Universidade de Aveiro Departamento de Biologia
2018

CARLA FILIPA NORTE GOMES

ANÁLISE DO MICROBIOMA DE AMOSTRAS DE FEZES HUMANAS - CONTRIBUTO PARA O DIAGNÓSTICO DO CANCRO COLORETAL

ANALYSIS OF THE MICROBIOME OF HUMAN STOOL SAMPLES - APPROACHING COLORECTAL CANCER DIAGNOSIS

DECLARAÇÃO

Declaro que este relatório é integralmente da minha autoria, estando devidamente referenciadas as fontes e obras consultadas, bem como identificadas de modo claro as citações dessas obras. Não contém, por isso, qualquer tipo de plágio quer de textos publicados, qualquer que seja o meio dessa publicação, incluindo meios eletrónicos, quer de trabalhos académicos.

Universidade de Aveiro Departamento de Biologia
2018

CARLA FILIPA NORTE GOMES

ANÁLISE DO MICROBIOMA DE AMOSTRAS DE FEZES HUMANAS - CONTRIBUTO PARA O DIAGNÓSTICO DO CANCRO COLORETAL

ANALYSIS OF THE MICROBIOME OF HUMAN STOOL SAMPLES - APPROACHING COLORECTAL CANCER DIAGNOSIS

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação científica da Doutora Catarina R. Marques, Investigadora Convidada do Departamento de Biologia da Universidade de Aveiro e do CESAM.

Dedico esta dissertação à minha querida família que sempre me apoiaram e em memória do meu falecido avô paterno que inspirou-me nesta pesquisa.

o júri

presidente

Prof. Doutor Mário Guilherme Garcês Pacheco
Professor Auxiliar com Agregação do Departamento de Biologia da Universidade de Aveiro

Mestre Juan Carlos Mellídez Barroso
Médico Oncologista e Diretor do Serviço de Oncologia do Centro Hospitalar do Baixo Vouga E.P.E., Aveiro

Doutora Catarina Pires Ribeiro Ramos Marques (orientadora)
Cientista Convidada do Departamento de Biologia da Universidade de Aveiro e membro integrado do Centro de Estudos do Ambiente e do Mar (CESAM)

agradecimentos

Antes de mais nada, gostaria de expressar a minha gratidão à Catarina Marques, a minha orientadora, por ter-me aceitado na sua equipa de investigação e pela paciência, compreensão e orientação que ela forneceu ao longo deste percurso de pesquisa científica.

Gostaria de agradecer às minhas colegas de pesquisa e amigas, Mafalda e Diana, por tornarem esta experiência e o ambiente de laboratório alegres quando os momentos do nosso trabalho se tornaram mais difíceis.

Gostaria de agradecer à minha família (mãe e pai) pelo apoio infinito ao longo deste processo stressante, mas recompensador, e por acreditarem nas minhas capacidades acadêmicas. À minha irmã, Cristina, por ser o exemplo e dar-me a coragem para enfrentar o desafio de estudar num país estrangeiro.

Agradeço ao Francisco por ter-me aturado em épocas mais stressantes e por ser o meu refúgio sempre que precisasse.

Também agradeço aos voluntários que participaram neste estudo, bem como aos médicos, cirurgiões, enfermeiros e administrativos, pela excelente colaboração e disponibilidade, especialmente aqueles dos serviços de Cirurgia e Medicina do Trabalho do Centro Hospitalar do Baixo Vouga, E.P.E, Aveiro.

Finalmente, a todos aqueles que participaram e ajudaram diretamente ou indirectamente neste processo, um grande obrigado.

palavras-chave

biomarcadores de cancro colorectal, estadiamento de cancro, grupos clínicos, bactérias, fezes, microbioma, disbiose, 'ómicas', metagenómica, gene 16S rRNA, diagnóstico clínico

resumo

Do ponto de vista histológico, o cancro coloretal (CCR) reside na proliferação anormal de células epiteliais da mucosa do cólon, progredindo de adenoma a adenocarcinoma. Este cancro continua a ser o terceiro com maior incidência e mortalidade mundialmente. É causado por um acúmulo de mutações genéticas e silenciamento epigenético, para além de outros fatores de risco intrínsecos e extrínsecos. Devido às altas taxas de incidência e mortalidade, têm vindo a ser criadas, implementadas e otimizadas ferramentas de diagnóstico e prevenção. No entanto, continua premente a necessidade de desenvolver ferramentas que forneçam um diagnóstico cada vez mais precoce, rigoroso e sensível. Neste sentido, os objetivos desta dissertação consistiram em (1) desenvolver o estado da arte acerca das ferramentas de diagnóstico de CCR, (2) resumir as aplicações "ômicas" para indentificar biomarcadores microbianos relacionados com CCR e (3) comparar o microbioma de pacientes portugueses com CCR e indivíduos saudáveis. Atualmente, o diagnóstico de CCR tem vindo a ser conduzido por procedimentos mais (*e.g.,* colonoscopia) ou menos (*e.g.*, técnicas de imagem, biomarcadores moleculares) invasivos. Muito recentemente, a procura de biomarcadores microbianos através de ferramentas "ómicas" tem sido uma alternativa, principalmente devido à relevância do microbioma nas homeostase metabólica e fisiológica, assim como no funcionamento do sistema imunitário do hospedeiro. Assim, ao microbioma intestinal tem sido atribuído um papel ativo na evolução do CCR, podendo influenciar ou ser influenciado pela doença. Em particular, a análise metagenómica e metabolómica do microbioma associado a CCR em amostras de fezes tem estimulado a comunidade científica na procura de biomarcadores sensíveis, fidedignos, diferenciais, estáveis e precoces na deteção não invasiva da doença. Contudo, estes avanços carecem de uma representatividade para diversas áreas geográficas, dada o impacto cultural, genético e ambiental na incidência desta doença.

Neste sentido, realizou-se a análise do microbioma (com enfoque em Bacteria) em fezes de dois grupos clínicos constituídos por indivíduos Portugueses (pacientes com CCR e indivíduos saudáveis), através da sequenciação do gene 16S rRNA usando Ilumina MiSeq. Este estudo é um contributo para colmatar a lacuna de conhecimento existente sobre o microbioma associado a CCR na população Portuguesa. Apesar da estrutura do microbioma de fezes assumir padrões homogéneos entre indivíduos do mesmo grupo clínico, houve alguma variabilidade na abundância de *taxa* entre esses grupos e em diferentes estádios do CCR. Por exemplo, maiores abundâncias de *Prevotella*, *Alloprevotella*, *Sutterella*, *Desulfovibrio* e *Olsenella* observadas em amostras de CCR podem servir como biomarcadores microbianos. No futuro, o estudo será alargado a amostras populacionais maiores, assim como a outro tipo de amostras humanas e grupos clínicos, no sentido de identificar assinaturas microbianas sensíveis e específicas, que possam traduzir o desenvolvimento de CCR, reduzindo, assim, as taxas de incidência e mortalidade.

keywords

colorectal cancer biomarkers, cancer staging, clinical groups, bacteria, stool, microbiome, dysbiosis, biomarkers, “omics”, metagenomics,16S rRNA gene, clinical diagnosis

abstract

Histologically, colorectal cancer (CRC) resides in the abnormal proliferation of epithelial cells of the colon mucosa, progressing from adenoma to adenocarcinoma. This cancer continues to be the third with the highest incidence and mortality worldwide. It is caused by an accumulation of genetic mutations and epigenetic silencing, in addition to other intrinsic and extrinsic risk factors. Due to the high rates of incidence and mortality, diagnostic and prevention tools have been created, implemented, and optimized. However, the need to develop tools that provide an increasingly early, rigorous and sensitive diagnosis remains a pressing need. In this sense, the objectives of this dissertation were: (1) to develop the state of the art about CRC diagnostic tools, (2) to summarize "omic" applications in order to identify microbial biomarkers related to CRC and (3) to compare the microbiome of Portuguese patients with CRC and healthy individuals. Currently, the diagnosis of CRC has been driven by more (*e.g.,* colonoscopy) or less (*e.g.,* imaging techniques, molecular biomarkers) invasive procedures. Recently, the search for microbial biomarkers through "omic" tools has been an alternative, mainly due to the relevance of the microbiome in the metabolic and physiological homeostasis, as well as in the functioning of the host immune system. Thus, the intestinal microbiome has been assigned an active role in the evolution of CRC, being able to influence or be influenced by the disease. In particular, the metagenomic and metabolomic analysis of the CRC-associated microbiome in stool samples has stimulated the scientific community in the search for sensitive, reliable, differential, stable, and early biomarkers in the non-invasive detection of the disease. However, these advances lack representativeness for several geographic areas, given the cultural, genetic, and environmental impact on the incidence of this disease.

In this sense, the microbiome (with a focus on Bacteria) was analyzed in feces of two clinical groups constituted by Portuguese individuals (patients with CRC and healthy individuals), through the sequencing of the 16S rRNA gene using Ilumina MiSeq. This study is a contribution to fill the gap of existing knowledge about the microbiome associated with CRC in the Portuguese population. Although the structure of the fecal microbiome assumes homogeneous patterns among individuals of the same clinical group, there was some variability in the abundance of *taxa* between these groups and at different stages of CRC. For example, *Prevotella, Alloprevotella, Sutterella, Desulfovibrio* and *Olsenella* observed in CRC samples can serve as microbial biomarkers. In the future, the study will be extended to larger population samples, as well as to other types of human samples and clinical groups, in order to identify sensitive and specific microbial signatures that can translate the development of CRC, thus reducing incidence rates and mortality.

Table of Contents

CHAPTER III: STOOL MICROBIOME STRUCTURE IN HEALTHY *VERSUS* COLORECTAL CANCER PATIENTS TO TRACK POTENTIAL MICROBIAL BIOMARKERS OF DISEASE: A PORTUGUESE COHORT STUDY

List of Figures

List of Tables

Abbreviations

AhpC	Alkyl Hydroperoxide Reductase C
Akt	Protein Kinase B
ALX4	Alx Homeobox 4
APC	Adenomatous Polyposis Coli
AvrA	SPI-1 Effector Protein (*S. enterica*)
BMP3	Bone Morphogenetic Protein 3
BRAF	B-RAf Proto-Oncogene Serine/Threonine Kinase
CA19-9	Carbohydrate Antigen 19-9
CA-125	Carbohydrate Antigen 125
CAP	Constrained Analysis of Principal Components
CD 11c/ ITGAX	Integrin Alpha X
CD 33	Siglec-3
CD 206	Cluster of Differentiation 206, Mannose Receptor
CD209	Cluster of Differentiation 209, DC-SIGN
CEA	Carcinoembryonic Antigen
CHBV	Centro Hospitalar Baixo Vouga
CIMP	CpG Island Methylator Phenotype
COGs	Clusters of Orthologous Groups
CRC	Colorectal Cancer
CT	Computed Tomography
CXCL10	C-X-C Motif Chemokine 10
CYFRA21-1	Cytokeratin- 19 Antigen (fragment)
CYR61	Cysteine Rich Angiogenic Inducer 61
DCBE	Double Contrast Barium Enema
2D-DIGE	Two Dimensional Difference Gel Electrophoresis
DGGE	Denaturing Gradient Gel Electrophoresis
DKK3	Dickkopf Related Protein 3
EPCAM	Epithelial Cellular Adhesion Molecule
ERK	Extracellular Regulated Kinase
FadA	Adhesion protein FadA (*F. nucleatum*)
FDA	US Food and Drug Administration
FDG	Fludeoxyglucose
FIT	Fecal Immunochemical Test
FliC	Flagellar Filament Structural Protein
FusA	Elongation Factor G
GapA	Glyceraldehyde-3-Phosphate Dehydrogenase A
GBP-1	Guanylate Binding Protein 1
GC-MS	Gas Chromatography Mass Spectrometry
gFOBT	Guaiac-Based Fecal Occult Blood Testing
GREM1	Gremlin 1
HITChip	Human Intestinal Tract Chip
HMP	Human Microbiome Project
^{1}HNMR	Proton Nuclear Magnetic Resonance Spectrometry
IBD	Inflammatory Bowel Disease
IBS	Irritable Bowel Syndrome
iFOBT	Immunological Fecal Occult Blood Testing
IGFBP2	Insulin-like Growth Factor Binding Protein 2
iHMP	Integrative Human Microbiome Project
IL-6,-8	Interleukin 6, 8
KEEG	Kyoto Encyclopedia of Genes and Genomes
KRAS	Kirsten ras Oncogene Homolog

LG/MS	Liquid Chromatography/ Mass Spectrometry
MAPK	Mitogen-Activated Protein Kinase
MBD1	Methyl-CpG Binding Domain Protein 1
MLH1, 2, 6	Mut L Homologs 1, 2, 6
MPIDB	Microbial Protein Interaction Database
MRI	Magnetic Resonance Imaging
MSH2	Mut-S Protein Homolog 2
MSI	Microsatellite Instability
NDRG4	NDRG4 Family Member 4
NGS	Next Generation Sequencing
NPTX2	Neuronal Pentraxin 2
NSDT	Nematode Scent Detection Test
PET	Positron Emission Tomography
PfK	Phosphofructokinase
PI3K	Phosphatidylinositol-3 Kinase
PIK3CD	Phosphatidylinositol-4, 5-Bisphosphate 3-Kinase Catalytic Subunit α
PKM2	Pyruvate Kinase M2
PKS	Polyketide Synthase
PMS1	Postmeiotic Segregation Increased 1
PURPL	p53 Upregulated Regulator of p53 Levels
RANTES	Regulated on Activation, Normal T-Cell Expressed, and Secreted
RARB	Retinoic Acid Receptor Beta
RET	Rearranged During Transfection
RT-qPCR	Reverse Transcription Quantitative Polymerase Chain Reaction
SDC2	Syndecan 2
SCFA	Short Chain Fatty Acid
SEER	Surveillance, Epidemiology, and End Results
SMAD4	SMAD Family Member 4; Mothers against Decapentaplegic Homolog 4
SNORA42	Small Nucleolar RNA 42
TCF21	Transcription Factor 21
TGF-β	Transforming Growth Factor β
Tig	Trigger Factor
TOF-MS	Time-of-Flight Mass Spectrometry
TrEMBL	Translation of the EMBL Nucleotide Sequence
TNF	Tumor Necrosis Factor
TNM	Clinical Tumor-Node-Metastasis
TP53	Tumor Protein p53
T-RFLP	Terminal Restriction Fragment Length Polymorphism
USG	Ultrasonography
VIM	Vimentin
VOCs	Volatile Organic Compounds
Wnt	Wingless-type MMTV Integration Site Family Member

CHAPTER I

General Introduction

CHAPTER I: GENERAL INTRODUCTION

1. An Overview of Colorectal Cancer

1.1 Histopathological Development

Colorectal cancer (CRC), also known as bowel cancer, is characterized by an abnormal or dysregulated cell proliferation originating in the colon and/or rectum tissues, which normally assist in the excretion of fecal matter, as well as final nutrient and water uptake[1]. The most common predecessor stages that may lead to CRC are the formation of non-malignant inflammations/adenomas (polyps) developed from gland cells on the lining or in the mucosa layer of the colon/rectum[2,3], which may evolve into pre-cancerous adenomas (*e.g.*, adenomatous polyps) and ultimately to adenocarcinomas (proliferative, metastatic, and invasive tumor) (Figure 1). These adenomas, in contrast to hyperplastic and inflammatory polyps, can and often do turn malignant, which are then classified as invasive adenocarcinomas. The presence of these specific, threatening neoplasms can be observed by the growth progression from the lining into the walls of the colon or rectum and the worst case scenario migration, or metastasis, via the lymph or blood to other tissues and organs of the body[3] (Figure 1). These adenocarcinomas arising in the mucosa epithelium have been found to be accounted for almost all of the diagnosed CRCs,[4–6] and are graded as well, moderately, or poorly differentiated, or undifferentiated with respect to the manifestation of glandular formations[7]. The severity at the time of diagnosis is determined by oncologists and pathologists using two frequently used systems of classification: (1) the clinical tumor–node–metastasis (TNM) staging in which the lettering indicates the original site of the primary tumor, the lymph node proliferation, or spreading to distant organs, being the increased numbering a reflection of unfavorable prognosis[8]; (2) the surveillance, epidemiology and end results staging (SEER) system that is characterized by the following categories: *in situ*, localized, regional, distant, and unknown[9]. Other elements considered in tumor classification may include the extent to which the malignancy involves regions in its periphery, percentage of bowel obstruction, and invasiveness pattern[7].

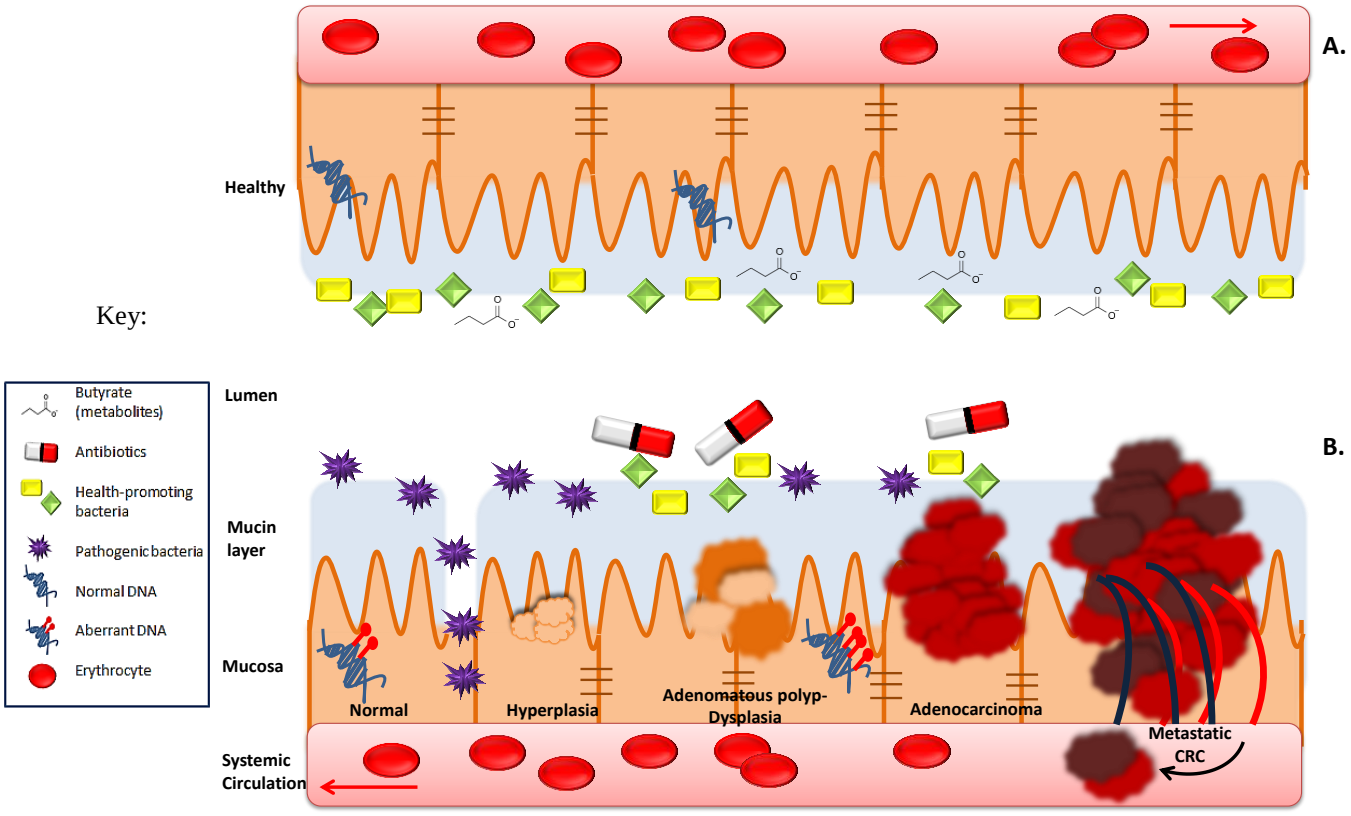

Figure 1: Colorectal cancer development and progression (A.) Normal intestinal epithelium maintenance by health-promoting bacteria and the production of their beneficial metabolites (*e.g.,* butyrate) that guarantee mucin barrier integrity conservation and anti-inflammatory characteristics. **(B.)** Colorectal cancer development caused by epi/genomic aberrations (*e.g.,* microsatellite and chromosomal instability), mucin-degrading bacteria, inflammation, environmental agents (*e.g.*, antibiotics) that reduce intestinal homeostasis and promote cell aberrant proliferation.

1.2 Molecular, Genetic, Epigenetic Characterization

With consistent advancements in biotechnological, molecular biology, and cytological tools, not only is CRC examined at a histopathological and clinical level, but also on its molecular, genetic, and epigenetic microscopic underpinnings (Figure 1). Cancer pathogenesis is commonly due to the accumulation of tumor suppressor gene mutations that inactivate them, such as adenomatous polyposis coli (*APC*)[10,11], guanylate binding protein 1 (*GBP-1*)[12], rearranged during transfection (*RET*) proto-oncogene[13], transcription factor 21 (*TCF21*)[14], and methyl-CpG binding domain protein 1 (*MBD1*)[15]; or oncogenic mutations that stimulate an aberrant behavior, such as Kirsten ras oncogene homolog (*KRAS*), phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit α (*PIK3CA*), and B-RAf proto-oncogene serine/threonine kinase (*BRAF*)[11,16,17]. These genomic alterations affect usual inter/intracellular communication, namely multiple signaling

cascades such as Wnt, MAPK/PI3K, TGF-β, and TP53 pathways already reviewed by a few authors[10,18–20]. The presence of APC or β-Catenin mutations and crosstalk with the MAPK/PI3K pathway results in an overactive or dysregulated Wnt pathway, causing transcriptional activation of CRC-associated genes, hence enhancing proliferative properties instead of maintaining its normal function in regulating a balanced intestinal stem cell growth and the differentiation of various intestinal cell types[21]. Likewise, the mutational silencing of the receptors or intracellular mediators of the TGF-β pathway initiates proliferative cancer progression nulling its role in tumor suppression[22]. Furthermore, through genome-wide analysis and next generation sequencing technology (NGS), genetic/epigenetic instability, such as microsatellite instability (MSI)[17,20,23–25], chromosomal instability[20,25], and CpG island methylator phenotype (CIMP)/genome-wide hypermethylation[17,20,25,26], have been highlighted and interconnected to various cancers, primarily in CRC. These instabilities are characterized by hypermutability, epigenetic silencing of DNA repair system-encoding genes, anomalies in chromosome number, and the presence of inactivating methyl groups near or on the promoter regions of tumor suppressor genes[17,20,23–26].

1.3 Epidemiological Importance

Out of the world's estimated 18.1 million new cancer cases in 2018, 1.8 million cases are attributed to CRC in both sexes, making it the third most common in males after lung and prostate cancers, and second most common in females after breast cancer[27]. Taken together, CRC is the third most frequent cancer detected and diagnosed, and the fourth most common cause of mortality that is cancer-affiliated on a global level with an alarming prospective increase to approximately 2.2 million new cases and 1.1 million deaths by 2030[28,29]. The incidence and mortality numbers are hugely high, and continue to rise, in more developed continents, such as Australia, Europe, and North America, than less developed areas, mainly because of notably significant differences in economy, culture, and lifestyle. In Portugal alone, which has conformed to such lifestyle changes[30,31], the raw incidence rate of malignant colorectal tumors was 70.3 *per* 100,000 individuals in 2010[32]. Moreover, colorectal mortality rates for both genders combined

reached 36.1 *per* 100,000 individuals in 2014, with the rates of survival shown to decrease significantly with time after diagnosis[32,33].

1.3.1 Etiological Risk Factors

There are numerous risk factors that have been associated with increased predisposition in acquiring CRC as aforementioned briefly: age, gender, ethnical and genetic background or predisposition, medical history, behavior and lifestyle choices, and exposure to environmental carcinogens[34] (Figure 2). Age is a risk factor associated with multiple cancers and other diseases, primarily because of the consequences of cellular aging. The macroscopic signs of aging observed in the older generation reflect the breakdown of normal cellular functioning, organelle damage, and less efficient quality control and repair mechanisms, both at transcriptional and translational levels[35]. Just as CRC occurs more frequently in individuals older than the age of 50 with the exception of the early onset on those with hereditary susceptibilities (mentioned below)[6,34,36], ethnical or racial group variances with genetic proneness among African Americans and, surprisingly, Japanese Americans, in comparison to Caucasians/whites, display the highest and an increased incidence of CRC[3,37].

Lynch syndrome (*MLH1, MLH2, MLH6, PMS2, EPCAM),* hereditary mixed polyposis (*GREM1*), juvenile polyposis (*SMAD4)*, and familial adenomatous polyposis (*APC*) are some of several autosomal dominant hereditary conditions prompting the development of colorectal tumorigenesis, by demonstrating microsatellite instability, large rearrangements, CpG island methylation, and base duplication[38,39]. Individuals who suffer from inflammatory bowel disease (IBD) (*e.g.,* Crohn's disease and ulcerative colitis), type 2 diabetes mellitus, obesity, or have multiple familial CRC occurrences, also have a greater tendency for colorectal malignancies[3,34].

In contrast to fish, the consumption of red and processed meats tend to have a pro-CRC effect because of toxic heme and high lipid content inducing tumorigenesis, although these diet-cancer relationships show gender-specific inconsistencies[3,40,41]. In contrast, vegetables and high fiber content diets have been shown to have protective and anti-CRC effects, not only because of the presence of flavonoids (antioxidants), but also because they promote stool bulk for defecation, and the microbial production of colonic

health-promoting metabolites [*e.g.,* short-chain fatty acids (SCFAs), butyrate][42]. In addition, regular alcohol consumption, as well as exposure to cigarette smoke inhalation with dose-dependent intakes have a directly proportional link with the development of advanced adenomas and multiple adenomas (in alcohol consumers), comparatively to nonalcoholic drinkers and nonsmokers[3,34,43–46].

Recent attention has been drawn to the notion of microorganisms playing a role as a potential influencing factor in carcinogenesis and IBD via environmental contaminants or gut modifications due to daily dietary habits, stress, and overuse of antibiotics[47–49] (Figures 1 and 2). Nevertheless, CRC is not defined by a single causal factor, but by a combination of additive and co-acting risk factors, including variable biologic/genetic features and/or environmental exposures that augment the chances of disease manifestation.

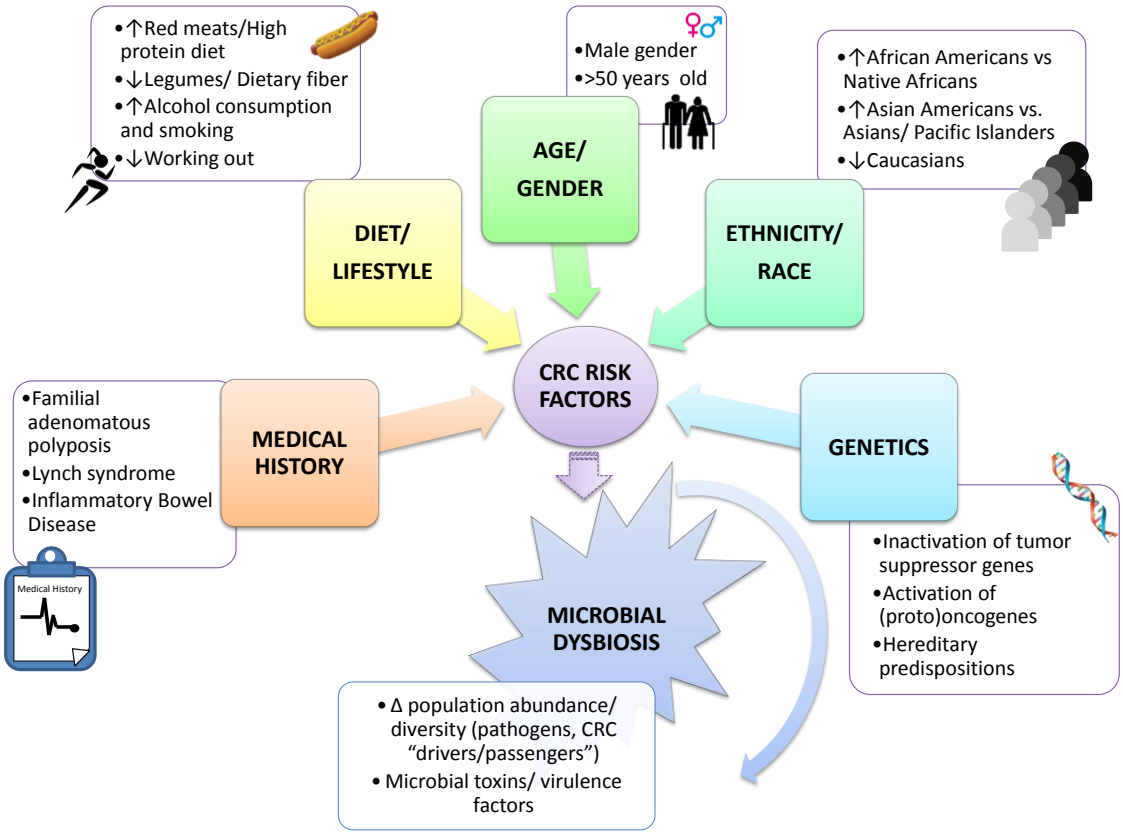

Figure 2: Summary of contributing risks for CRC onset

2. Colorectal Microbiome and Cancer

2.1 Microbial Symbiosis and the (Healthy) Microbiome

The microbiota, microflora, or gut flora are terms used to describe the massive number of diverse microorganisms (*e.g.,* bacteria, yeast, viruses) that have co-evolved and symbiotically inhabited different human body niches (*e.g.,* skin, genital tracts, oral cavity), particularly in the gut[50–52]. The microorganisms colonizing the intestinal mucosa, specifically on the loose outer mucus gel layer composed of mucins, provide humans with energy (*e.g.,* due to caloric storage) and nutrients for normal cellular functioning, anabolism of vitamins (*e.g.,* K and B9), fermentation of indigestible material (*e.g.,* fibers), synthesis of fatty acids (*e.g.,* butyrate), a protective barrier against potential pathogen penetration and invasion, and enhancement of the host's immune defenses (second immunological system)[53–56]. In return, the mutualistic microorganisms are provided with a sheltered environment rich in macromolecules for their metabolism and growth provided from the host's food intake[57]. Hence, host and resident microbiota are essential players for both species' coexistence and any disturbance in the delicate equilibrium with respect this bi-dynamics relationship leads to detrimental outcomes *(cf.* Section 2.2).

The environments in which the microorganisms adapted to reside and thrive in are reflections of their morphological and physiological characteristics[18,58]. In the colorectal region of the gastrointestinal tract, there is an interesting gradient of microorganisms diversity and abundance associated with the organ properties, such as variations in pH, moisture, oxygen concentrations, digestive passageway rates, nutrient/substrate levels, and condition of the mucosa[18,57–62]. The majority of the microorganisms colonizing the gut are obligate anaerobes due to an absence of oxygen, but there are cases of facultative/tolerant anaerobes (*e.g*., Enterobacteriaceae), and possible aerobes in the rectum, proximal to the anal orifice[57,61]. Consequentially, the lower gastrointestinal tract is mostly dominated by two specific phyla of bacteria and one archeon species: Bacteroidetes (mostly Gram negative bacteria) and Firmicutes (mostly Gram positive bacteria), and the *Methanobrevibacter smithii* Archaea[52,61,63]. Of the thousands of *taxa* that populate the large intestine, the most common bacterial genera under healthy

conditions include *Bacteroides*, *Clostridium*, *Ruminococcus*, *Eubacterium,* and *Bifidobacterium*[64]*.*

In addition to the microbiota, the human microbiome, not only refers to the diversity and characterization of microbes, but also to the collection of genetic information of the microbes that live symbiotically in various environmental niches[52,62,65]. Just as the Human Genome Project aimed to accomplish and decipher the entire DNA sequence in hopes of discovering the mysteries of complex genetic disease and jumpstart the era of personalized medicine, the Human Microbiome Project (HMP) intended to gain insight on the microbial diversity that exists among healthy individuals in areas such as the skin, gut, and vagina, and understand the underlying mechanisms by which microbes produce beneficial and unfavorable outcomes to their host, using "multi-omic" large scale analysis, datasets, and genomic profiling technology (*cf.* Chapter II)[52,62,66].

Out of all of the human body areas, the microbiota or microbiome of the gut has by far gained most attention recently, with an ever increasing 10,000s of publications on the Pubmed search engine dated from 2011 to 2016[65]. For example, Goodrich and her colleagues[67] aimed to understand the interplay between the modifiable gut microbiome and the hereditable host's genomic variations, using both identical and fraternal twins with a common initial environment. In this study, the dominant phyla observed were Firmicutes, Bacteroidetes, and Proteobacteria and, in contrast to fraternal twin pairs, identical twin pairs with indistinguishable genomes had significantly more related microbiotas with respect to taxonomic diversity[67]. With respect to hereditability, the family *Christensenellaceae* (order Clostridiales, phylum Firmicutes), in contrast to phylum Bacteroidetes (responded to environmental dietary modifications), was found to be the most heritable taxonomic group and was associated with leaner, healthier individuals[67].

2.2 Microbial Dysbiosis and CRC

In healthy conditions, the resident microbiota present in the human gut serve the purposes aforementioned; however, in disease scenarios, the microbial symbiosis transforms itself into dysbiosis, a term used to illustrate the complex ecosystem unbalance via shifts in microbial population abundances, diversity (*e.g.*, permanent to

transient/foreign pathogens), functional structure, and biosynthetic profiles namely relying on the expression of microbial toxins and virulence factors that eventually lead to disease states (*e.g.*, IBD and cancer)[52,68] (Figure 2). In a 1, 2-dimethylhydrazine-induced tumor-forming mouse model experiment conducted by Sun *et al*.[11], alterations in the microbial gut composition were observed. Even though in all samples the principal phyla (of eight observed) were Bacteriodetes and Firmicutes, the abundance of Bacteriodetes was higher in experimental *versus* control groups in comparison to Firmicutes. Moreover, out of the remaining 6 phyla identified, Deferribacteres was absent in experimental groups[11]. In addition, Clostridiaceae, Peptostreptococcaceae, and Sutterellaceae families were only present in the tumor-induced experimental groups, thereby demonstrating a shift in microbial population and abundances on different taxonomic levels[11]. In concordance with Sun *et al*.[11], Gao *et al*.[68] proved that the occurrence of Firmicutes, Bacteroides, and Fusobacteria (genera *Lactococcus*, *Fusobacterium*, *Escherichia-Shigella*, and *Peptostreptococcusten*) was frequent amongst colorectal patients, while Proteobacteria was the most detected in the control group. However, some discrepancies exist between results such as in a study developed by Kostic *et al*.[69], where a depletion of the phyla Firmicutes and Bacteroidetes and an enrichment of Fusobacteria (*e.g.,* mainly *Fusobacterium nucleatum*[70] compared to less common *F. necrophorum*, *F. mortiferum* and *F. perfoetens* species), was observed in CRC patients (tissue and resected metastases). Additionally, fragilysin-producing *Bacteroides fragilis* (enterotoxic *B. fragilis*), an obligate anaerobe and gut colonizer of healthy individuals (becomes an opportunistic pathogen) proven by a statistically significant isolation between CRC patients and healthy individuals, has been associated not only with diarrhea and IBD, but also as a driver in CRC[71,72]. For example, a high detection rate of the *bft* gene encoding fragilysin, the enterotoxin stimulating oncogenic transcription and thus carcinogenesis promotion through Tcf-dependent β-catenin nuclear mechanism pathway activation within the colonic epithelium, was observed among colorectal individuals in comparison to the controls[71].

In contrast to Herpes Simplex virus, Cytomegalovirus and Epstein-Barr virus where there was no evidence demonstrating an oncogenic involvement in CRC

development[73], Human Papillomavirus (HPV), a sexually transmitted viral disease and anal/cervical cancer establishing agent, has demonstrated conflicting results with respect to CRC correlation[74–76]. Despite only few studies reporting the association of HPV DNA to CRC[74–76], a higher percentage of HPV-16 DNA in CRC tissue samples was identified and compared to the control group with a larger incidence in the rectum area[74]. Overall, mainly bacteria have been associated to CRC development, such as *Fusobacterium* spp. (*e.g.*, *F. nucleatum*, *F. necrophorum*, *F. mortierum*, *F. perfoetens*), *Streptococcus* spp. (*bovis[infantarius]/ gallolyticus*), *Clostridium septicum*, *Slackia*, *Enterococcus faecalis*, and *Escherchia coli*[72,77]. No single microorganism or risk factor has been attributed as the causal factor for colorectal pathogenesis; however, a combinatorial microbial network relationship with inter-individual differences may be the ultimate key to CRC development[78].

CRC location can influence the structure and community shifts of colon microbiome. Tumors more frequently occur in the left or descendent colon segment, but also in the rectum, though with variable degrees of adenocarcinomas development[6,79]. Flemer and colleagues[78] demonstrated an insignificant difference with respect to the microbiome community between malignant and non-malignant matched tissues within the same patient. However, significant differences were detected between the distal (descending colon and rectum; *e.g.*, *Alistipes*, *Akkermansia*, *Halomonas*, and *Shewanella*) and proximal (ascending colon) microbiota, which could be associated with the increased percentages of left-sided intestinal tumors[78]. In another study, the microbiome of distal tumors presented higher diversity and also abundance of specific *taxa* (*Fusobacterium*, *Escherichia, Shigella*, and *Leptotrichia*) compared to proximal tumors[68]. Contrasting to the left-sided predominance of CRC occurrence, biofilms, tightly aggregated bacterial communities known for their adhesive and inflammatory properties, have been shown to be associated with right-sided colorectal lesions and the risk is five times greater for CRC development for intestinal tissues sustaining biofilms[80,81].

3. CRC Diagnostics

The appearance of cancer is not a novel outbreak plaguing humanity. It is rather an ancient disease with fossilized recordings as early as the mid-Mesozoic era from dinosaur remains that indicated the presence of osteoma metastatic tumors[82], but also there was an evidence of osteoblastic lesions possibly arising from metastatic prostate cancer, as highlighted in radiographies of Egyptian mummy's remains (c. 285–30 B.C.)[83]. These historical discoveries, however, could only be achieved by modern groundbreaking technologies, which in turn can help with the development of other novel detection methods for implementing preventative and curative measures. In attempt to reduce incidence and poor prognosis, as well as to increase overall survival rates of CRC patients a great deal of research has been devoted to new detection methods. For that, high throughput technology, advanced imaging systems, and contemporary surgical devices have been employed throughout time to discover reliable monitoring targets, essentially based on the detection of histopathological, molecular and/or microbial disease markers (Figure 3).

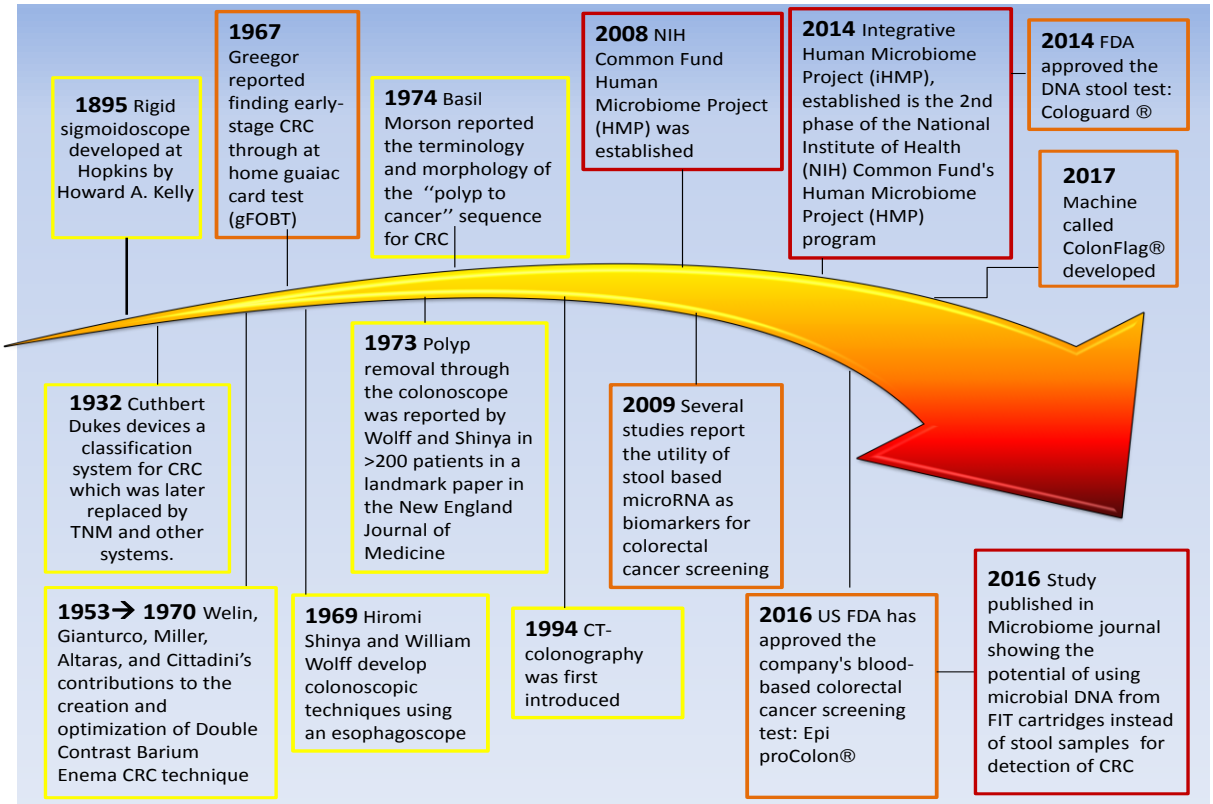

Figure 3: Historical CRC diagnostic timeline summary. [66,84–90] Yellow, orange, and red boxes indicate the histopathological-, molecular-, and microbial-based diagnostic findings and contributions over time, respectively.

3.1 Histopathological Examination

3.1.1 Invasive/ Direct Visualization Procedures

The currently most ideal procedure for preventative CRC diagnostics is the colonoscopy, also referred to as a lower endoscopy, first developed and performed by two medical professionals, William Wolff and Hiromi Shinya, in the Beth Israel Medical Center, New York City, with published successful results in the early 1970s[91–94]. Those individuals evidencing clinical manifestations suspected of CRC or asymptomatic individuals possessing factors with increased risk (*e.g.,* age >50, past familial history) undergo a process directed at histological colon analysis by the insertion of a fiber optic colonoscope through the anus along the full length of the colon in order to track abnormalities observed in the mucosal tissue and remove potential precancerous polyps through a polypectomy[91,93,95]. Early polypectomies during Wolff's and Shinya's time were established using a primitive wire loop snare that attached to a tube with an eyepiece and lens on opposite ends[91,94]. Since then, the colonoscope has been improved to bypass the anatomical maneuvering difficulties encountered by the gastroenterologist, as well as attempting to eliminate discomfort, procedural bleeding, intestinal wall tearing by overstretching the tissue, and obtain better imaging quality detection through high definition monitoring and narrow band technologies, water immersion, air-insufflation, and novel panoramic cap devices for full spectrum and multi-view analyses[96]. Approximately 25% of precancerous polyps, especially flattened lesions, are overlooked during a colonoscopy, due to the folding nature of the large bowel, and so increasing the observation area is imperative for colorectal risk reduction[97]. Hence, Rubin *et al*.[97] evaluated the addition of the Third Eye® Panoramic Cap extension in comparison to the standard instrument[98]. The authors' findings pointed out an improved detection with a 300º wide-viewing range and an acceptable overall polyp detection rate of 44%[97]. Other device extensions such as the Endocuff and EndoRings, have demonstrated improvements for accurate colonoscopy performance[99].

The colonoscopy procedure as a whole, along with its advancements throughout time, has been proven to be the ideal screening method with studies demonstrating a ~60% colorectal risk decrease[100], a variable polyp detection rate with a median of

~43%[101], and polyp detection miss rates ranging from 2.1-26% with lower oversight rates in larger polyps (>10mm) *versus* smaller ones (1-5mm)[102]. Additionally, the implementation of an optimized protocol has allowed for 83% and 89% CRC incidence decrease and mortality decline, respectively[103]. However, there are some inevitable limitations and restrictions in undergoing a colonoscopy that can hamper its use frequency and success. First, the cost and invasiveness of the technique requires highly-trained personnel with expertise on polyp removal (some may only be partially removed what may enhance the development of interval cancers) and colonoscope maneuvering, as to prevent intestinal perforation and post-procedural bleeding. Moreover, there is the need for advanced equipment, optimal pre-procedural bowel cleansing of the patient to achieve precise detection results, and sedative/anesthetic administration whenever needed[96,104–108].

Although oftentimes confused as the equivalent to a colonoscopy, a sigmoidoscopy, both flexible and rigid, differs with respect to examining only the lower portion of the large intestine, specifically the rectum, sigmoid, and descending colon, which are the colon areas most vulnerable to and frequently affected by CRC[6,92,94,109,110]. Although successful in finding aberrations in the distal region of the colon with high detection rates and low perforation rates[111,112], a thorough colonoscopy examination must be done to confirm the diagnosis and to identify any other abnormalities potentially present in the proximal regions of the large bowel.

3.1.2 Imaging Techniques

Complementary to the standard endoscopic techniques used by gastroenterologists for direct intestinal irregularity visualization, imaging techniques used by radiologists, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), ultrasonography (USG), and double-contrast barium enema (DCBE), have evolved to diagnose and detect diseases such as CRC and give indications for further invasive and noninvasive testing[91]. Colon directed CT scanning, also known as a virtual colonoscopy, allows for the visualization of two-dimensional and even three dimensional images via multiple cross-sections of the large intestine interior lining

through the utilization of low radiation dosage and by the insertion of a CO_2 insufflator at the rectum entry[89,110]. Although the conventional colonoscopy is the gold standard of colorectal diagnostics, virtual colonoscopy is the preferred method both in procedure and pre-procedural preparation by 73.9% of average-risk patients, after undergoing both methods (CT and colonoscopy) in a retrospective questionnaire-based study[108]. This CT preference is despite the risks of radiation exposure, possible sensitivity to the contrast media, fatigue, and overall discomfort the patients may experience, the compliance result being fundamental to the success of colorectal screening[108]. However, virtual colonoscopies, although less sensitive and accurate than invasive methods, but effective in proximal adenocarcinoma detection, are reserved for patients who are unable to undergo a conventional colonoscopy due to daily administration of anti-coagulants (*e.g.,* in elderly patients) and risk of perforation in invasive colonoscopies, or for those with an obstructed colon by tumor size, hence preventing the passage of a colonoscope[89,110,113]. In addition, virtual colonoscopies are optimal for strategic surgical planning[110], TNM staging, and precise localization of tumors[113,114]. In a study conducted by Pullens *et al*.[113], symptomatic and asymptomatic patients succumbed to a virtual colonoscopy after an incomplete colonoscopy, in which an additional 27 pre-malignant and malignant polyps, including one flattened (sessile) early adenocarcinoma were detected by CT and overlooked during the conventional method. Thereby, it suggests to be an advantageous, preliminary alternative technique for patients with restrictions to other diagnostic methods abovementioned or in combination to such standard methods (*e.g.,* colonoscopy).

In some ways, MRI, a detailed high-contrast and resolution scanning system used to obtain clear cross-sectional images of internal organs [115], serves similar purposes in CRC pre-surgical planning and staging of malignancies for most appropriate therapy administration, as does CT scanning[116]. However, MRI is reserved primarily for distal-rectal cancers and, as opposed to CT that is optimal for colonic abnormality observations, MRI is more sensitive in distinguishing the relative progression of tumors, the differentiation from normal mucosa, and degree of extension or even proximal organ metastasis[116]. Although limited on patient size, Nerad *et al.*[117] demonstrated promising

results showing the alternative usage of MRI towards CRC diagnostics with high accuracy in tumor localization, high specificity and sensitivity of T3-T4 staging of rectal tumors[118], but also colonic tumors at the mucosa, and extension of lymphatic and serosal involvement. Nevertheless, a drawback to this study was the low detection percentages of metastases occurring beyond the intestinal wall and in nearby organs[117].

Another imaging method with an ability to acquire CRC's physiological features is PET, where positrons emitted by a radiotracer, ^{18}F-fludeoxyglucose (FDG), interact with electrons releasing photons in opposite directions that are detected by a scanner to measure metabolic activity and create a three dimensional image of the internal organ[119]. Kunawudhi *et al*.[120] directed a comparative diagnostic analysis using both noninvasive preoperative imaging and invasive colonoscopy in order to determine the accuracy and precision in detecting pre-carcinomas/carcinomas with the presence of the following positive lesions: three adenoma types (villous, tubular and tubulovillous), serrated hyperplastic polyp/hyperplastic polyposis, and colorectal carcinoma. Although there were significant positive predictive values for colonic neoplasias, the cancer detection rate and sensitivity were low in this study with almost all being false negatives, presenting a size from <5mm to <10mm. This uncertainty indicates that multiple colorectal diagnostic techniques should be applied to prevent failures on the detection of adenoma to adenocarcinoma progression and transition[120]. In this context, to compare the efficiency of colorectal staging, two imaging methods, MRI and ^{18}F-FDG-PET/CT, were used to scan the entire bodies of patients at stage 3 and 4 of CRC, as to evaluate nodal involvement and metastases occurrence[121]. The outcomes indicated more efficient detection of nodal involvement using PET-CT scanning, increased number of liver metastases detected using MRI, but increased number of lung metastases using PET-CT in agreement with Yu *et al.*[122], and equal detection rates of bone metastases using both scanning methods[121]. Therefore, one cannot discount the effectiveness of either CRC diagnostic imaging technique, as there are advantages and disadvantages of usage to both[121].

USG of the abdomen is another method to detect CRC, although not one of the most popular, but bypasses the popular limitations of other diagnostic tools, such as duration, difficulty, and health risk levels to the patient[123]. Nevertheless, although

ultrasounds for colorectal imaging fall short in detecting smaller lesions and those of the rectal ampulla, abdominal ultrasounds have shown to have high sensitivity, specificity, positive and negative predictability of neoplastic detection with the majority of carcinomas identified in leftmost region of the gastrointestinal tract as demonstrated by Martínez-Ares *et al.*[123]. For those CRC patients exhibiting liver metastases, the efficiencies, sensitivities, and specificities of contrast enhanced USG and computed tomography were compared by Rafaelsen *et al.*[124]. Identical sensitivities and negative predictabilities, as well as similar specificities were observed for both methods with the exception of a higher positive predictability percentage observed in the USG approach, indicating a conceivable alternative strategy for metastatic detection[124]. Despite the positive results obtained for CRC detection, staging rectal cancer using endorectal USG has been controversial[125]. Asraf *et al.*[125] have demonstrated unsatisfactory endorectcal staging results with high percentages of inaccuracy due to the inability to precisely measure invasion depths with USG, which is a crucial factor for proper surgical and therapeutic administration. However, Halligan[126] challenges these results[125] by mentioning the demanding technical and interpretative skill requirements during examination and that UGS, in fact, has utility in distinguishing early *versus* late staged cancer.

DCBE, although a method often replaced by CT virtual colonoscopy currently attributed to less discomfort and more patient tolerability of the overall procedure[127], it is a radiographic technique using two contrast types as the name itself suggests, enhancing visualization of the X-rays taken of the organ. DCBE has, in older literature, demonstrated high sensitivity, low percentage of false negatives, and low oversight rates with higher probability of failure in small lesion detection (<3cm)[128,129]. However, more recent investigations have been skeptical upon the usage of this technique in detecting CRC[130]. This is due to contradictory results concerning significant undetected malignant lesion rates with higher prevalence of occurrence in rightmost portion of the colon, as well as the rectum[130].

3.2 Molecular- and Cellular-Based Detection

Biomolecular testing to identify specific genes, proteins, and other factors unique to CRC has been trending in biomedical investigation. The identification of genetic CRC biomarkers has led to the development of a stool (host) DNA test named Cologuard® that was approved in 2014 by US Food and Drug Administration (FDA). This test is convenient for patient acceptance and serves as an effective strategic preliminary test[131]. The Cologuard® test was developed based upon the laboratorial research conducted by Imperalie *et al*.[132] in which specific genetic and epigenetic aberrations associated with CRC are detected. One of them is the *KRAS* mutation, which has been frequently present in tissue, blood, and fecal DNA samples[133], but NDRG family member 4 (*NDRG4*) and bone morphogenetic protein 3 (*BMP3*) promotor region methylation, and *β-actin* reference gene, can also be quantified. Results of significantly higher sensitivity, specificity, and detection efficiency of premalignant polyps and carcinomas are observed with Cologuard® when compared to the fecal immunochemical test (FIT), a immunoassay test specific for human hemoglobin commercially available for clinical practice[132]. Within a year of access to the Cologuard® stool test, another FDA approved noninvasive molecular test for CRC detection called Epi proColon® was released, in which a simple blood sample would suffice in effectively determining the presence of methylated Septin9 CRC marker from circulating tumor DNA[134]. This marker is involved in numerous cell survival processes and apoptosis[134]. In addition to the inhibition of the Septin9 gene via hypermethylation of promoter region, six other hypermethylated genes, namely ALX homeobox 4 (*ALX4*), *BMP3* (a genetic marker previously mentioned[132]), neuronal pentraxin 2 (*NPTX2*), retinoic acid receptor beta (*RARB*), syndecan 2 (*SDC2*), and vimentin (*VIM*) have been highlighted with variable sensitivities and specificities, as to broaden the spectrum of potential molecular biomarkers for CRC detection[135,136]. An automatic diagnostic system sequentially performing genomic DNA extraction, purification, amplification, mutation detection, and interpretation, has been proposed as a possible effective equipment towards CRC detection. This is performed via the recognition of *KRAS*, *BRAF*, *PI3KCA* mutational biomarkers (often synergistically present) and can be applied to frozen, formalin-fixed or paraffin-embedded tissues[137]. It has been found to have better

precision, sensitivity, accuracy, while requiring a lower effort and time than traditional Sanger sequencing[137]. Similarly, a recent machine called ColonFlag® calculates and incorporates patients' clinical information such as gender, age, and blood count to efficiently differentiate CRC patients *versus* healthy individuals. Tumors can be early-flagged by approximately six months to one year prior to the time it would be detected through histopathological analysis, besides doing that with higher accuracy, which is a limitation of most CRC diagnostic methods[138].

In addition to DNA-based biomarkers, attention has likewise been drawn to noncoding RNAs (ncRNAs), such as circular RNAs, long noncoding RNAs, and microRNAs (miRNAs). Hundreds of dysregulated levels of circular RNAs, which are involved in normal genetic regulation and expression, have been distinctively identified in colorectal tumorigenesis and poor prognosis conditions with has-circRNA-103809 and has-circRNA-104700 downregulation being of the most interest, yet with unknown mechanisms[139]. Long noncoding RNAs have regulatory roles and, just as other RNA classes, are controlled by transcription factors like p53[140]. These specific ncRNAs have been selectively identified in CRC tissue, one of which was designated by Li *et al*.[140] as p53 upregulated regulator of p53 levels (*PURPL*). *PURPL* serves as a p53 suppressor and its expression is self-regulated by the same transcription factor complex[140]. By maintaining low p53 levels, this specific RNA has been shown to jumpstart tumorigenesis and cancer progression with mechanisms yet to be fully determined[140]. miRNAs are the control system of gene expression at a post-transcriptional level and when up- or downregulated, it may result in mishaps in cell differentiation, propagation, and overall survival[141]. In colorectal tumors, dysregulated levels of several miRNAs, which are rare in healthy conditions, have been associated with tumor molecular traits with potential diagnostic and therapeutic value, such as *TP53* and *KRAS* mutations, CpG island methylator phenotype, and microsatellite instability[142]. In a study conducted by Niu *et al*.[143], blood circulating miRNAs including has-miR-93-5p, has-miR-25-3p, and has-miR-106b-5p in plasma samples of CRC patients were the key candidates attributed to the disease, as far as highly conserved miRNA markers and specificity towards CRC was demonstrated when compared to non-small cell lung cancer and breast cancer. Less explored ncRNAs are small nucleolar RNAs (*e.g*.,

SNORA42[144]) , which have demonstrated upregulated expression in CRC patients with prognostic potential, exhibiting cell proliferative, tumorigenic, and metastatic properties. Notwithstanding follow-up validation studies need to be performed in order to establish a consensus on core RNA references that could be commercially used in addition to Cologuard® and Epi proColon® DNA tests.

Protein antigenic CRC biomarkers, although somewhat less informative, specific, or precise in comparison to circulating DNA-based biomarkers, have indeed been identified. For instance, carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA 19-9) are two antigens assuming high levels in the sera of CRC patients *versus* those with nonthreatening intestinal diseases[145] and healthy individuals[146]. Just as Septin9 has been included as a gene biomarker for CRC screening, CEA is a scientifically and clinically accepted protein biomarker currently applied for non-asymptomatic CRC blood testing due to a reasonable predictability of seven months to two years prior onset of the disease[146]. As to boost the performance of CEA protein biomarker testing, additional wide spectrum protein or antigenic profiles could enhance sensitivity and improve the success of CRC detection rate, which was one of the objectives of Thomas *et al*.[146]. However, the group obtained unsuccessful and limited results with respect to known lung and ovarian cancer biomarkers, cytokeratin-19 antigen fragment (CYFRA 21-1) and carbohydrate antigen 125 (CA-125), as potential CRC biomarkers, thereby excluding both of these antigens from early CRC diagnostics altogther[146]. In addition to CA 19-9 and CEA, Cyr61 has been found to be a potential biomarker in distinguishing with high sensitivities and specificities between CRC and healthy patients' sera, as well as having a positive correlation with advanced stage and cancer progression[147]. Likewise, a variety of potential unique protein signatures, but representative of the disease nature, such as insulin like growth factor binding protein 2 (IGFBP2), Dickkopf related protein 3 (DKK3), and pyruvate kinase M2 (PKM2), among others, have been accessed with overall enriched sensitivity, specificity, and better performance for effective CRC detection when used in combination and not individually[148,149]. The expansion of a multi-"omics" approaches that address the heterogeneity of CRC, although already considered, still remains to be developed, as to drastically improve CRC tracking[149].

Additionally, there are two other methods of occult blood testing on stool samples available and briefly aforementioned for the clinical detection of asymptomatic CRC, which are the guaiac-based fecal occult blood testing (gFOBT) and immunological fecal occult blood testing, also known as the fecal immunochemical test (iFOBT/FIT)[134]. Multiple published articles have described and compared both occult blood tests[150–155]: (1) gFOBT chemically measures symptomatic bleeding, indicating the presence of hidden blood (heme) in stools caused by friction against larger/advanced neoplasms, and (2) iFOBT/FIT immunologically finds blood elements (globin moiety) in stool samples via antibodies. With respect to compliance, sensitivity, specificity, and predictability in advanced colorectal neoplasm screening, and region, most studies indicate iFOBT as a better alternative to gFOBT, by presenting a better detection rate amongst distal *versus* proximal tumors[150–155]. As primary evidence, one gFOBT (*e.g.*, Hemoccult-II) and three iFOBT tests (*e.g.*, FOB-Gold, Magstream, and OC-Sensor) were compared and results demonstrated better efficiency for CRC and advanced neoplasm detection with higher rates for all three iFOBT tests in comparison to the Hemoccult-II gFOBT test[156,157]. Notwithstanding, when these tests are paralleled to stool DNA and blood protein biomarker tests, both iFOBT and DNA tests demonstrated similar results in detection[136,148,158].

Metabolic biomarkers, which are the hallmark by-products of cellular respiration and energy metabolism, have been of current interest to oncological researchers in order to aid possible predictive presence of CRC due to the imbalanced nature of the disease. Since one of the many CRC characteristics (*e.g.,* angiogenesis, hypoxia) includes a glycolytic metabolism, Jerzak *et al.*[159] demonstrated a relationship between higher carbon dioxide levels and venous deoxygenated blood. This is because of suspected increased glycolytic activity and higher serum osmotic concentration due to high levels of the solutes (sodium and urea) in the blood and urine of CRC patients exhibiting metastasis *versus* those with local tumor manifestations[159]. Although in practice this prospective method of CRC detection, especially for metastatic cancer, would be economical and simple, the underlying biological mechanism by which these results are based upon remain undetermined and future studies have yet to confirm the association with a larger

population size[159]. Moreover, Qiu *et al*.[160] obtained from tumor and adjacent non-tumor specimens 15 distinguishable metabolic biomarkers upregulated and downregulated in diverse metabolic pathways that have been associatively attributed to CRC. These metabolic biomarkers include lactate, glycerol, and glutamate, with relatively increased energy and nutrient supply roles, aspartate, β-alanine, uracil, myristate, palmitoleate, hypoxanthine, and kynurenine involved in macromolecular synthesis, and 5-oxoproline, 2-aminobutyrate, cysteine, myo-inositol, and putrescine with homeostatic redox maintenance for the regulation of oxidative stress[160]. When comparing the two studies mentioned above[159,160], lactate demonstrated contradictory results (significant positive results in one but not in the other study) in association to the diseased samples, indicating variations depending on the origin of sample (plasma *versus* tissue), which need to be considered, addressed, and validated in future metabolic research.

A more infrequent contemporary proposed cancer detection method has been through the utilization of enhanced olfactory senses of certain animals to detect specific chemicals, volatile organic compounds (VOCs), emitted from various body sample types of diseased *versus* healthy individuals[161,162]. As an example, Sonoda *et al.*[163] studied the detection accuracy of watery stool and breath samples of CRC patients using the scent of a trained Labrador Retriever with results indicating a nearly perfect sensitivity and specificity for both colorectal sample types *versus* controls, even at early stages. Despite the advantages and recent advances in identification of VOCs associated with cancer, transitioning to the usage of animals for cancer detection in the clinical environment has been challenging due to the disadvantage of relying on a living animal: lifespan, cost and time demand training, and the interspecies variances in scent detection, limiting the flexibility of practical use[163]. Besides canines, the model organism *Caenorhabditis elegans* has gained attention in uniquely identifying the presence of CRC using the same method[164]. The Nematode Scent Detection Test (NSDT) created by Hirotsu *et al*.[164] resulted in *C. elegans* directing itself towards colorectal and gastric cancer urine and tissue samples via specific chemical odors but evading the healthy controls. This test demonstrated similar sensitivities and specificities even in early cancer detection, as did

the canines[163,164]. However, in contrast, the NSDT offers a more inexpensive, clinical friendly and practical advantage in comparison to using canines[163,164].

Furthermore, the possibility of exhalation breath tests for CRC detection via differences in VOCs, using gas chromatography-mass spectroscopy (GC-MS) coupled with sensor analysis by nanoarray, and ion mobility spectroscopy has made the scientific community enthusiastic in that it would be a noninvasive, cost-effective, and optimal method for patient acquiescence, overcoming most barriers of current diagnostic methods[165,166]. Amal *et al*.[165] showed a higher concentration of both acetone and ethyl acetate in CRC patients *versus* healthy individuals, and the inverse for ethanol and 4-methyl octane concentrations with 85% and 94% sensitivity and specificity detection rates, respectively. However, just as in canine studies, there are still discrepancies between the identification of these CRC-associated VOCs in these nascent investigations, which have yet to be addressed in larger sample studies, mainly due to slight variations in the machinery used, the origin of the chosen sample and collection method, and most importantly, the samples' heterogeneity[165]. Similarly to the possibility of using breath tests for CRC detection, a recent fecal gaseous study introducing a novel sensor detection device, SCENT A1, via metabolic changes and peroxidation of tissues amongst diseased *versus* healthy volunteers has been explored with the same purpose, but the study is still ongoing with only partial results available[167].

Although many molecular detection discoveries and screening tests have gained much interest and popularity in the scientific community and medical field, thus transitioning laboratorial work into clinical practice, there are only few tests available for commercial use. Much future effort and validation remains to be done in order to maximize the public's appeal, through the development of cost-effective and efficient options that may further lower the cancer incidence rates.

3.3 Microbial-Based Screening Signatures

Various microorganisms have been referred to be associated with multiple cancers including cervical, intestinal, oral, gastric[77]. Samples from multiple origins, most commonly from mucosa and feces, have been used and tested as a prospective

alternative or complementary to frequently used CRC and IBD diagnostic methods due to the confirmed pathogenic microbes and microbial shifts occurring in colorectal diseases and cancer, *per* opposition to healthy microbiomes[168] (*cf.* Section 2.2). Bacterially-based detection of CRC, have not only been relying on the taxonomic diversity, but also on molecular products (*e.g.,* RNAs, proteins, metabolites) produced by this group of microorganisms. A variety of metabolites produced by key bacteria have been related with tumorigenesis or a benefic action towards intestinal health through proliferative reduction and promotion of apoptotic behavior have been found: SCFAs primarily butyrate, acetate, and propionate and higher butyric acid- and butyrate-producing bacteria have been identified in healthy *versus* CRC individuals, as expected[169] (*cf.* Chapter II for more details). Progressively, with an economical and practical aim, a prospective dual usage of FIT technology combined with a microbial detection approach has been considered to be a more effective and cheaper methodology for CRC detection[84]. Additionally, the adversity towards patients providing multiple stool samples and the hassle in separate test sample processing is thereby eliminated[84]. Results yielded preserved microbial communities amongst the original stool sample and the residues obtained from FIT cartridges, as well as discriminatory bacterial abundances between CRC *versus* controls, and equal predictability of the two methods, which can be taken advantage of in future screening protocols[84]. Also, to increase detection sensitivity and specificity in combination to FIT, fecal quantification by quantitative polymerase chain reaction (qPCR) of a broadly accepted CRC bacterial biomarker, *Fusobacterium nucleatum*, has been proposed to most accurately detect advanced neoplasms, 75% of which are missed by FIT alone[170,171]. Rectal swabs and brushes have also been considered and reviewed for microbial studies in IBD[172]. With the growing interest in microbial community dysbiosis, this approach could be applicable to CRC research (mainly rectal) using the outer mucus gel layer that is rich in commensal microbial species for DNA extraction and metagenomic analysis instead of fecal samples and colonoscopic biopsies[172]. Preliminary *ex vivo* cancer rodent models and human patient samples have already focused the efficiency, specificity, and sensitivity of rectal swabbing involved with host tumor gene expression, whereas cattle (*e.g.,* sheep, cow) recto-anal mucosal

swabbing tests have been directed at microbial analysis, giving rise for future complimentary applicable results specific to human CRC[173–175].

A metagenomic-transcriptomic screening method, PathoChip, a microarray identifying a broad range of key microorganisms related to tumorigenesis has been developed and validated with PCR and sequencing techniques. Although performed on oropharyngeal squamous cell carcinoma, and head and neck carcinoma tissue samples it could be applicable to other cancers and diseases that are microbe associative[176]. One drawback with this screening method is the reliance on already available sequences, which may limit less common, unknown, or unsequenced species detection, yet this is balanced by the advantage of being quick, cost-effective, customizable, sensitive, and specific[176].

Urine and saliva (oral swabbing) are less common samples used for CRC screening relying on microbially-based biomarkers related with the products produced by their metabolism[172]. Early studies performed by Cummings and colleagues in the 1970s evaluated the effect of hosts' diet on the gut microbial metabolism on both fecal and urine samples. Meat protein intake was directly proportional to the presence of total phenolic levels, carcinogenic metabolites, whilst fiber intake reverted these levels[177]. Similarly, in addition to tissue and plasma metabolic profiling of CRC *versus* healthy individuals, an indirect microbial dysbiosis can be observed through varying levels of the metabolites p-cresol and p-hydroxyphenylacetate that were obtained in the urine samples of CRC patients. As such, this demonstrates the possible shift of specific bacteria fermenting these metabolites, what can be linked to disease manifestation[178]. A representative array of oral pathogens have also been referred as possible microbial biomarkers for the differentiation amongst CRC tumors, benign polyps, and control groups[179]. Oral bacteria such as *Haemophilus, Parvimonas, Prevotella, Alloprevotella, Lachnoanaerobaculum, Neisseria,* and *Streptococcus* have been shown to be more abundant in CRC patients *versus* healthy individuals, what was confirmed by the microbiome of matched tissue samples of these patients[179]. The presence of oral biofilm species like *Streptococcus* and *Actinomyces* favor the aggregation of other bacterial species via adhesins or other virulence proteins/molecules that have been isolated from

colorectal tumors with similar environmental conditions, thereby suggesting an association with CRC morbidity[180].

4. Dissertation Objectives and Structure

The main aims of this dissertation are the following:

- Perform the state-of-the-art on relevant colorectal carcinogenesis aspects as well as approach the various histopathological-, molecular-, and microbial-based diagnostic or preventative screening options clinically available or undergoing ongoing laboratorial research;
- Provide a background literature review on the current "omics" research with valuable outcomes or applications towards the identification of potential colorectal cancer microbially-based biomarkers for the early detection of the disease;
- Compare the microbiome structure of fecal samples obtained from Portuguese CRC patients and healthy donors, using 16S rDNA gene sequencing NGS technology;
- Offer supplemental data from the Portuguese community for the prospective development of a noninvasive and effective microbial-based biomarker of CRC as diagnostic alternative that could eventually serve clinics needs in the future.

In order to achieve the abovementioned objectives, the dissertation is divided into four chapters as follows:

- Chapter 1: General Introduction;
- Chapter 2: "Omics" applied on stool microbiome in order to track new biomarkers for colorectal cancer
- Chapter 3: Stool microbiome structure in healthy *versus* colorectal cancer patients to track potential microbial biomarkers of disease;
- Chapter 4: Final Conclusions and Future Perspectives;

5. Scientific Outputs produced within the MSc Research

The work performed along the Master allowed the production of different types of scientific outputs as discriminated below:

Poster Presentation

1.) Gomes, C., Soares, A., Marques, C. Diversity of the Microbial Community (microbiome) in Stool Samples of Healthy Individuals *versus* Colorectal Cancer Patients. *IV PostGrad Symposium in Biomedicine,* ibiMed, UA. p.40 (Poster)

Research/Review Articles

2.) Review article on the contextual basis of Chapter 1: "Colorectal cancer screening and diagnosis: a histopathological-, molecular-, and microbial-based approach" to be submitted in *Clinical Colorectal Cancer* Journal

3.) Review article on the contextual basis of Chapter 2: "Omics" applied on stool microbiome in order to track new biomarkers for CRC" to be submitted in *The International Journal of Medical Microbiology*

4.) Original research article on the contextual basis of Chapter 3: "Stool microbiome structure in healthy *versus* colorectal cancer patients to track potential microbial biomarkers of disease". This study is integrated into a more comprehensive research project, in which outcomes will be joined into one sole paper.

CHAPTER II

"Omics" Applied on Stool Microbiome in Order to Track New Biomarkers for Colorectal Cancer

CHAPTER II: "OMICS" APPLIED ON STOOL MICROBIOME IN ORDER TO TRACK NEW BIOMARKERS OF COLORECTAL CANCER

Abstract

There is a clinical need to determine reliable noninvasive biomarkers for the diagnosis and therapeutic monitoring of colorectal carcinogenesis and cancer due to its ever increasing incidence rates. The application of multiple "omics" to CRC research has been extensive in systems biology and cancer genomics, but is yet to be evaluated on microbiome-based avenues, given the role of colorectal microbiome and microbial dysbiosis on CRC. Although metagenomics has had a great impulse in microbial biomarker discovery in stool samples, downstream metatranscriptomics, metaproteomics, metabolomics, and future "multi-omic" tools can prospectively aid in completing the current knowledge of underlying disturbed mechanisms for tumorigenesis. Therefore, this review hones in on potential candidate microbially-based biomarkers under study in CRC patients, as well as microbial signatures capable of providing early-warning detection of high CRC risk.

Keywords

microbial signatures, colorectal cancer diagnostic, metagenomics, metatranscriptomics, metaproteomics, metabolomics

1. Omics - the today's tools for tomorrow's detection of CRC

A relatively recent active area of medical microbiological research revolves on what is known as the "omics", a contemporary term describing the interplay of classical molecular biology and bioinformatics through the application of high-throughput NGS technologies, and advanced specialized databases and/or online libraries[181–184]. "Omics" has been explored in different branches of microbial molecular biology and biochemistry, either targeting genes and their modifications (genomics and epigenomics)[185,186], RNA transcripts (transcriptomics)[187], protein expression (proteomics)[188], metabolites (metabolomics)[189,190], lipids (lipidomics)[191], and carbohydrates (glycomics)[192]. The term, "meta-omics", oftentimes present in the current literature, refers to the application of

"omics" towards the understanding of dynamic interactions in complex (eco)systems (*e.g.,* microbial communities) under certain circumstances. For instance, meta-omics enables a more robust and interactive analysis of the flux of genes, proteins and metabolites governing community dynamics and ecological relationships on the microbiome of a given sample (*e.g.,* stool) undergoing specific conditions[193]. The implications of those biomolecules on health and complex multifaceted diseases, such as cancer, have been proven to enclose a major role[181,184,194–196]. Therefore, "(meta-)omics" have been increasingly applied to identify and quantify macromolecules (*e.g.,* DNA, RNA, proteins, lipids), as well as their structure/composition, function, multi-molecular interplay/networks, and key role in different genetic, biochemical, metabolic and physiological pathways that may trigger or influence human health and disease *status*[194,197].

CRC is one of the deadliest cancers worldwide and its incidence rates have been increasing over the last years[28,29,33,198]. One of the CRC research avenues has been lately relying on omics tools to unravel the influence of microbiome dysbiosis and host-microbiome bidirectional relationship on disease pathogenesis. Microbial dysbiosis occurs whenever there is a disequilibrium in the microbiome, via shifts in populations density and diversity, that may disrupt the immune response, enable the proliferation of pathogenic microbes responsible for the release of toxins and virulence factors, hence ultimately leading to the onset of disease (*cf.* Chapter I: Section 2.2)[52,68]. Therefore, understanding the underlying biological mechanisms by which the gut microbiota maintains host's health and how its dysregulation leads to CRC pathogenesis, through the use of large-scale genomic profiling technology, interlinked with robust bioinformatics and statistical analyses, is essential.

Successful efforts and steps have been done in this direction by the HMP[52,62,66] aiming at characterizing healthy microbiomes, but also by the Integrative Human Microbiome Project (iHMP), which focuses on the influence of the microbiome on host health and diseases. Many studies have been performed also to prove the role of gut microbiome in CRC development[199–201]. Most of them rely essentially on omics, as far as the isolation of CRC-associated microbes, not only would be too unrepresentative, as it

would be quite impractical to reproduce the natural and heterogeneous tumor microenvironment[202–204]. Overall, the microbiome dysbiosis associated with CRC has been mainly characterized by microbial taxonomic fingerprint of different sample types (*e.g.,* stool, mucosa)[170,205–208]. Taxonomical differences have been essential in identifying possible microbial biomarkers for clinical use (*e.g.*, diagnosis)[209,210]. However, research directed to specific microbial genes, proteins, metabolites, as well as their functions, in promoting or affecting disease development would provide a more complete description of underlying unclear carcinogenic mechanisms[169]. Moreover, considering colorectal carcinogenesis (*i.e.*, healthy through inflammation and adenoma, up to carcinoma stages) in differential microbial biomarker selection is of high value, as it will contribute for defining early-warning, sensitive, and accurate biomarkers to monitor and stop CRC development at precursor stages[211–215]. These biomarker features should be incremented by ability to be screened in easily-obtainable and non-invasive human samples[211,216–220]. Despite numerous microbial dysbiosis studies linked to CRC being based upon invasively obtained mucosal tissues[78,221,222], utilizing noninvasive samples like urine, blood and/or feces would be the optimal and most practical option for CRC diagnosis through microbial-based biomarkers. Feces, in particular, is the most representative noninvasive sample type of the colonic microbial community, given its intimate contact with the microbiome of host intestinal mucosa[223,224].

After initial screening, verification, and development of the most significant and highest quality potential biomolecular suggestive leads in a sample, the ultimate success of a validated clinical biomarker for CRC and precursor adenoma diagnosis is dependent on its specificity, sensitivity, reproducibility, and early-warning detection in easily-obtainable and non-invasive human samples[211,216–220]. Thus, CRC stool biomarkers uncovered through molecular testing (*e.g*., Cologuard® targets DNA of abnormal colon cells) have gained interest in the general population, due to its convenience and cost-effectiveness when compared to the adversity of undergoing invasive tests/biopsies like/obtained by colonoscopy[131–134,219,225,226]. However, this stool based test is limited to the detection of tumoral molecular aberrations and excludes the microbiome influence

on colorectal carcinogenesis, which has yet to be implemented in an approved diagnostic/ detection clinical test[132].

The main objective of this review is to highlight the "multi-(meta-)omic" advancements and research, that provided promising cues on potential microbial biomarkers of CRC in fecal human samples (Table 1 and Figure 4). The purpose of these biomarkers would serve as future noninvasive alternatives to CRC screening, diagnostics, therapy monitoring, and ultimately to personalized medicine as tools for prospective precise drug treatments (Figure 4).

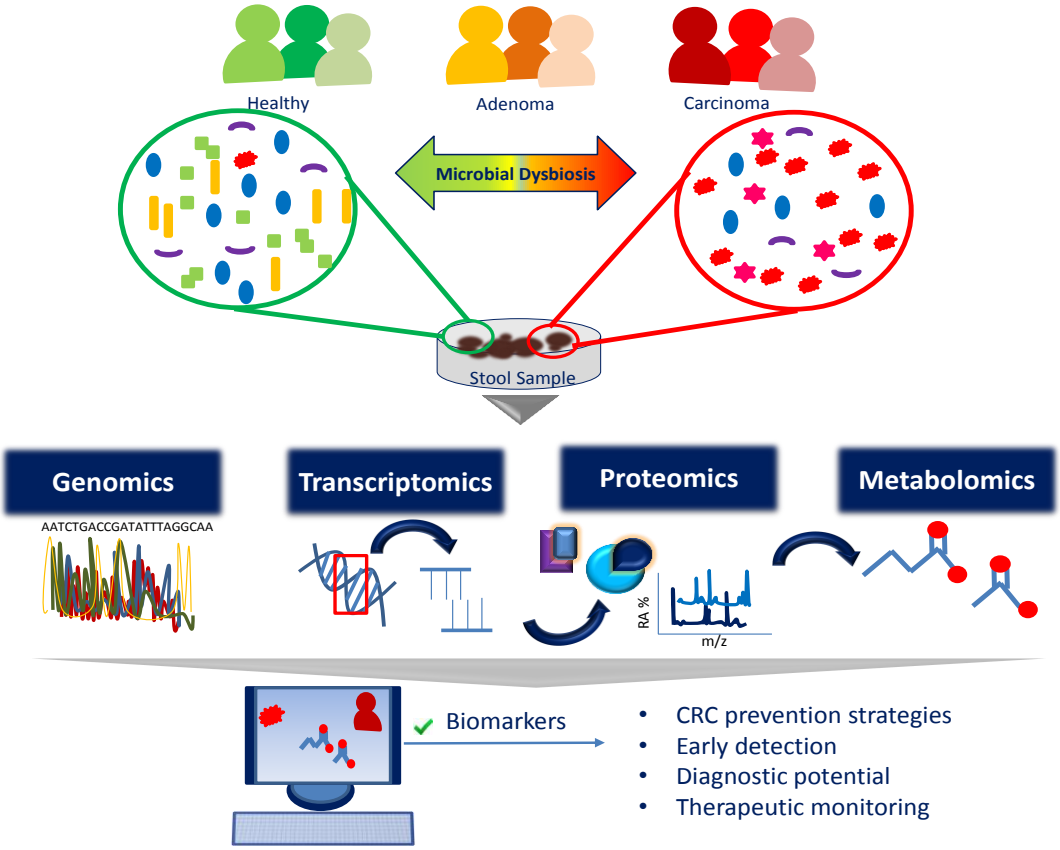

Figure 4: Omic applications on the stool samples of healthy, adenoma, and carcinoma patients for the detection of microbial-CRC biomarkers

2. Metagenomics

Metagenomics is the most mature and advanced "omics" and it has been frequently applied to in-depth study of intestinal microbiome in order to elucidate the collective genes and genetic potential enclosed by the microbial community inhabiting that biological niche[227,228]. Metagenomics, either classical [*e.g.*, DGGE, terminal restriction

fragment length polymorphism (T-RFLP), Sanger sequencing][229–231] or high-throughput [*e.g.*, shotgun sequencing, pyrosequencing, single molecule real time (SMRT) sequencing][229,232–235], tools have been applied for the sequencing of specific genes (*e.g.*, 16S rDNA gene) or the whole-genome of the microbial community[236]. In particular, metagenomics have been quite valuable towards the identification of gut microbial imbalances resulting from structural shifts in specific *taxa* or its abundance[227,237,238], microbial genes[227,238,239], and intervention in host-intestinal microbiome relationship-based pathways[227,240].

In order to understand the microbiome profile associated to a disease state, one must comprehend the gut microbial community characteristics of healthy individuals, which was early addressed by Gill *et al.*[228]. Although limited on patient sample size (two individuals), the microbial DNA and encoded metabolic functions in stool samples of healthy individuals were sequenced, identified, functionally compared, and deeply scrutinized through shotgun sequencing. Multiple strains of *Bifidobacterium longum* and *Methanobrevibacter smithii* and two bacterial groups, Firmicutes and Actinobacteria, were highly representative within the obtained sequencing reads[228]. When mapped against the functional genes databases KEGG (Kyoto Encyclopedia of Genes and Genomes) and COGs (Clusters of Orthologous Groups), the sequenced reads highlighted a general overrepresentation of sugar (*e.g.,* glucose, galactose, fructose, arabinose, mannose, and xylose), amino acids (*e.g.,* lysine, phenylalanine, tyrosine, tryptophan, glutamate, valine, leucine, methionine), nucleotides (*e.g.,* purine and pyrimidine), and coenzymes (*e.g.,* vitamin B6, biotin, thiamine), as well as molecules involved in lipid transport (*e.g.,* steroids, glycerolipids, fatty acids), SCFA biosynthesis and metabolic processing (*e.g.,* acetate, butyrate), metabolism (*e.g.,* ATP and ATPases, starch and sucrose metabolism, detoxification of toxic gaseous and xenobiotics waste, anabolism of vital amino acids and vitamins) with some discrepancies among the fecal samples of the two healthy individuals studied[228]. Along with the launch of the HMP, complementary studies based on healthy individuals with larger population sizes were accessed. The characterization of the gut microbiome genetic potential obtained from stool samples of healthy individuals have mainly identified the presence of Bacteroidetes, Firmicutes,

Actinobacteria*, Dorea/Eubacterium*/*Ruminococcus*, *Bifidobacteria*, Proteobacteria, and *Streptococci/ Lactobacilli* bacterial groups[241,242]. The microbial genes found in these samples were not only housekeeping genes (*e.g.,* carbon metabolism and amino acid synthesis, RNA/DNA polymerase and ATP synthase), but also specific genes associated with the homeostasis and host health maintenance. The latter are not present in the host's genome and encode proteins involved in adhesion to host's molecules (*e.g.,* collagen, fibrinogen, fibronectin), or in the breakdown of mucosal lining, carbohydrates (*e.g.*, glycosyl hydrolases), and glycans (*e.g.,* pectin, rhamnose, sorbitol pathways) obtained from dietary components[241,243]. In general, the healthy gut microbiome has consisted of an increased microbial gene activity in carbohydrate transportation and its metabolism as well as factors involved in the host's immunity while decreased genomic content in cellular motility, secondary metabolic and lipid synthesis[244].

With the healthy structural and functional genomic understanding accessible, evaluation of disease states like CRC became possible to tackle. Zackular *et al*.[245] characterized and compared the microbiome of fecal samples of three distinct clinical groups - healthy, adenoma, and carcinoma – to find microbiome shifts with clinical relevance for CRC diagnosis. The authors demonstrated significantly increased abundances of Ruminococcaceae, *Clostridium*, *Pseudomonas,* and Porphyromonadaceae, and decreased abundances of known producers of SCFAs associated with colonic health (*e.g.*, *Bacteroides*, Lachnospiraceae)[228] in the adenoma-carrying patients. In the carcinoma group, however, the most represented bacteria were *Fusobacterium*, *Porphyromonas*, Lachnospiraceae, and Enterobacteriaceae, most of which associated with the stimulation of CRC development and progression, whilst a decrease of beneficial bacterial groups could be determined similarly to the profile observed in the adenoma group[245]. Feng *et al*.[200] had also identified variability in genera and gene richness, as well as virulence gene abundance in the microbiome of stools from healthy, adenoma, and carcinoma clinical groups. Moreover, the carcinoma group exhibited metagenomic linkage groups, which correctly reflected and classified a higher level of *Bacteroides* and *Parabacteroides* species as well as *Alistipes putredinis, Bilophila wadsworthia, Lachnospiraceae bacterium, Fusobacterium* sp., *Parvimonas micra, Gemella morbillorum,*

Peptostreptococcus stomatis, and *Escherichia coli versus* healthy samples represented by higher abundances of *Ruminococcus*, *Bifidobacterium,* and *Streptococcus* consistent with other findings mentioned[200,245]. Feng and colleagues [200] also investigated the interference of host diet on intestinal microbial genes. A fiber-rich diet was correlated with the healthy group microbiome through the presence of metagenomic linkage genes and associative bacteria that synthesize fiber-degrading enzymes and fiber-binding domains, which contribute for colonic health[200,214]. On the contrary, meat protein-rich consumptions were associated with the CRC group. Likewise, the detrimental effects of red meats, such as beef, processed meats, and pork, have been associated to the onset of CRC risk, due to heme iron cytotoxicity and high lipid content aiding in tumor formation[3,40,41]. This heme-triggered abnormal proliferation of intestinal epithelia relies on the gut microbiota, particularly sulfide-producing and mucin-degrading *Akkermansia muciniphila*, permitting intestinal vulnerability to cytotoxic agents and other opportunistic bacteria[246]. In contrast, a fiber-rich diet has anti-CRC effects that among other effects (*e.g.*, presence of flavonoids, increased stool bulk for easier excretion) stimulate the growth of beneficial bacteria and in turn the microbial synthesis of beneficial compounds, such as butyrate[42].

As such, the identification of microbial biomarkers of CRC from metagenomic analysis should take into account the interference of other co-acting intrinsic and extrinsic (risk) factors. Host diet, ethnic/racial and genetic traits[3,37] can hence influence intestinal microbiome, either in healthy[249] or CRC-diagnosed[247,248] individuals. Although based on a small sample size, Brim *et al.*[247] identified different microbiome profiles (mainly bearing on the differential abundance of *Bacteroides*, Firmicutes and Proteobacteria groups) in healthy- and adenoma-derived fecal samples, respectively. However, despite the taxonomic difference displayed amongst the two study groups through Human Intestinal Tract Chip (HITChip) phylogenetic microarray and 16S rRNA gene barcoded 454 pyrosequencing analyses, the metagenomic analysis, which would be more indicative of the microbiome influence on cancer progression, did not reveal significant differences[247]. In a combined fecal and tissue analysis, 14 differentiating sequences (9 non-redundant ones) were detected from matched sequence reads and mapped to *Streptococcus* sp. VT_162, *Acinetobacter baumanii* AC12, and *Sphingomonas* sp. PM2-P1-29. Thus, these

bacterial species were potential biomarkers allowing to distinguish healthy from adenoma-carrying individuals[248]. Only *Streptococcus* sp. VT_162, an orally-residing bacterial species, was proven and validated to be statistically significant due to its increased presence with respect to advanced adenoma and carcinoma stages[248]. Using 16S rRNA gene pyrosequencing, another study highlighted taxonomic differences between four ethnic groups, suggesting a significant overrepresentation of bacterial members of the Ruminococcaceae family and Firmicutes phylum, and underrepresentation of Lachnospiraceae among African Americans, which in comparison to other ethnic groups (*e.g.*, Caucasians), is the race presenting the highest CRC incidence[249–251]. Indeed, these bacterial *taxa* have been indicative of gastrointestinal lesions and CRC initiation and development[249].

In summary, the application of metagenomic tools on the analysis of the stool microbiome, have been strengthening the potential of specific microbial *taxa* or microbial (enzymatic) genes as biomarkers for CRC detection. Among the microbe-based biomarkers, the ones potentially evidencing a more reliable diagnostic character are *Fusobacterium nucleatum* (subspecies *vincentii* and *animalis*)[203,210,214,252,253], *Akkermansia muciniphila* [169], *Citrobacter farmer*[169], *Parvimonas micra*[210], *Solobacterium moorei*[210], *Porphyromonas* spp.[203,214,252], *Peptostreptococcus stomati*[214], and *Bacteroides/Prevotella* genera and its association with increased IL17 immunoreactive cells[254]. In turn, within the microbial genes showing ability to be applied as markers of CRC, it has been shown the efficiency of targeting the presence of *clbA*+ bacteria containing the polyketide synthase (*pks*) gene of *Escherichia coli* and *F. nucleatum*[253,255], butyryl-CoA dehydrogenase from *F. nucleatum*[210], *rpoB* gene encoding RNA polymerase subunit β from *P. micra*[210], RNA-directed DNA polymerase from an unidentified bacterial species[210], transposases from *Peptostreptococcus anaerobius*[210], aminomethyltransferase gene[203], tryptophanase gene[203], peptide methionine sulfoxide reductase msrA/msrB genes[203], and putative membrane protein gene[203]. Overall, these microbial genes encode enzymes or proteins with major role or influence towards, genotoxicity (double stranded DNA damage) and increased mutagenicity, adapted senescence-associated secretory phenotype for enhanced proliferation, lowered butyrate oxidation state, reduced transcriptional fidelity,

oxidative damage, anti-apoptotic behavior, and loss of epithelial cell integrity for facilitated invasion and metastasis[256–259].

3. Metatranscriptomics

While metagenomics provides essential DNA sequencing data for cataloguing the existent (active and non-active) genes of the microbiome, metatranscriptomics further enables the understanding of the underlying expression dynamics and which genes and encoded metabolic pathways are active at different cellular conditions, environmental factors and/or host conditions at a particular time[260]. The recently evolved metatranscriptomic RNA-based technologies have been exponentially applied for understanding the dynamic and complex ecological relationships and shifts in the intestinal microbiome that may govern or be governed by the host-microbiome interplay in CRC and other inflammatory bowel diseases[204,261–264]. Among the techniques explored in this context, quantitative reverse transcription PCR (RT-qPCR)[78,265,266], probe-based microarrays, and RNA sequencing (RNA-Seq; *i.e.*, sequencing of cDNA libraries constructed from mRNA transcripts) have been frequently used[267,268]. However, unlike metagenomic approaches, microbial metatranscriptomics has some disadvantages. Some techniques require previous knowledge on the genetic sequences (*e.g.*, microarray probes). Besides, the low percentage of microbial mRNA transcripts comparatively to rRNA and tRNA, together with its instability due to the absence of degradation-sheltering poly-A tail, makes the isolation of mRNA quite challenging[269–272]. Moreover, and despite the great efforts in reducing the costs of metatranscriptomic analyses[272], still the available techniques are expensive and time-consuming. In spite of this, much attention has been shed on the application of transcriptomics to discover new disease biomarkers in gastrointestinal cancer studies. Although metatranscriptomic studies on the stool-associated microbiome have been applied on healthy individuals for evaluating the influence of multiple environmental and/or host-intrinsic factors (*e.g.*, races, diet, lifestyles)[273–276], the influence of several gastrointestinal (*e.g.,* Irritable Bowel Disease) [49] and metabolic disorders (*e.g.,* diabetes, obesity)[277] have also been characterized.

Bacterial RNA-seq was applied by Arthur *et al.*[255] to analyze the inoculated *E. coli* transcriptome from germ-free rodent stool samples with respect to CRC-associated inflammation and carcinogenesis. The expression of *pks* island genes and operons in *E. coli* transcriptome were enhanced in combination with the appropriate conditions (inflammation) that, in conjunction, promoted tumorigenesis, suggesting the potential of using this mRNA as a CRC biomarker[255]. This mRNA encodes an enzymatic complex capable of colibactin synthesis, a genotoxin impairing DNA and triggering the acceleration of CRC tumorigenesis once *pks* harboring bacteria come into close contact with the host's intestinal epithelia at disturbed opportunistic conditions[212,255].

Notwithstanding, the application of metatranscriptomic techniques for discovering microbiota-derived CRC signatures have been mostly explored towards the identification of differential expression of human genes as a response to the proliferation/presence of specific bacteria and/or microbiome dysbiosis. In this context, the expression of cytokine/chemokine genes and *RANTES* (regulated on activation, normal T-cell expressed, and secreted)[270], as well as the expression of Notch and Wnt/β-catenin members[278], were verified in stool samples of *Helicobacter bilis*–infected Smad3$^{-/-}$ mice and in mice receiving stool of CRC patients, respectively. However, the expression of these genes or regulation pathways presents low specificity if a clear bacteria-CRC carcinogenesis relationship is intended. Other research studies, in turn, could evidence a more direct linkage. *F. nucleatum* proliferation, for instance, has been associated with colorectal adenoma and carcinoma[279,280]. Kostic *et al.*[281] concluded that *F. nucleatum* contributes for a pro-inflammatory tumor microenvironment capable of generating intestinal abnormalities in rodent models. They further determined by RNA-Seq potential *Fusobacterium*-associated human CRC markers relying on the expression of pro-inflammatory genes encoding tumor-associated macrophages (*CD209*, *CD206/MRC1*, *IL-6*, *IL-8*, *CXCL10*), myeloid-derived suppressor cells (*CD33, IL-6*) and dendritic cells (*CD11c/ITGAX*, *CD209*, *TNF*, *CD80*) enrolled in tumor development and angiogenesis[281]. In another combined transcriptomic-proteomic approach, using real time RT-PCR and western blotting, the *cyp27a1* mRNA of microbial origin involved in apoptotic (dys)regulation was significantly overexpressed in fecal samples, with consequent

enhanced protein synthesis in tissue samples of CRC patients, comparatively to healthy individuals. Therefore, it can be potentially exploited as a prospective biomarker for convenient, efficient, and noninvasive CRC detection[282,283].

4. Metaproteomics

Despite the great venues brought to light with metatranscriptomics, the transcripts expressed by a CRC-(non)constrained intestinal microbiome might not be actually translated into functional proteins. Metaproteomics can hence offer a concise overview on the translated proteins, their composition, structure and post-translational modifications under specific circumstances, and as a result of specific interactions. Although individual protein analysis is beneficial in identifying potential protein-based signatures that differentiate between healthy and disease states through conventional methods, it may not always reflect the heterogeneous nature of a sample type due to inadequate or insufficient targeted proteins within a collected sample[284–286]. Protein "biomarker networks", which hone on protein-protein interactions and associations to distinguish phenotypes and cellular processes[284–286], along with the aid of metaproteomic technology, could potentially offer a means to discover microbial disease-based biomarkers. Such biomarkers would be valuable in achieving early-detection strategies, improved/sensitive diagnostic measures, and the identification of additional stable biomarkers from existing ones in an attempt to robustly diagnose CRC or assess treatment efficiency[284–288]. Perhaps, novel microbial biomarkers can be an addition to ongoing cancer-based proteomic tests that have been already tested and verified by ELISA, such as the multiplex fecal protein biochip[289].

Notwithstanding, the application of metaproteomics for cataloguing the microbial CRC proteome and the respective functional pathways is still on its infancy[290], despite their exploitation for discovering microbial protein biomarkers of other bowel diseases, like IBD[291–293]. Through two-dimensional difference gel electrophoresis (2D-DIGE) and LC-MS/MS applied on fecal samples of Crohn's disease patients, were revealed abundant microbial proteins (*e.g.*, AhpC, FusA, PPi-dependent PfK, TonB-dependent receptors) mainly from *Bacteroides* spp., depleted proteins (*e.g.*, GapA, flagellins FliC, Tig) from

Firmicutes and *Prevotella* spp.[291]. AhpC, FusA, PPi-dependent PfK, TonB-dependent receptor proteins have been implicated in the regulation/protection of oxidative stress, protein homeostasis (production, conformation, and repair mechanisms), carbon metabolism via glycolytic and phosphate signaling pathways, nutrient uptake, and immunity, which are dysregulated in disease conditions and can promote a conducive environment for increased barrier permeability for opportunistic pathogens to invade and colonize the mucosa[291].

Few attempted proteomic studies revolving on microbially-derived CRC biomarkers have been directed to colon mucosa (FadA- *Fusobacterium nucleatum; Escherichia coli*)[294] and serum/plasma (*e.g.*, AvrA- *Salmonella enterica;* anti-Fn antibodies)[295,296] being stool samples overlooked[218], namely due to hurdles regarding protein extraction and separation[297]. Nevertheless, the generated knowledge on bacterial proteins from other host samples, like colon mucosa, has been giving insights on the prospective use of stool samples for CRC diagnosis[295,297,298]. FadA, an adhesion protein and virulence factor expressed on *Fusobacterium nucleatum* and *Escherichia coli*, serves as an example of a CRC target because of its involvement in colonic epithelia invasion and consequentially bacterial colonization and inflammation induction[294]. This proteomic detection strategy could be complementary to *F. nucleatum* and *E. coli* metagenomic detection in stool samples[253], as a distinguishable CRC signature.

However, the discovery of microbial CRC-specific biomarkers through proteomics is quite challenging, given the vast diversity of the intestinal microbiota and, consequently, the numerous proteins produced by them that are far from being all with annotated functions or even inventoried in the available microbial protein databases [*e.g.,* uniprotkb/ Swiss-Prot protein knowledgebase[299], Translation of the EMBL nucleotide sequence (TrEMBL)[300], microbial protein interaction database (MPIDB)[301,302]] [290,303,304]. Besides, many proteins with potential for disease biomarker purposes are oftentimes less abundant, of smaller size, and more unstable than dominant (host and food-derived) proteins, which coupled with the lack of large-scale tools to hone in on these specific proteins and reference genomes for unknown microbial species, may rise further difficulties in their detection, identification, quantification, and analysis[290,303–305]. Despite

its potential, metaproteomic studies on stool samples, among others, have been limited due to hurdles in its workflow: difficulties in protein separation, mass spectrometry analysis, database usage and data analysis/comprehension[306]. Optimization of these methodological steps in human biological samples, advancement of more sensitive instrumentation, current proteomic data validation, completion of existing protein sequence databases, and development and implementation of analytic tools such as MetaProteomeAnalyzer could be fundamental in obtaining differential protein concentrations in complex, heterogeneous samples (stool) that functionally infer specific diseased states[303,306,307].

Nonetheless, these challenges have been progressively alleviated with the onset of dual metagenomic and metaproteomic technologies[303,304,308], especially when integrated with complementary (omic) approaches. In a combined mathematical, biochemical and proteomic framework, directed to tumor microenvironment conditions, host and microbial protein expression and microbiome composition, Dick (2016)[309] highlighted a potential disease-progression biomarker in stool samples based on the reduced oxidation states of carbon in the bacterially-expressed cancer proteome comparatively to that of healthy individuals. Undoubtedly, additional metaproteomic studies on CRC-associated microbiome are essential to understand the underlying dynamics of CRC progression and how the microbiota plays its role in this pathological turnover, as a way to discover clinically worthy microbial biomarkers to aid in early-detection diagnostic protocols.

5. Metabolomics/Metabonomics

A metabolite is by definition a substance, usually a small molecule with low molecular weight that is produced in a metabolic pathway, as a result of gene expression and protein synthesis[310–312]. Thus, metabolomics devotes to the total quantification, characterization and dynamics network of these molecules at different levels of organization, often with the support or integration of high-throughput next generation sequencing (metagenomics and metatranscriptomics) data and bioinformatic tools[310–314]. Metabonomics, often used interchangeably with metabolomics, is defined as the quantification of the biometabolic responses to disease-causing stimuli or genetic

modification due to populational variations with respect to genetic instability and environmental stress[314,315]. Although studying metabolites is not a novel strategy due to its advantage of being the global functional representation of the total influencing factors within systems biology[315,316], metabolomics has been promising for clinical applications (*e.g.*, cancer biomarker discovery). This is especially because it can target a variety of sample types and methods without the need of a reference genome[313,317–322].

The interplay of nutrient exchange via gut microbes' food breakdown with certain molecules absorbed at the intestinal mucosa and others excreted as waste byproducts has gained a great deal of interest with respect to gastrointestinal cancers, because the metabolites produced are significantly distinctive and reflective upon a diet that is conducive to disease[246,323]. It is of broad knowledge that the gut microbiota assumes a major role in the homeostasis of intestinal health by providing the host with energy and nutrients for normal cellular functioning as well as a protective barrier against potential pathogen invasion[53–56]. As such, any changes of the microbiota due to external (*e.g.*, diet and lifestyle)[246,324,325] or internal (*e.g.*, genetic instability, disruption of immune system)[326,327] factors, will constrain the overall gut health, and vice-versa.

In a metabonomic study conducted by Phua *et al.*[328] using gas chromatography/time-of-flight mass spectrometry (TOF-MS) and strategic data fusion through orthogonal partial least squares discriminant analysis, CRC-diagnosed and healthy volunteers were discriminated with respect to the metabolic profiles obtained in collected mucosal and fecal samples. Fructose and linoleic acid (tumor-derived), as well as nicotinic acid (gut-derived), were found to be decreased in CRC fecal samples in comparison to healthy individuals[328]. This could be explained by the overexpression of fructose transporters that use this carbohydrate, the reduced regulation of oxidative stress and mitochondrial dysfunction, genomic instability, and a reduction in nicotinic acid-synthesizing bacteria (*e.g., Bifidobacteria*) as often recorded in studies involving CRC patients[328]. Additionally, in a GC-MS-based analysis honing on gut metabolites synthesized by key microbial species in stool samples of healthy individuals *versus* CRC-affected patients, Weir and colleagues[169] found elevated levels of butyrate, glycerol, poly/monounsaturated fatty acids and ursodeoxycholic acid (bile acid) in healthy stool samples, while increased levels

of acetate (acetic acid), valeric acid, isobutyric acid, isovaleric acid concentrations, benzeneacetic acid, propionic acid, myristic acid, vitamin B5, and a cholesterol derivative were determined in malignant tumor-bearing volunteers. Thereby, the authors proved the influence of shifts on the microbial community in the interchange of metabolites and the catabolism of food components. A dysbiotic intestinal microbiota can hence carry out specific metabolic reactions in CRC individuals, as well as it can be responsible for the uptake and/or production of these metabolites from the microbiome-host tumor metabolism[169,329]. In a joint metabolic fingerprinting study conducted on fecal samples, although there was no differentiation between gut microbiota diversity and abundance amongst the control and cancer groups, the metabolic analysis suggested a distinction between the latter[329]. In agreement with the results obtained by Weir *et al*. [169], Wang and colleagues[329] demonstrated a significantly increased concentration of acetic acid, valeric acid, and isovaleric acid in CRC stool samples in comparison to healthy ones, but in disagreement with the expression levels of butyric and isobutyric acid. The authors mentioned that the increase in acetate and propionate in CRC stool samples is due to the depletion of butyric acid-producing bacteria (*e.g.*, *Ruminococcus* and *Pseudobutyrivibrio* spp.) in the large bowel[329]. Both studies obtained similar results with respect to the healthy controls: higher concentrations of unsaturated fatty acids (oleic acid, elaidic acid and linoleic acid), glycerin, monoacyl glycerol, and ursodesoxycholic acid correlated with *Ruminococcus*, but once again the authors obtained different results with respect to the expression of myristic acid and pantothenic acid[169,329]. They described the abundance of *Bacteroides*, *Dialister, Pseudobutyrivibrio*, *Fusobacterium,* and *Ruminococcus* associated with the free fatty acid levels, whilst the levels of phenylalanine, glutamic acid, serine, and threonine were associated with *Phascolarctobacterium* and *Acidiminobacter*[169,329]. Sinha *et al.*[330] used pyrosequencing for microbiota taxonomic classification and high-performance liquid phase chromatography and gas chromatography coupled with tandem mass spectrometry for metabolite identification [*e.g.*, amino acids, carbohydrates, fatty acids, androgens, xenobiotics, SCFAs]. In an attempt to access the differential correlation between specific fecal microbial species with certain identified metabolites, which were more significant in CRC samples, they demonstrated the relevance of Proteobacteria and

Actinobacteria representatives[330]. For example, higher levels of *Fusobacterium* and *Porphyromonas* were correlated with considerably strong associations between *p*-hydroxy-benzaldehyde and palmitoyl-sphingomyelin metabolites in CRC stool samples[330]. In a rodent model, the influence of broadly known CRC risk factors, *i.e.*, fat-enriched diet and genetically-based obesity, on the microbiome-metabolome dynamics was assessed[331]. It was concluded that a high-fat diet presented a stronger influence on fecal metabolome-microbiome, as a decrease amount of lactate and its producing bacteria (*Lactobacillales*) was observed, comparatively to mice fed with low-fat diets[331]. Differences were also observed in 2-oxindole-3 acetate and adenosine concentrations with lower levels of adenosine detected in the obese mice, which have been found to be pro-carcinogenic via inflammation[331]. Ijssennagger *et al.*[246] proved the onset of colorectal tumor formation and its association to toxic heme exposure on the colonic mucosa[3,40,41]. They showed it was dependent on the hydrogen sulfide-producing and mucin-degrading bacteria present in the gut microbiota, as well as on their action in damaging the protective mucus barrier[246]. A significant decrease in trisulfides, a potential biomarker produced by bacterial sulfides stimulated by the presence of heme, was observed thus increasing the mucosal permeability[246].

Of all the metabolites discussed, SCFAs have gained much attention in colonic health and disease. They are the end products formed via intestinal bacterial fermentation from undigested food material (mainly carbohydrates and fiber). SCFAs contribute directly to the overall intestinal homeostasis as key regulators of human metabolism through microbiota-colonic cross-talk, as well as they are constrained by diet changes[332,333]. Due to their importance in human health, SCFAs have been evaluated in multiple CRC studies relying on metabolomics to discover microbially-produced prospective biomarkers[169,329]. Chen *et al.*[334] used GC to assess the SCFAs influenced by diet and gut microbial changes in the fecal samples of patients diagnosed with advanced adenomas (*i.e.*, precursors of CRC). Besides the taxonomic differences between the advanced adenoma and healthy control groups, metabolic differences were highlighted in the adenoma group: decreased dietary fiber intake with lower acetate, butyrate, and propionate, as well as their respective producing bacteria, which are essential SCFAs and microbes for colonic health

and oftentimes observed in healthy individuals[334]. Epidemiological factors suchlike ethnicity or race can influence diet-based SCFAs, as levels of acetate, butyrate, and propionate were more decreased in stool samples of African Americans than in those of American Indians, Hispanics, and Caucasians[249]. Similarly, Ou *et al.*[335] studied SCFAs and bile acids produced by microbial metabolism, using GC and liquid chromatography (LC)-MS in the healthy fecal samples of two distinctive populations with dietary differences, African Americans and rural native Africans, with high and low risk for colorectal cancer development, respectively. They determined an increased concentration of primary and secondary bile acids (*e.g.*, cholic acid, chenodeoxycholic acid, deoxycholic acid, and lithocholic acid), which promote inflammation and proliferation, in the African American fecal samples and the inverse for native Africans with increased SCFAs (*e.g.,* acetate, propionate, and butyrate) that induce anti-inflammatory and anti-proliferative responses[335]. As to highlight the relevance of obtaining early-warning predictive microbial biomarkers that discriminate healthy *status* from initial stages of CRC, fecal metabolic profiling (obtained through proton nuclear magnetic resonance spectroscopy, ^{1}HNMR) on early cancer stage was conducted by Lin *et al.*[336]. A reduction in SCFAs (*e.g*., acetate, butyrate, propionate in agreement with other studies[249,334–336]), glucose, and glutamate, as well as augmented levels of amines (*e.g.*, proline, isoleucine, leucine, valine, alanine, dimethylglycine), succinate, glutamate, and lactate were observed, which are involved in microbial dysbiosis and disturbed glycolysis, glutaminolysis, and SCFAs metabolism networks[336].

6. Conclusion

The “omics” has been a contemporary technological approach to research the underlying genetic, metabolic and physiological processes that cannot be understood and explained superficially through traditional, non-high throughput sequencing methods. This is to no exception in CRC fecal microbiome-based studies in order to find noninvasive yet effective, sensitive, specific, preventative biomarkers for this disease. Although metatransciptomics and metaproteomics are still not deeply applied in this field and

more research has to be invested in fecal sample studies, preliminary metagenomics and metabolomic approaches offer great promise in biomarker discovery and strategic, preventative or diagnostic measures. Aggregating multiple “omics” could be advantageous in creating a more complete biomolecular test for complex diseases, such as CRC. Despite the individual advances made in each “omic” area, the interactome of the host’s gastrointestinal tract and the gut microbiota in comparative studies interlinking the different “omics” discussed in this review could be of great value in developing future broader-range screening alternatives and ultimately aid in filling the gaps in colorectal carcinogenesis.

Table 1: Summary of Potential "Omic" Based Microbial Biomarkers for CRC, Polyp Precursors, and High Risk Individuals

Potential microbially-derived biomarkers	Type of biomarker	Study organism	Sample origin	Technical approach	Healthy *vs* disease progression diagnostic	Ref
METAGENOMICS						
Ruminococcaceae; *Clostridium; Pseudomonas;* Porphyromonadaceae	Taxonomic	Human	Stool	V4 hypervariable region using Illumina MiSeq	Increase in adenoma	245
Bacteroides; Lachnospiraceae; Clostridiales; *Clostridium*	Taxonomic	Human	Stool	V4 hypervariable region using Illumina MiSeq	Decrease in adenoma and carcinoma	245
Fusobacterium; Porphyromonas; Lachnospiraceae; Enterobacteriaceae	Taxonomic	Human	Stool	V4 hypervariable region using Illumina MiSeq	Increase in carcinoma	245
Bacteroides; Parabacteroides; Alistipes putredinis; Bilophila wadsworthia; Lachnospiraceae*; Fusobacterium* sp.*; Parvimonas micra; Gemella morbillorum; Peptostreptococcus stomatis; Escherichia coli*	Taxonomic	Human	Stool	Meta- genomic analysis by Illumina platform	Increase in carcinoma	248
Ruminococcus; *Bifidobacterium; Streptococcus*	Taxonomic	Human	Stool	Meta- genomic analysis by Illumina platform	Increased in healthy individuals	200
Streptococcus sp. VT_162	Taxonomic	Human	Stool Tissue	16S rRNA gene; Human Intestinal Tract microarray (HITChip) and Pyrosequencing	Increased in carcinoma and adenoma	248
Firmicutes; Firmicutes/Bacteroidetes; Ruminococcaceae	Taxonomic	Human	Stool	16S rRNA gene pyrosequencing	Increased in population with high CRC risk	249
Lachnospiraceae	Taxonomic	Human	Stool	16S rRNA gene pyrosequencing	Decreased in population with high CRC risk	249
Fusobacterium nucleatum (subspecies *vincentii* and *animalis*)	Taxonomic	Human	Stool Tissue	Shotgun sequencing, Illumina HiSeq, and real-time qPCR using the Microbial DNA qPCR Assay	Increased in carcinoma	203,210,214,252,253
Peptostreptococcus stomati	Taxonomic	Human	Stool Tissue	Shotgun sequencing and Illumina HiSeq	Increased in carcinoma	214

Porphyromonas asaccharolytic	Taxonomic	Human	Stool Tissue	Shotgun sequencing and Illumina HiSeq	Increased in carcinoma	214
Akkermansia muciniphila; Citrobacter farmeri	Taxonomic	Human	Stool	V4 hypervariable region pyrosequencing	Increased in carcinoma	169
Parvimonas micra; *Solobacterium moorei*	Taxonomic	Human	Stool	Metagenome-wide association study using Illumina HiSeq, and TaqMan probe-based qPCR	Increased in carcinoma	210
Bacteroides/Prevotella genera and its association with increased IL17 immunoreactive cells	Taxonomic	Human	Stool Tissue	Pyrosequencing, real-time qPCR, Immunohistochemistry	Increased in carcinoma	254
clbA+ bacteria; polyketide synthase (*pks*) *(E. coli)*	Enzymatic gene	Human Rodent	Stool Tissue	qPCR; V6 hypervariable region sequencing using Illumina HiSeq	Increased in carcinoma	253,255
RNA-directed DNA polymerase	Enzymatic gene	Human	Stool	Metagenome-wide analysis using Illumina HiSeq, TaqMan probe-based qPCR	Increased in carcinoma	210
Peptostreptococcus anaerobius transposases	Enzymatic gene	Human	Stool	Metagenome-wide analysis using Illumina HiSeq, TaqMan probe-based qPCR)	Increased in carcinoma	210
P. micra rpoB encoding RNA polymerase subunit β	Enzymatic gene	Human	Stool	Metagenome-wide analysis using Illumina HiSeq, TaqMan probe-based qPCR)	Increased in carcinoma	210
F. nucleatum butyryl-CoA dehydrogenase	Enzymatic gene	Human	Stool	Metagenome-wide analysis using Illumina HiSeq, TaqMan probe-based qPCR	Increased in carcinoma	210
Aminomethyltransferase; tryptophanase; peptide methionine sulfoxide reductase msrA/msrB; putative membrane protein	Enzymatic gene	Human	Stool	Shotgun sequencing by Illumina HiSeq	Increased in carcinoma	203
METATRANSCRIPTOMICS						
RANTES when infected with *Helicobacter bilis*	Cytokine Chemokine RNA	Mice	Stool	Semiquantitative real-time RT-PCR	Mediated mucinous adenocarcinoma development	270
HES1, MATH1, ELF3, KLF4, IL17 and *IL17R (Bacteroides* and *Coprococcus)*	cell renewal mRNA	Mice	Stool Mucosa	Quantitative RT-PCR	Increased aberrant cryptic foci and promoted pre-neoplastic lesion and carcinogenesis	278
CD209, CD206/MRC1, IL-6, IL-8 CXCL10, CD33, CD11c/ITGAX, TNF CD80 (Fusobacterium spp.	Immune mRNA	Human	Stool	qPCR and RNA-Seq	Increased in adenoma and carcinoma	281

association)						
cyp27a1	mRNA	Mice Human	Mucosa Stool	real time RT-PCR and Western blotting	Increased in carcinoma	282,283
polyketide synthase (*pks*) (*ClbG, ClbH, ClbL, ClbM* and *ClbN) (E.coli)*	mRNA	Mice	Stool	Illumina RNA-seq.	Increased in carcinoma	255
METAPROTEOMICS						
Oxidation State of carbon of microbial proteins	oxidative post-translational modification	Human	Stool	Data composed of multiple proteomic studies	Decreased in carcinoma	309
METABOLOMICS						
nicotinic acid [reduction in nicotinic acid-synthesizing bacteria (*e.g., Bifidobacteria*)]	Metabolite	Human	Stool Mucosa	GC/ time-of-flight mass spectrometry	Decreased in carcinoma	328
acetate (acetic acid), valeric acid, isobutyric acid, isovaleric acid , benzeneacetic acid, propionic acid, myristic acid, vitamin B5, phenylalanine, glutamic acid, serine, and threonine due to depletion of butyric acid-producing bacteria (*e.g.*, *Ruminococcus* and *Pseudobutyrivibrio* spp.,); *Phascolarctobacterium* and *Acidiminobacter*	Metabolites	Human	Stool	GC-MS	Increased in carcinoma	169,329
butyrate, glycerol, poly/monounsaturated fatty acids and ursodeoxycholic acid (bile acid) associated with *Ruminococcus*	Metabolites	Human	Stool	GC-MS	Increased in healthy samples	169,329
p-hydroxy-benzaldehyde and palmitoyl-sphingomyelin associated with *Fusobacterium* and *Porphyromonas*	Cell shedding metabolite	Human	Stool	Pyrosequencing and high-performance liquid phase chromatography; gas chromatography coupled with tandem mass spectrometry (HPLC-GC/MS-MS)	Increased in carcinoma	330
p-aminobenzoate (PABA) and conjugated linoleate-18-2N7 (CLA) accoiated with *Clostridia*, Lachnospiraceae	Inflammation and innate immunity metabolites	Human	Stool	Pyrosequencing and HPLC-GC/MS-MS	Decreased in carcinoma	330

Lactate associated with *Lactobacillales*	Intracellular metabolite of glucose	Mice	Stool Mucosa	Ultrahigh performance liquid chromatography/tandem mass spectrometry (UHPLC/MS/MS) and GC	Decreased in high fat diets (associated with CRC)	[331]
adenosine	Anti-inflammatory role metabolite	Mice	Stool Mucosa	Ultrahigh performance liquid chromatography/tandem mass spectrometry (UHPLC/MS/MS) and GC	Decreased in obese mice (CRC risk factor) thus pro-carcinogenic via inflammation	[331]
acetate, butyrate, propionate accoaited with *Clostridium* spp., *Faecalibacterium prausnitzii*. *Roseburia*, and *Eubacterium* spp., Lachnospiraceae	anti-inflammatory and anti-proliferative metabolites	Human	Stool	GC and LC-MS	Decreased in adenomas and increased in healthy native rural African population (not associated with the CRC-related diet)	[249,334,335]
primary and secondary bile acids (*e.g.*, cholic acid, chenodeoxycholic acid, deoxycholic acid, and lithocholic acid) associated with 7a-dehydroxylating bacteria	inflammation and proliferation promoting metabolites	Human	Stool	GC and LC-MS	Increased in African Americans (CRC ethnical) risk	[335]

CHAPTER III

Stool Microbiome Structure in Healthy *versus* Colorectal Cancer Patients to Track Potential Microbial Biomarkers of Disease: A Portuguese Cohort Study

CHAPTER III: STOOL MICROBIOME STRUCTURE IN HEALTHY *VERSUS* COLORECTAL CANCER PATIENTS TO TRACK POTENTIAL MICROBIAL BIOMARKERS OF DISEASE: A PORTUGUESE COHORT STUDY

Abstract

With the ever increasing worldwide CRC incidence and mortality rates, research keeps pushing forward the understanding of the disease, the potential factors interacting in its pathogenesis and the discovery of sharper early-warning detection tools. The relationship between the gut microbiome dysbiosis and colorectal carcinogenesis has been increasingly explored, but still there is a gap concerning the coverage of different geographical areas, which may *per se* influence gastrointestinal microbiome and host-microbiome interactions. In order to determine and compare the diversity of the bacterial communities present in the fecal samples of 6 Portuguese CRC patients *versus* 6 healthy individuals, as well as its variations in different CRC stages, Illumina MiSeq was applied for 16S rRNA sequencing. A loss in bacterial diversity was clearly witnessed in the CRC group when compared to the healthy group. Moreover, the two clinical groups presented overall homogenous bacterial community structures at a taxonomic level though certain *taxa* abundance differences could also be noticed. As such, genera from Bacteriodetes (*Prevotella*, *Alloprevotella*), Proteobacteria (*Sutterella*, *Desulfovibrio*), and Actinobacteria (*Olsenella*) could serve as potential microbial biomarkers due to their increased abundance or exclusivity in CRC patients. The variability detected between clinical groups should be deeply studied by deep sequencing analysis to find out significantly consistent bacterial species with biomedical relevance, namely for disease control.

Keywords

Metagenomics, 16S rRNA gene sequencing, clinical groups, microbial signatures, colorectal cancer staging, clinical diagnosis

1. Introduction

The gut microbiota and/or microbiome, which encompasses the vast complexity of microorganisms (*e.g.,* bacteria, archaea, viruses) and their collective genetic material [50–52,62,65], is vital to maintaining human gut homeostasis (*e.g.,* nutrient exchange, host immunity, pathogen protection)[53–56] and once disturbed can be used to understand its involvement in promoting gastrointestinal disease (*e.g.*, IBD, CRC)[52,68]. Therefore, significant changes in the microbial community sustaining different molecular/biochemical (*e.g.*, toxins, virulence factors) and functional (*e.g.,* modulation of host immune system) traits have been reported as influencing factors in CRC onset and development[52,68]. Comparing the microbiome of diseased *versus* healthy individuals, even across CRC evolution (*e.g.,* the adenoma-carcinoma development and progression), has allowed for differential microbial biomarker identification and quantification for alternative and efficient, screening, diagnostic, and therapeutic avenues[200,207]. Also, the linkage between the microbiome and host intrinsic/extrinsic (risk) factors (*e.g.,* diet, race/ethnicity) that aid in creating a microenvironment conducive for carcinogenesis has also been addressed by other authors[247,331]. Nonetheless, further investigation is necessary to linearly transition these bacterial candidates into stable, reliable microbial biomarkers for the implementation of novel colorectal screening methods or in combination to existing ones [169,210,245].

In order to deeply unravel the association between the microbiome and CRC, a transition from conventional methods (*e.g.*, microbial culturing)[337] to modern molecular methods (*e.g.,* 16S, DGGE)[338,339], and NGS "omics" technology (*e.g.*, shotgun sequencing, pyrosequencing)[200,249] has been ongoing. Predominantly, metagenomics-based analyses characterizing the structural and functional microbial-CRC influence have been explored extensively [69,209,210,214] (*cf.* Chapter II: Section 2). Among the diverse microorganisms colonizing the gastrointestinal tract, mainly bacterial species (*e.g., Fusobacterium nucleatum*[222,279,294], enterotoxigenic *Bacteroides fragilis*[298], *Enterococcus faecalis*[340]) have been identified and associated with CRC, potentially supporting early detection strategies of the disease.

Although mortality rates have been slightly declining since 2012 due to innovations in CRC screening programs in the European community, CRC is still one of the most deadly cancers prospected for 2018[341]. However, populations from different geographical areas do assume varied ethnic backgrounds, genetic traits, culture, and lifestyle, which have all been considered influencing factors of the microbial-CRC relationship. Therefore, these factors should be taken into consideration for developing more robust diagnostic tools, as a way to decrease CRC incidence and mortality in the near future[342–345].

The northern region of Portugal presents a high incidence of CRC[346]. However, no studies yet provided a more comprehensive analysis of the CRC-associated microbiome in the affected Portuguese population, being only one study available for gastric cancer[347]. Thereby, fulfilling this gap will open great opportunities for the establishment of preventive or early detection measures for this particular geographic location[346], especially because specific diet, lifestyle, cultural habits, and other CRC-prone risk factors have been gradually adopted[348]. Thus, it is hypothesized that there might be a distinctive microbiome structure differentiating CRC and healthy Portuguese patients, what could be further explored as potential diagnostic tools. As such, the aim of this study is to determine and compare the microbiome profiles of fecal samples obtained from healthy individuals *versus* CRC-affected patients, in the center of Portugal. The identification of particular microbial *taxa* on CRC and health *status*, as well as CRC- associated microbial shifts in different disease stages will be analyzed in order to contribute in the quest of finding new and potential microbial-based CRC biomarkers.

2. Material and Methods

2.1 Clinical Groups and Samples

This study was approved by the Ethics Committee and Administration Council of the Centro Hospitalar do Baixo Vouga, Aveiro, E.P.E. (CHBV), being all the procedures in compliance with Helsinki Declaration. Consent forms were given and signed by each volunteer participant. Stool samples were obtained from 12 individuals of two clinical

groups from CHBV: (1) healthy group (H) constituted by volunteers without previous or current medical CRC, (2) patients diagnosed with colorectal carcinoma (C) prior undergoing surgery. For the C group, cancer number staging (mostly 1, 2, 3, 4 stages and associated substages) was obtained based on the TNM classification system. Handmade sterile kits with illustrative instructions were given to the control group for fecal collection. Upon reception, the fecal samples were aliquoted, immediately frozen in liquid nitrogen and subsequently stored at -80ºC until future usage.

2.2 Microbial DNA Extraction from Stool Samples

DNA extraction from all fecal samples (6 healthy individual and 6 CRC patient samples) was performed using the QIAamp® PowerFecal® DNA Kit (Qiagen, Germany). The samples were thawed in ice and approximately 250 mg of each fecal sample was used for microbial DNA extraction, as recommended. The DNA extraction protocol was performed according to manufacturer instructions, being the extracted DNA stored at -20ºC. The NanoDrop spectrophotometer was used in order to determine the quantity and quality of the extracted DNA.

2.3 16S Library Preparation for Next Generation Sequencing

In order to prepare the 16S gene library, the hypervariable region V4 was selected to identify the bacterial *taxa* constituting the microbiome of the stool samples. Two separate amplicon and index attachment PCRs and two primer/dimer clean up stages were conducted followed by quantification and quality analysis (using NanoDrop spectrophotometer and agarose gel electrophoresis), according to the recommended Illumina MiSeq system instructions[349]. Non-template controls were included in the analysis in order to remove potential contaminant OTUs from sequenced samples. In summary, the 1st PCR consists of the amplification of the 16S gene V4 region of each genomic DNA (gDNA; *ca.* 10 ng) sample with specific forward and reverse primers attached to overhang adapters (1μM each). This PCR was amplified using NEBNext® High Fidelity Master Mix with the following thermocycler run program: (1) initial denaturation of 1 min at 98 °C, (2) 25 cycles of denaturation for 15s at 98 °C, annealing for 30s at 55 °C,

and extension for 30s at 72 °C, (3) final extension for 5 mins at 72 °C, and (4) ∞ hold at 15°C. The 2nd PCR allows for the adherence of tag indices and sequencing adapters for the identification of each sample, when pooled. The 2nd PCR run included the following: (1) initial denaturation of 1 min at 98 °C, (2) 8 cycles of denaturation for 15s at 98 °C, annealing for 30s at 55 °C, and extension for 30s at 72 °C, (3) final extension for 5 mins at 72 °C, and (4) ∞ hold at 15°C. A final clean up stage was performed in order to purify the final library. The prepared 16S library (equimolar paired-end amplicons) was sequenced using Illumina MiSeq sequencer.

2.4 Analysis of Sequencing Data

Sequence reads were demultiplexed and stripped of the index and adapters. The analysis of the sequences was performed with the USearch tool developed by Robert C. Edgar[350], being then generated an OTU table. Taxonomic prediction of the OTUs was obtained with the RDP database and pipeline[351]. Alpha diversity was evaluated through the number of OTUs (richness), Shannon index, Chao index, and phylogenetic diversity parameters, which were determined with QIIME and/or Rhea pipeline. These parameters were depicted in boxplots[352] for each clinical group considered and compared through the application of Kruskall-Wallis analysis of variance ($p<0.05$). Beta diversity was also computed with QIIME commands, being calculated the weighted-Unifrac similarity distance between samples, which takes into consideration phylogenetic relations and *taxa* abundance. Constrained Analysis of Principal Components (CAP) plots were thereby created in RStudio[352] environment, being the samples plotted in the 2-dimensional graphical space according to their similarity and constraining variables (clinical group and cancer staging). For this constrained analysis ANOVA was applied to test for statistically significant differences between the levels of the factors/variables considered. For taxonomic analysis, the relative abundance of microbial *taxa* (at phylum, family and genus *per* phylum) in the stool samples of volunteers from H and C clinical groups, and at different cancer staging were represented in barplots. Whenever *taxa* were represented with an abundance lower than 0.5% or 0.1%, they were merged together and allocated to the non-assigned ('NA') category. These analyses were also performed in RStudio[352] environment.

3. Results and Discussion

Although stool samples can be the most representative and noninvasive for microbial community analyses towards the evaluation of CRC *status*, they may pose some challenges regarding the efficiency of DNA extraction due to the presence of host DNA, undigested food particles, proteins, and other inhibitors of downstream NGS procedures[353,354]. In this study, adequate DNA concentrations were obtained ranging from 62.4-170.8 ng μL^{-1} (Table 2). Indeed, the purity of DNA observed from the A_{260}/A_{280} ratio ranged from 1.64-1.83 (Table 2), most of which were ~1.8, indicating a credible microbial DNA quality without a notable presence of proteins or other contaminants that can interfere with PCR amplification processes[353]. Although the A_{260}/A_{280} readings demonstrated the typical values of DNA purity for the majority of the cases, the A_{260}/A_{230} ratio measurement did not exemplify the ~1.8-2.2 consensus purity values in all cases, which may require further DNA extraction optimization[353]. With the highest quality of DNA established (A_{260}/A_{280}), metagenomic sequencing data can be achieved and analyzed using a series of measures that allow comparisons between the healthy and carcinoma groups, such as alpha/beta diversity, and taxonomic relative abundances.

Table 2: Quality and quantity of gDNA extracted from stool samples of CRC-diagnosed (C) and healthy (H) individuals.

	Stool	***Quantity and quality of DNA (NanoDrop Spectrophotometer)***					
Sample	**Mass (g)**	**[DNA] (ng/µL)**	**A_{230}**	**A_{260}**	**A_{280}**	**A_{260}/A_{280}**	**A_{260}/A_{230}**
C1	0.25974	76.8	1.071	1.536	0.842	1.82	1.43
C2	0.25038	123.3	1.374	2.466	1.353	1.82	1.79
C3	0.25568	103.2	1.145	2.063	1.128	1.83	1.80
C4	0.19147	100.9	1.341	2.019	1.122	1.80	1.51
C5	0.17180	68.0	0.795	1.359	0.759	1.79	1.71
C6	0.23200	62.4	0.891	1.249	0.688	1.81	1.40

H1	0.22003	134.7	1.742	2.693	1.498	1.80	1.55
H2	0.25803	126.8	2.030	2.536	1.449	1.75	1.25
H3	0.25629	74.7	1.208	1.493	0.859	1.74	1.24
H4	0.25596	104.1	1.988	2.082	1.270	1.64	1.05
H5	0.26103	149.2	1.870	2.984	1.674	1.78	1.60
H6	0.25441	170.8	1.870	3.416	1.869	1.83	1.83

In order to address the overall alpha diversity within and between both clinical groups, the OTU richness, phylogenic diversity (PD), Shannon's diversity and Chao indices, were applied and analyzed. These indices are statistical estimates of species diversity, abundance and evenness, and phylogenetic (evolutionary) relationships between *taxa* found in a given community[355–358]. Healthy stool samples had a generally greater bacterial diversity, given the significantly higher number of OTUs and Chao values, when compared to carcinoma samples, demonstrating a loss of bacterial diversity/richness (Figure 5). Although PD and Shannon diversity exhibited differences amongst both clinical groups, they were non-significant in this study (Figure 5). These results associating CRC individuals with a globally reduced bacterial community diversity concur with Ahn *et al*.[252], demonstrating that Shannon diversity index of fecal microbiome differed amongst healthy and CRC volunteers (extraction kit used: QIAamp® PowerFecal® DNA Kit; sequencing system: 454 Roche FLX Titanium pyrosequencing system). Similarly, Gao *et al*.[68], which obtained statistically significant Shannon diversity results proving a loss in bacterial diversity in the CRC group, but non-significant differences between healthy *versus* CRC tumor groups with Chao index in a mucosal study. Unlike this trend, Russo *et al*.[359] reported insignificant variations in the number of OTUs, Chao, and Shannon diversity between CRC and healthy stool samples, and even attained higher diversity in CRC biopsy samples of mucosa. The authors used a different DNA extraction kit (PowerLyzer® PowerSoil® DNA Isolation Kit) among other factors, what might have ended up in variations on the obtained DNA concentration and purity, what may influence

downstream NGS (Illumina MiSeq) applications, accounting for such variability in results[359]. Even in conventional advanced adenoma manifesting patients, Peters *et al*.[360] observed a decreased microbial diversity (**<** no. OTUs, Shannon's diversity) compared to those individuals without any indication of harboring CRC precursors. Although Peters *et al.* [360] evidenced a distinctive microbial diversity in premature colorectal tumorigenesis, Goedert *et al*.[361] obtained no significant variations (OTUs, Shannon, Chao, PD) between fecal samples of adenoma and healthy groups. However, further studies have to be performed in order to confirm these associations to accurately prevent CRC onset early in its tracks.

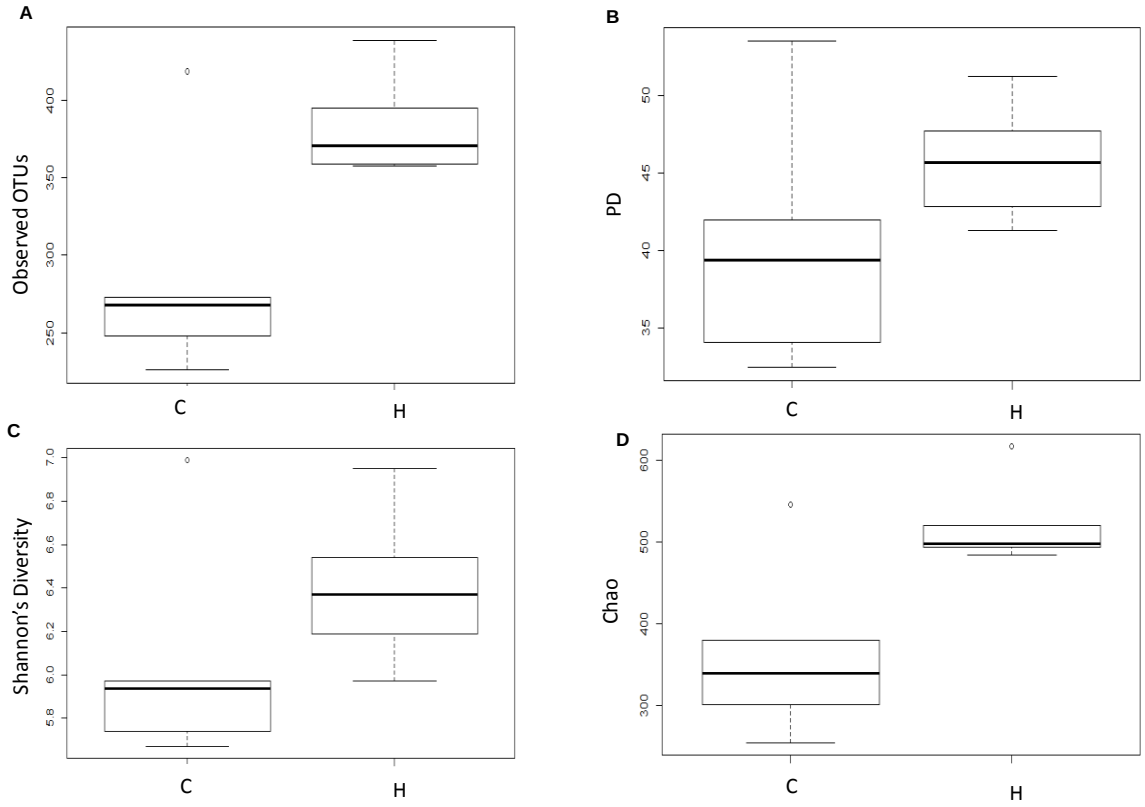

Figure 5: Box-plots of alpha-diversity indices comparing the stool microbiome amongst the stool samples of Carcinoma (C) and Healthy (H) clinical groups. **(A)** Observed OTUs (Kruskal-Wallis chi-squared = 4.3485, df = 1, p-value = 0.03704), **(B)** Phylogenic Diversity (PD) (Kruskal-Wallis chi-squared = 3.1026, df = 1, p-value = 0.07817), **(C)** Shannon's Diversity index(Kruskal-Wallis chi-squared = 3.4029, df = 1, p-value = 0.06508), and **(D**) Chao index (Kruskal-Wallis chi-squared = 4.3333, df = 1, p-value = 0.03737)

In addition to alpha diversity, beta diversity, a measurement of how variable (similar or different) the microbiome structure is between the stool samples of both clinical groups and amongst the CRC group along different CRC stages[358], was evaluated using CAP. The microbial *taxa* profiles in the stool samples of CRC-diagnosed individuals were clearly distinguished from the healthy (Figure 6A), which is in accordance with other authors[210,214,254]. Clear CRC-microbiome variations have even been shown by CAP, not only between CRC and healthy individuals, which was significant in this study, but also gradually between different CRC development and progression noted by Zeller *et al.*[214], as observed from early to late stages of CRC (Figure 6B). However, in this study the microbiome differences observed amongst CRC staging were not statistically significant ($p>0.05$). Nevertheless, the ability to separate both clinical groups, even when CRC patients are at different disease stages, may offer a highly significant diagnostic/screening tool based on the microbial profiles.

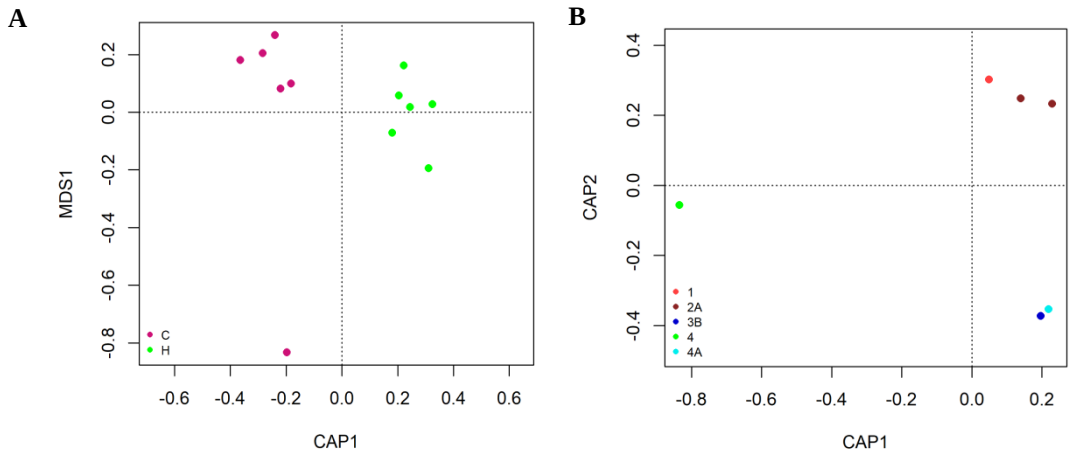

Figure 6: Constrained Analysis of Principal Components (CAP) of the stool microbiome plotted according to the variation between **(A)** healthy (H) and CRC (C) clinical groups (explained overall variation: 13.9%; ANOVA on the constrained ordination method: F=1.611, P=0.012), and **(B)** CRC stages diagnosed in the CRC group of patients (explained overall variation: 85.7%; ANOVA on the constrained ordination method: F=1.499, P=0.133)

The relative abundances of bacteria for both clinical groups were evaluated on multiple taxonomic levels: phylum, family, and genus (analyzed *per* phylum). At a phylum level, the microbiome of stool samples in this research displayed 5 principal phyla

irrespective of the clinical group, and they can be ranked according to their decreasing order of relative abundance as: Firmicutes > Bacteroidetes > Proteobacteria > Actinobacteria > Fusobacteria (Figure 7). As such, the microbiome between healthy and carcinoma samples were constituted by similar phyla, being Firmicutes and Bacteroidetes the most represented and whose equilibrium is pivotal in modulating disease manifestation and development. Nonetheless, the Firmicutes/Bacteroidetes ratios were slightly inverted in healthy and carcinoma groups, once Firmicutes was more represented in the healthy group and Bacteroidetes was overrepresented in the carcinoma group. Bamola *et al*.[362] found similar results in the stools of CRC and IBD patients. Healthy patients with a restricted vegetarian diet possessed a greater abundance of members from the Firmicutes phylum followed by Bacteroidetes, being the former known for butyrate-producing bacteria that aid in protecting the regular functioning of the colonic epithelia. In turn, diseased patients and healthy non-vegetarians exhibited the reversal trend[362]. However, this relationship is multivariable and lacks consistency with other findings[68]. Proteobacteria has been reported to be increased in colorectal adenomas[363] as well as in CRC clinical scenarios, and was shown to be progressively increased with cancer staging (Figure 8), in agreement with Kinross *et al.*[364]. Fusobacteria has had a strong associative effect with colorectal carcinogenesis and has been a well-established candidate to be a taxonomic biomarker for the disease[68–70,279,281]. Additionally, Actinobacteria, although not as abundant as other predominant phyla, is an underlying agent in preserving gut homeostasis, especially because some members (*e.g.,* SCFA producing-Bifidobacteriaceae) are capable of conserving the structure of the intestinal barrier, prone to toxic substances and opportunistic pathogens[365]. Although Fusobacteria has been linked to CRC and Actinobacteria with healthy states, the former was underrepresented in CRC patients (Figure 7).

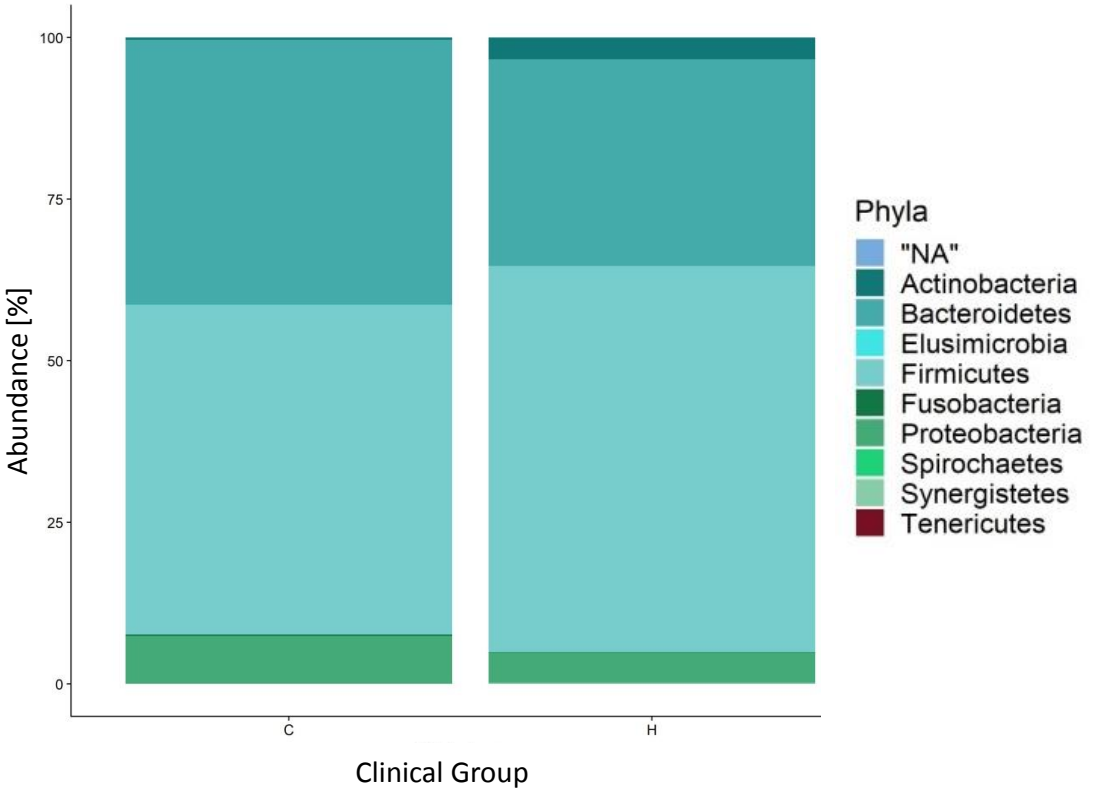

Figure 7: Phyla-level relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

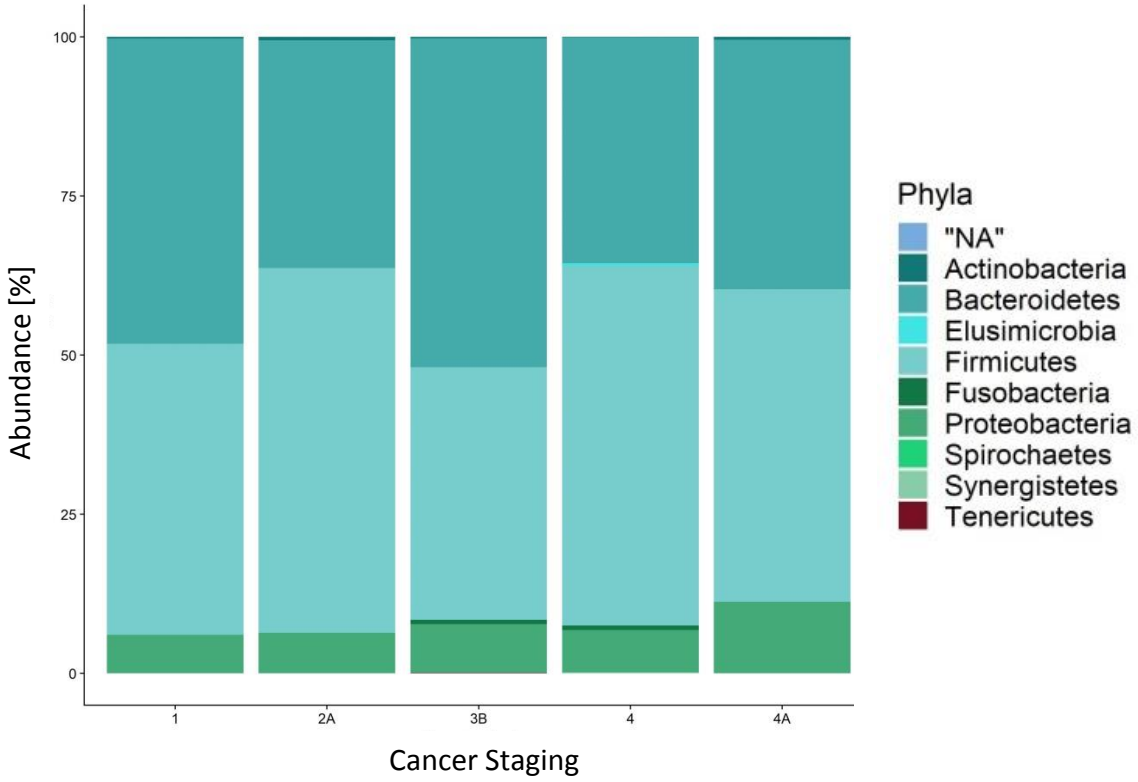

Figure 8: Phyla-level relative abundance of bacteria from stool samples of CRC-diagnosed patients, according to the diagnosed cancer stage.

At the family level, the most abundant observed in this study were the following: Ruminococcaceae (H>C), Lachnospiraceae (H>C), Bifidobacteriaceae (H>C), Prevotellaceae (H<C), Eubacteriaceae (H<C), Enterococcaceae (H<C), Porphyromondaceae (H≈C), Rikenellaceae (H≈C), and Bacteroidaceae (H≈C). Some distinctive variability was hence observed between the clinical groups (Figure 9). Similarly to our outcome, Ruminococcaceae and Lachnospiraceae families presented reduced abundances in mucosal samples of IBD patients, as well as in cancerous tissue samples of CRC patients relative to the intestinal lumen (stool and rectal swabs), adjacent noncancerous or pre-malignant tissue, thereby suggesting its potential relevance as markers of intestinal dysbiosis[366–368]. Flemer *et al*.[179] proved a negative correlation between Lachnospiraceae members and a series of oral residing pathogens (*Porphyromonas*, *Parvimonas*, and *Fusobacterium*) implicated with CRC occurrence, as well as a negative association with a high fat/sugar diet, which is favored by these oral pathogens. The prevention of harmful colonic bacterial species colonization may be impeded by beneficial bacterial communities, such as species of the Lachnospiraceae family[179]. Similarly, the antibiotics ruminococcin A and C synthesized by *Ruminococcus gnavus*, a Ruminococcaceae species, has evidenced anti-pathogenic activity against *Clostridium septicum*[369,370]. Despite this species being a regular inhabitant of the gut, it possesses virulence factors that can influence colorectal tumorigenesis and other gastrointestinal diseases[371]. The probiotic benefits of members of Bifidobacteriaceae and Lactobacillaceae families once stimulated with prebiotics, have been demonstrated to have CRC retarding effects, enhanced barrier integrity, pro-apoptotic stimulation, and anti-inflammatory properties[372]. In contrast, an enrichment of Prevotellaceae abundance has been found in the mucosa and stool of CRC patients as observed by Chen *et al*.[366]. The Prevotellaceae family includes opportunistic pathogens acting under appropriate conditions (*e.g.*, enhanced permeability of colon epithelia and infiltration of harmful bacteria/metabolites), and that have been linked to metabolic disorders (*e.g.*, obesity, high-fat diets)[366]. Similarly, Eubacteriaceae has been increased in CRC cases when compared to control groups[366,373]. Moreover, Enterococcaceae family members, such as *Enterococcus faecalis*, have been implicated in colon malignancies due to the production of superoxides, hydrogen peroxide, and

hydroxyl radicals exhibiting genotoxicity on intestinal epithelial cells and genomic instabilities (*e.g.,* chromosomal instability)[340]. Although between clinical groups the relative abundance of Bacteroidaceae is herein similar (Figure 9), pathogens such as the fragilysin-producing toxigenic *Bacteroides fragilis*, were previously detected to be pronounced in colorectal tumors, and related with carcinogenesis through the induction of an uncontrolled pro-inflammatory response, and acting as a driver in disrupting colonic epithelia integrity via the β-catenin/Wnt pathway[374,375]. In agreement with a Spanish cohort mucosal study conducted by Allali *et al*.[345], Desulfovibrionaceae, a sulfate-reducing, hydrogen sulfide producing, aerotolerant bacterial family and member of the Proteobacteria phylum, was more prominent in CRC patients when compared with healthy individuals, although it was not a major family represented (Figure 9). Hydrogen sulfide has been indirectly shown to promote colonic cell proliferation and differentiation, mediated by PI3K/Akt and ERK signaling, and downregulation of p21 and nitric oxide[376].

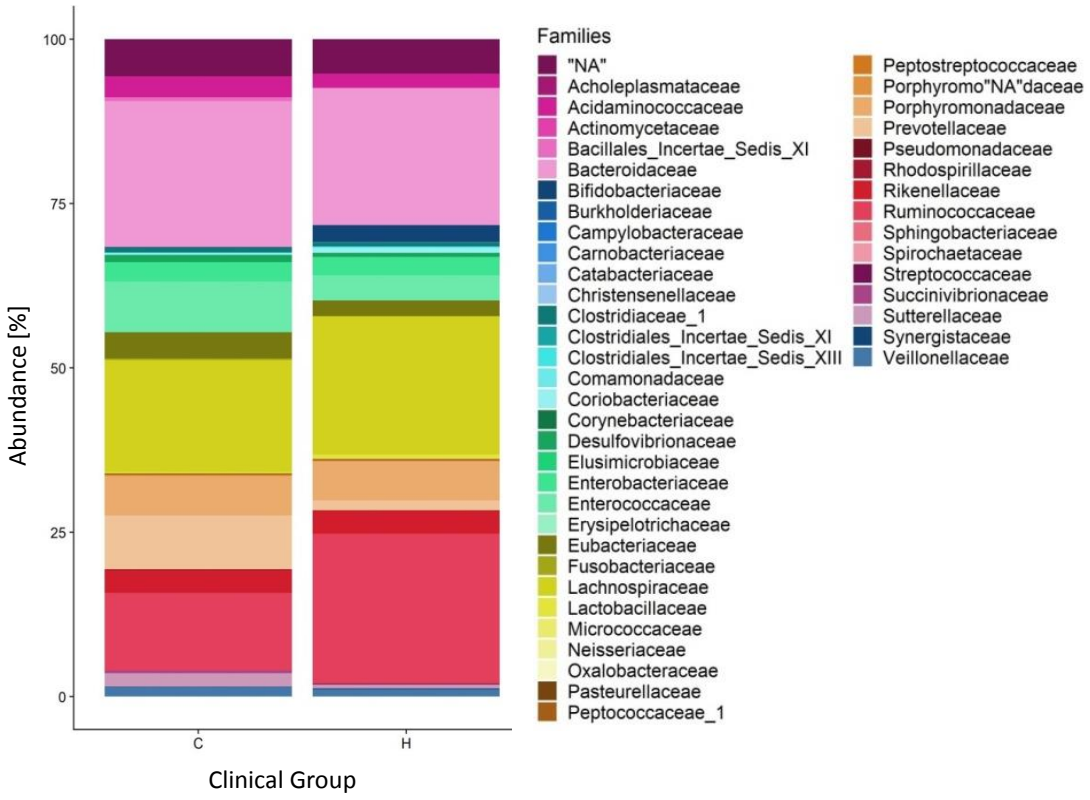

Figure 9: Family-level relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

At the genus level, the most dominantly observed were *Ruminococcus*, *Roseburia*, *Oscillibacter*, *Lachnospiracea incertae sedis*, *Faecalibacterium*, *Eubacterium*, *Enterococcus*, *Clostridium* XIVa, *Blautia*, *Acidominococcus* (Firmicutes) (Figure 10), *Bacteroides*, *Alistipes*, *Parabacteroides*, *Prevotella*, *Alloprevotella* (Bacteroidetes) (Figure 12), *Morganella, Sutterella, Desulfovibrio, Succinovibrio*, *Parasutterella*, *Proteus* (Proteobacteria) (Figure 14), *Bifidobacteria, Eggerthella* (Actinobacteria) (Figure 16), and *Fusobacterium* (Fusobacteria) (Figure 18). Healthy stool samples were overrepresented of

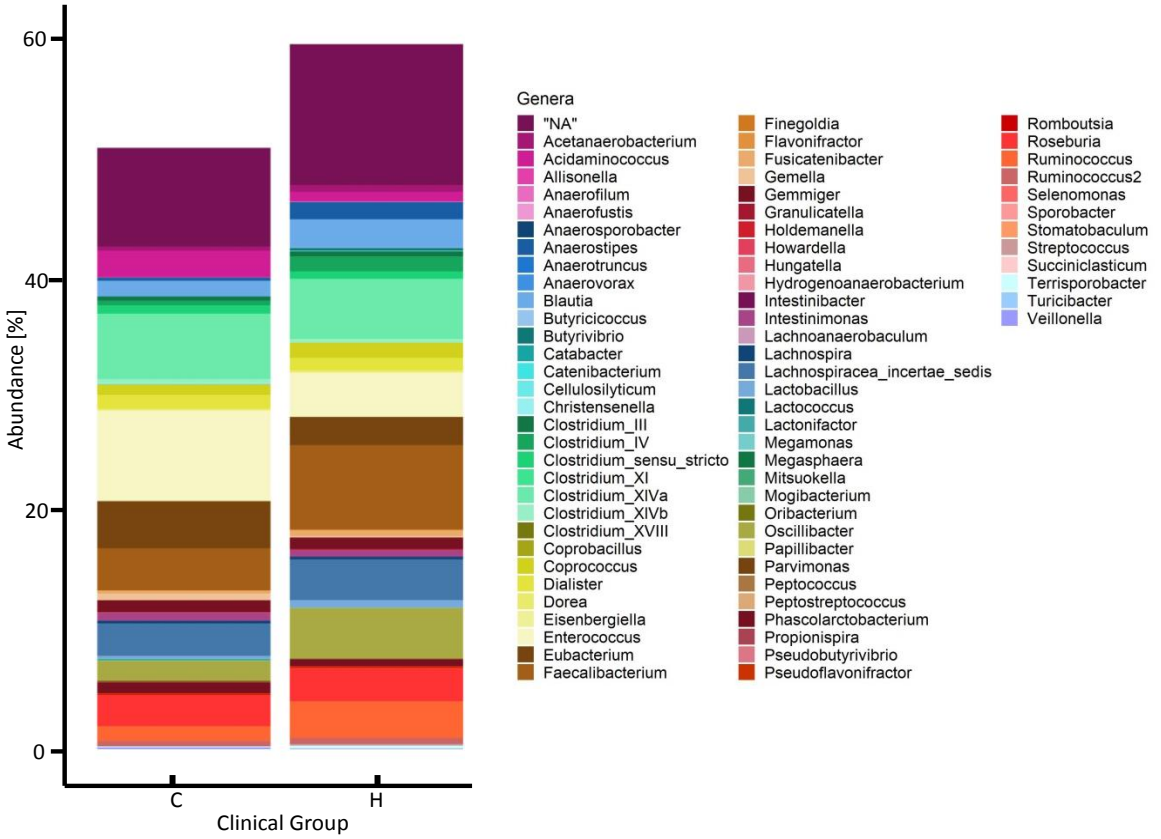

Figure 10: Genera-level (Firmicutes phylum) relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

Ruminococcus, *Oscillibacter*, *Lachnospiracea incertae sedis*, *Faecalibacterium*, *and Blautia* genera (belonging to the Fimicutes phylum); whereas the *Eubacterium*, *Enterococcus*, and *Acidaminococcus* genera were more representative of the carcinoma group. No drastic decreasing or increasing trends were detected along the progressive severity of the disease, but rather a variable fluctuation of *taxa* abundances (Figure 11). According to

the literature, members of *Clostridium*, *Roseburia*, *Faecalibacterium*, *Blautia*, and *Eubacterium* have been attributed to healthy *status* and thus reduced in diseased states, whereas *Enterococcus*, *Streptococcus*, *Helicobacter*, *Bacteroides*, *Porphyromonas*, *Parvimonas*, *Prevotella*, *Alistipes*, and *Thermanaerovibri*o, and *Fusobacterium* have been pronounced in CRC/adenomas[209,334,363,377,378]. Although *Clostridium* cluster XIVa (*e.g., Roseburia intestinalis*, *Eubacterium rectale)* was not differentially abundant in both clinical groups in this study, it has been most prevalent in healthy cases, serving as major sources of butyrate in intimate contact with the mucin layer[379]. However, certain *Clostridium* spp. such as *Clostridium difficile* and *Clostridium perfringens* have been considered to be opportunistic pathogens associated with CRC initiation and secondary bacteremia, contributing to poor prognosis of the patient[377,380]. On the other hand, *Faecalibacterium prausnitzii,* acquired from the stool samples of healthy donors, has been proposed as a probiotic vital to human health due to its butyrate producing capacity, absence of virulence factors, and anti-inflammatory properties[381]. *Acidaminococcus (e.g., Acidaminococcus fermentans),* a Gram-negative, anaerobic gastrointestinal tract residing genus, has been characterized for its glutamate fermentation and restricted use of amino acids for its metabolism[382]. As protein monomers, amino acids are essential for normal cellular activity in times of health and even more demanding for continuous cell survival in cancer cells[383]. Since high-protein/high fat diets have been linked to colorectal carcinogenesis and *Acidaminococcus* favor the building blocks of protein synthesis, it is not surprising that Andoh *et al.*[384] found a significant increase in *Acidaminococcus intestine*, in contrast to *Clostridium* spp., *Faecalibacterium* sp., and *Eubacterium* sp., in the stools of obese individuals when evaluated against leaner individuals.

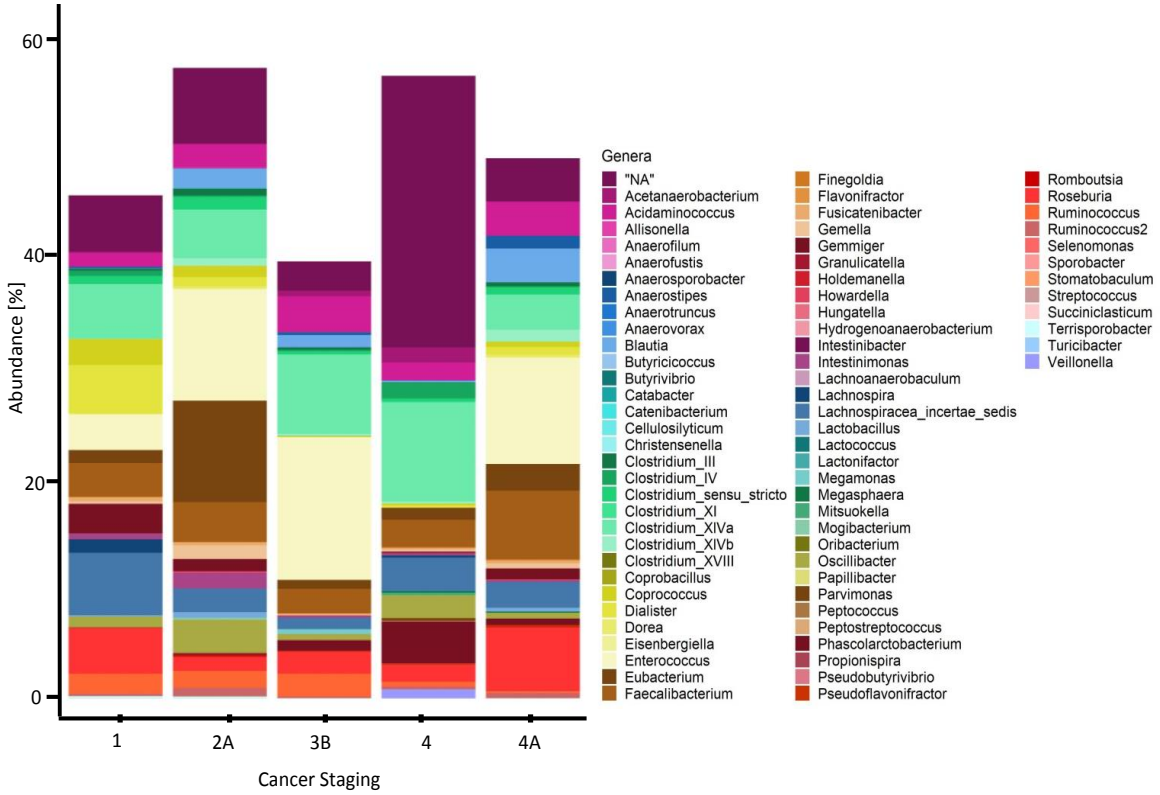

Figure 11: Genera-level (within the Firmicutes phylum) relative abundance of bacteria from stool samples CRC-diagnosed patients, according to the diagnosed cancer stage.

Genera belonging to the Bacteroidetes phylum demonstrated consistent relative abundances (*e.g., Bacteroides, Alistipes, Parabacteroides*) between both clinical groups, except an interesting higher abundance of *Prevotella* or the exclusive occurrence of *Alloprevotella* in the CRC group (Figure 12). Much like Firmicutes, a highly variable abundance was witnessed along the early and advanced colorectal stages, although *Prevotella* was more proliferated in early stage CRC (Figure 13). Although members of the *Prevotella* genus (*e.g., Prevotella copri*) are known SCFA producers, they have been considered to be pathobionts implicated in chronic IBD[385], Irritable Bowel Syndrome (IBS)-Diarrhea[386], and CRC[78], mainly driven through pro-inflammatory characteristics and disruption of the mucus gel layer by genes encoding mucin-desulfating sulfatase for increased permeability to the mucosa[387]. *Alloprevotella*, amongst other oral bacteria mentioned, has been associated with periodontal disease[388], but has yet to be fully characterized in CRC studies since the oral microbiota translocation route hypothesis has been implicated[179].

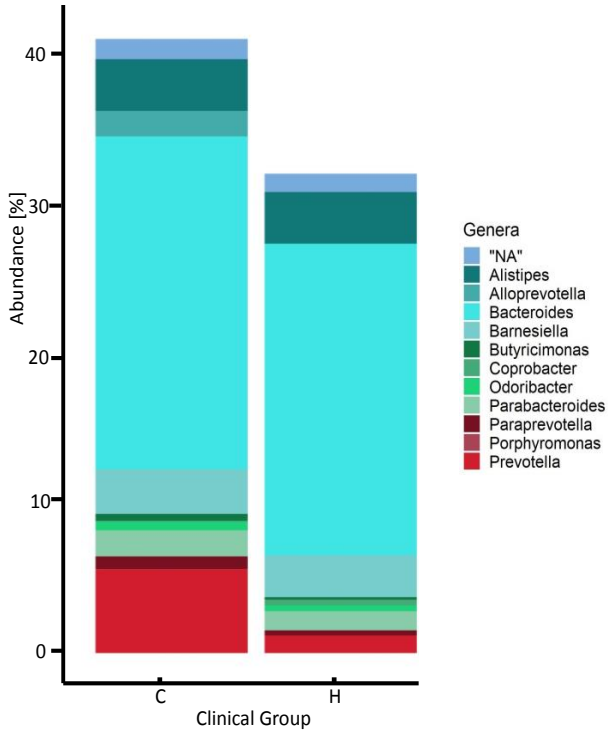

Figure 12: Genera-level (Bacteroidetes phylum) relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

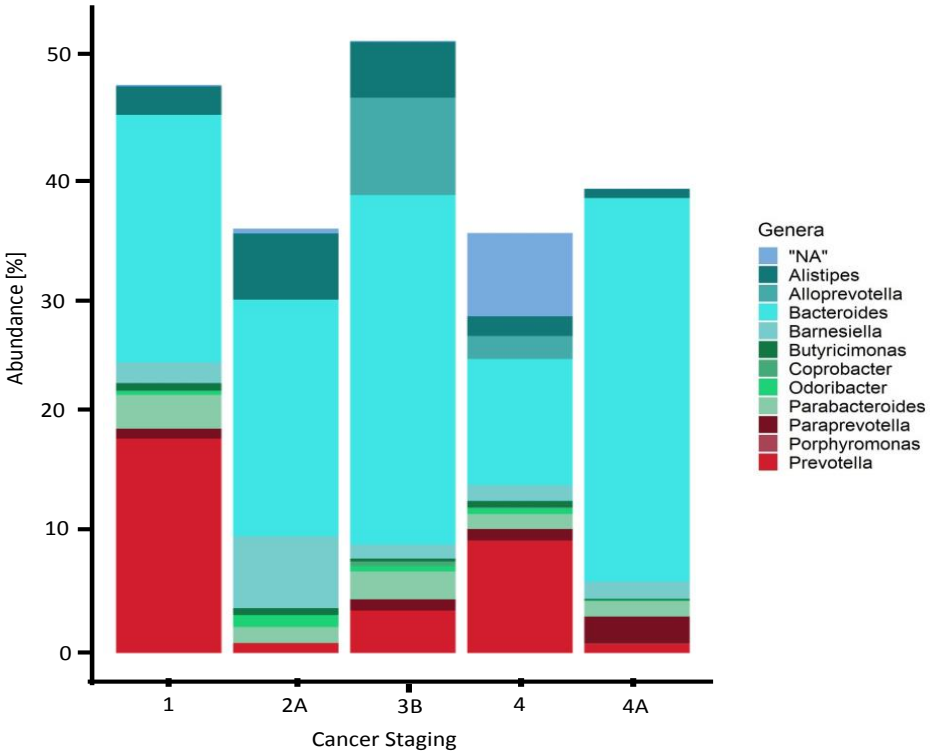

Figure 13: Genera-level (Bacteroidetes phylum) relative abundance of bacteria from stool samples CRC-diagnosed (C) individuals, according to diagnosed cancer stage.

Proteobacteria like *Desulfovibrio*, *Succinivibrio*, and *Sutterella* were differentially abundant in CRC *versus* healthy individuals, whereas *Parasutterella* and *Proteus* displayed greater abundances in the healthy group (Figure 14). *Morganella* showed no major abundance differences in both clinical groups (Figure 14), although decreased in late CRC stages (4 and 4A) (Figure 15). In contrast, *Sutterella* and *Desulfovibrio* abundances increased with advancing CRC stage (Figure 15). Despite not being major players in gut

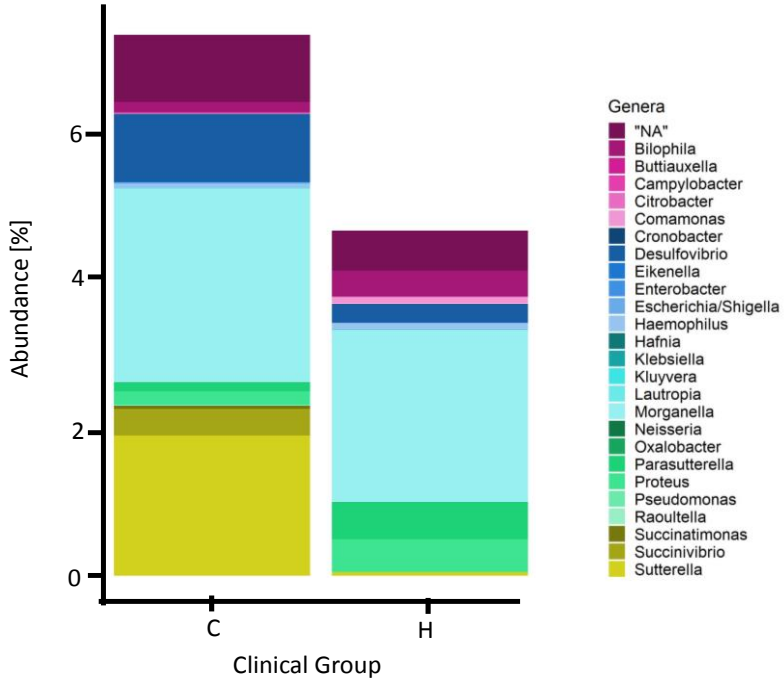

Figure 14: Genera-level (Proteobacteria phylum) relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

diseases due to their commensal relationship with the host, *Morganella* and *Proteus* have been considered to be opportunistic pathogens at inflammatory and immunosuppressive states presented in IBD and in early CRC development[389–391]. *P. mirabilis* has been characterized to possess genes encoding adhesion factors, flagellins for motility and invasion, hemolysins and metalloproteases destroying innate immune constituents, and lipopolysaccharide endotoxins, all of which are virulence agents that promote disease at the opportune conditions[391]. While *Desulfovibrio*, could be distinguished amongst both clinical groups, Balamurugan *et al.*[392] demonstrated no significant difference in the abundance of this genus in the stool samples of CRC patients compared to healthy donors

and individuals manifesting esophageal or gastric cancer. Nevertheless, the stimulatory growth of *Desulfovibrio* through a high fat diet has posed genotoxic effects to colonocytes through high hydrogen sulfide byproduct concentrations early on in the formation of adenomatous polyps[215], demonstrating the predictive value of *Desulfovibrio* in adenoma-CRC progression. Additionally, this genus has been indicative of colitis associated with the impairment of colon epithelia and long-term inflammation[393], which like the initial stages of adenoma, can indirectly contribute to CRC development. Likewise, *Sutterella* has been shown to be increased in the fecal samples of adenoma-bearing individuals[215] with slight pro-inflammatory action and moderate species-specific adhesion to intestinal cells[394]. However, the mechanisms by which *Sutterella* sp. and the host interact in CRC, especially in advancing CRC stages are still unclear due to conflicting results in CRC stage-specific affiliation[207,364]. Perhaps, *Sutterella*, amongst other genera, may act as a "passenger" genus that proliferates along with CRC carcinogenesis in order to maintain tumor metabolic functioning requirements, as hypothesized[364].

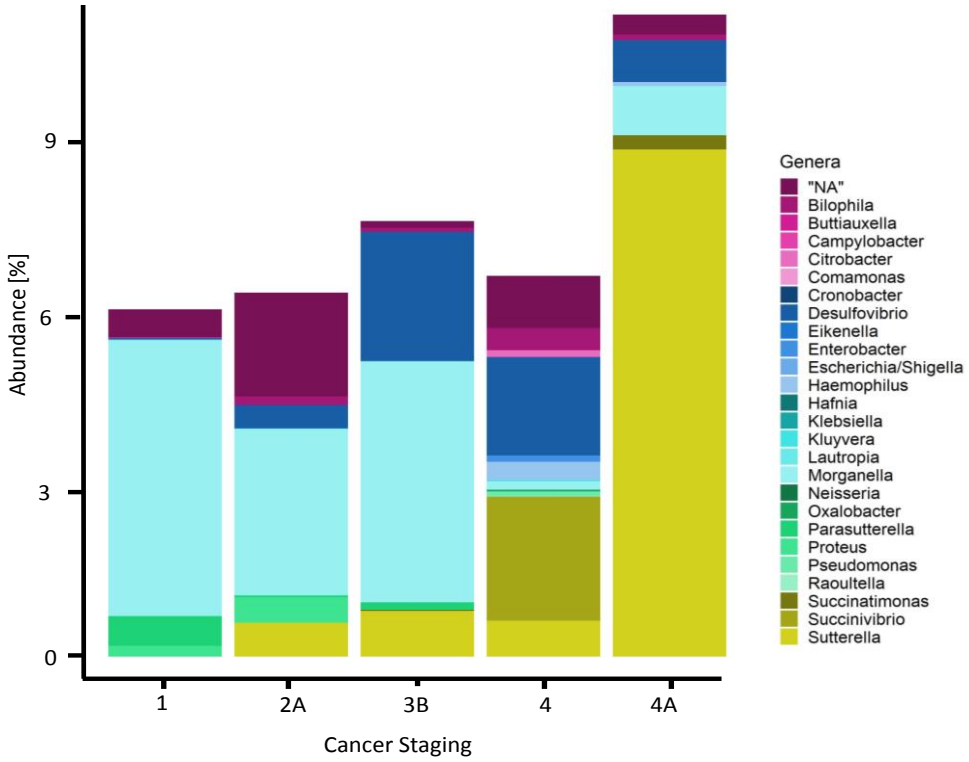

Figure 15: Genera-level (Proteobacteria phylum) relative abundance of bacteria from stool samples CRC-diagnosed (C) individuals, according to diagnosed cancer stage.

Although *Bifidobacterium*, belonging to the Actinobacteria phylum, was the most dominant genus in healthy individuals (despite its low abundance amongst the samples; Figure 16), it did not demonstrate a sequential significant decrease in CRC stages (Figure 17). *Bifidobacterium* has been well established in the literature for its protective effects on the colonic mucosa[372,395,396]. This genus has demonstrated the ability to enhance colonic barrier protective functioning in Caco-2 cells, by strengthening intestinal impenetrability via decreased pro-inflammatory cytokine and zonulin toxin (intestinal permeability regulator) levels[397]. Moreover, butanol extracted from *Bifidobacterium adolescentis* was shown to reduce colon cancer cell line (Caco-2, HT-29, and SW480) proliferation, increase macrophage stimulation, and cancer-specific apoptotic activity[395]. In contrast to *Bifidobacterium