).

277. German, J., Sanz, M. M., Ciocci, S., Ye, T. Z. & Ellis, N. A. Syndrome-causing mutations of the BLM gene in persons in the Bloom's Syndrome Registry. *Hum. Mutat.* **28**, 743–753 (2007).

278. de Voer, R. M. *et al.* Deleterious Germline BLM Mutations and the Risk for Early-onset Colorectal Cancer. *Sci. Rep.* **5**, 14060 (2015).

279. Gruber, S. B. *et al.* BLM heterozygosity and the risk of colorectal cancer. *Science* **297**, 2013 (2002).

280. Situ, Y., Chung, L., Lee, C. S. & Ho, V. MRN (MRE11-RAD50-NBS1) Complex in Human Cancer and Prognostic Implications in Colorectal Cancer. *Int. J. Mol. Sci.* **20**, 816 (2019).

281. Slattery, M. L. *et al.* Genetic variation in a metabolic signaling pathway and colon and rectal cancer risk: mTOR, PTEN, STK11, RPKAA1, PRKAG2, TSC1, TSC2, PI3K and Akt1. *Carcinogenesis* **31**, 1604–1611 (2010).

282. Boocock, G. R. B. *et al.* Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat. Genet.* **33**, 97–101 (2003).

283. Singh, S. A. *et al.* Breast cancer in a case of Shwachman Diamond syndrome. *Pediatr. Blood Cancer* **59**, 945–946 (2012).

284. Dhanraj, S. *et al.* Molecular characteristics of a pancreatic adenocarcinoma associated with Shwachman-Diamond syndrome. *Pediatr. Blood Cancer* **60**, 754–760 (2013).

285. Feldmann, J. *et al.* Functional consequences of perforin gene mutations in 22 patients with familial haemophagocytic lymphohistiocytosis. *Br. J. Haematol.* **117**, 965–972 (2002).

286. Zhang, K. *et al.* Hypomorphic mutations in PRF1, MUNC13-4, and STXBP2 are associated with adult-onset familial HLH. *Blood* **118**, 5794–5798 (2011).

287. Sieni, E. *et al.* Familial hemophagocytic lymphohistiocytosis may present during adulthood: clinical and genetic features of a small series. *PloS One* **7**, e44649 (2012).

288. Voskoboinik, I., Thia, M.-C. & Trapani, J. A. A functional analysis of the putative polymorphisms A91V and N252S and 22 missense perforin mutations associated with familial hemophagocytic lymphohistiocytosis. *Blood* **105**, 4700–4706 (2005).

289. Risma, K. A., Frayer, R. W., Filipovich, A. H. & Sumegi, J. Aberrant maturation of mutant perforin underlies the clinical diversity of hemophagocytic lymphohistiocytosis. *J. Clin. Invest.* **116**, 182–192 (2006).

290. Brennan, A. J., Chia, J., Trapani, J. A. & Voskoboinik, I. Perforin deficiency and susceptibility to cancer. *Cell Death Differ.* **17**, 607–615 (2010).

291. Chaudhry, M. S. *et al.* Missense mutations in the perforin (PRF1) gene as a cause of hereditary cancer predisposition. *Oncoimmunology* **5**, e1179415 (2016).

292. Whitworth, J. *et al.* Multilocus Inherited Neoplasia Alleles Syndrome: A Case Series and Review. *JAMA Oncol.* **2**, 373–379 (2016).

293. Stradella, A. *et al.* Does multilocus inherited neoplasia alleles syndrome have severe clinical expression? *J. Med. Genet.* **56**, 521–525 (2019).

294. Theodoratou, E. *et al.* Systematic meta-analyses and field synopsis of genetic association studies in colorectal cancer. *J. Natl. Cancer Inst.* **104**, 1433–1457 (2012).

295. Ou, S. *et al.* Fusobacterium nucleatum and colorectal cancer: From phenomenon to mechanism. *Front. Cell. Infect. Microbiol.* **12**, (2022).

296. Choucair, K., Radford, M., Bansal, A., Park, R. & Saeed, A. Advances in immune therapies for the treatment of microsatellite instability-high/deficient mismatch repair metastatic colorectal cancer (Review). *Int. J. Oncol.* **59**, 74 (2021).

297. Kaufman, B. *et al.* Olaparib monotherapy in patients with advanced cancer and a germline BRCA1/2 mutation. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **33**, 244–250 (2015).

298. Rubinstein, M. R. *et al.* Fusobacterium nucleatum promotes colorectal cancer by inducing Wnt/β-catenin modulator Annexin A1. *EMBO Rep.* **20**, e47638 (2019).

299. Guo, P. *et al.* FadA promotes DNA damage and progression of Fusobacterium nucleatum-induced colorectal cancer through up-regulation of chk2. *J. Exp. Clin. Cancer Res. CR* **39**, 202 (2020).

300. Hsueh, C.-Y. *et al.* Fusobacterium nucleatum impairs DNA mismatch repair and stability in patients with squamous cell carcinoma of the head and neck. *Cancer* **128**, 3170–3184 (2022).

301. Ono, T. *et al.* Fusobacterium nucleatum load in MSI colorectal cancer subtypes. *Int. J. Clin. Oncol.* **27**, 1580–1588 (2022).

302. Abed, J. *et al.* Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* **20**, 215–225 (2016).

303. Alon-Maimon, T., Mandelboim, O. & Bachrach, G. Fusobacterium nucleatum and cancer. *Periodontol. 2000* **89**, 166–180 (2022).

304. González-Morelo, K. J., Vega-Sagardía, M. & Garrido, D. Molecular Insights Into O-Linked Glycan Utilization by Gut Microbes. *Front. Microbiol.* **11**, 591568 (2020).

305. Bergstrom, K. *et al.* Defective Intestinal Mucin-Type O-Glycosylation Causes Spontaneous Colitis-Associated Cancer in Mice. *Gastroenterology* **151**, 152-164.e11 (2016).

306. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).

307. Yissachar, N. *et al.* An Intestinal Organ Culture System Uncovers a Role for the Nervous System in Microbe-Immune Crosstalk. *Cell* **168**, 1135-1148.e12 (2017).

308. Gagliardi, M. *et al.* A Gut-Ex-Vivo System to Study Gut Inflammation Associated to Inflammatory Bowel Disease (IBD). *Biology* **10**, 605 (2021).

APPENDIX A

LIST OF PUBLICATIONS during the PhD Program

1. Cugliari G, Allione A, Russo A, Catalano C, Casalone E, Guarrera S, Grosso F, Ferrante D, Sculco M, **La Vecchia M**, Pirazzini C, Libener R, Mirabelli D, Magnani C, Dianzani I, Matullo G. *New DNA Methylation Signals for Malignant Pleural Mesothelioma Risk Assessment.* Cancers (Basel). 2021 May 27;13(11):2636. doi: 10.3390/cancers13112636.

2. Barberis E, Joseph S, Amede E, Clavenna MG, **La Vecchia M**, Sculco M, Aspesi A, Occhipinti P, Robotti E, Boldorini R, Marengo E, Dianzani I, Manfredi M. *A new method for investigating microbiota-produced small molecules in adenomatous polyps*. Anal Chim Acta. 2021 Sep 22;1179:338841. doi: 10.1016/j.aca.2021.338841.

3. Pileggi S, **La Vecchia M**, Colombo E.A, Fontana L, Colapietro P, Rovina D, Morotti A, Tabano S, Porta G, Alcalay M, Gervasini C, Miozzo M, Sirchia S.M. *Cohesin Mutations Induce Chromatin Conformation Perturbation of the H19/IGF2 Imprinted Region and Gene Expression Dysregulation in Cornelia de Lange Syndrome Cell Lines*. Biomolecules. 2021 Nov 2;11(11):1622. doi: 10.3390/biom11111622.

4. Sculco M*, **La Vecchia M***, Aspesi A, Casalone E, Pinton G, Clavenna MG, Allione A, Grosso F, Libener R, Muzio A, Rena O, Baietto G, Parini S, Mirabelli D, Moro L, Magnani C, Ferrante D, Matullo G, Dianzani I. *Malignant pleural mesothelioma: germline variants may steer tailored treatment.* Eur J Cancer. 2022 Jan 12;163:44-54. doi: 10.1016/j.ejca.2021.12.023. (*co-first authors)

5. Sculco M, **La Vecchia M**, Aspesi A, Clavenna MG, Salvo M, Borgonovi G, Pittaro A, Witel G, Napoli F, Listì A, Grosso F, Libener R, Maconi A, Rena O, Boldorini R, Giachino D, Bironzo P, Maffè A, Alì G, Elefanti L, Menin C, Righi L, Tampieri C, Scagliotti GV, Dianzani C, Ferrante D, Migliore E, Magnani C, Mirabelli D, Matullo G, Dianzani I. *Diagnostics of BAP1-Tumor Predisposition Syndrome by a Multitesting Approach: A Ten-Year-Long Experience.* Diagnostics (Basel). 2022 Jul 13;12(7):1710. doi: 10.3390/diagnostics12071710.

6. Piantanida N, **La Vecchia M**, Sculco M, Talmon M, Palattella G, Kurita R, Nakamura Y, Ronchi AE, Dianzani I, Ellis SR, Fresu LG, Aspesi A. *Deficiency of Ribosomal Protein S26, which is mutated in a subset of patients with Diamond Blackfan anemia, impairs erythroid differentiation.* Front. Genet. 2022 Dec 12;13:1045236. doi: 10.3389/fgene.2022.1045236

7. Clavenna MG*, **La Vecchia M***, Sculco M, Joseph S, Barberis E, Amede E, Mellai M, Brossa S, Borgonovi G, Occhipinti P, Boldorini R, Robotti E, Azzimonti B, Bona E, Ferrante D, Manfredi M, Aspesi A, Dianzani I. *Distinct signatures of tumor-associated microbiota and metabolome in low-grade vs high-grade dysplasia colon polyps: inference of their role in tumor initiation and progression*. Cancers (Basel). 2023 Jun 6;15(12):3065. doi: 10.3390/cancers15123065 (*co-first authors)

UNIVERSITÀ DEL PIEMONTE ORIENTALE
DOTTORATO DI RICERCA
FOOD, HEALTH and LONGEVITY STUDIES

DECLARATION AND AUTHORISATION TO ANTIPLAGIARISM DETECTION

The undersigned Marta La Vecchia student of the *Food, Health and Longevity Studies* Ph.D course (XXXVI Cycle)

declares:

- to be aware that the University has adopted a web-based service to detect plagiarism through a software system called "Turnit.in",
- his/her Ph.D. thesis was submitted to Turnit.in scan and reasonably it resulted an original document, which correctly cites the literature;

acknowledges:

- his/her Ph.D. thesis can be verified by his/her Ph.D. tutor and/or Ph.D Coordinator in order to confirm its originality.

Date: 12/02/2024

Signature: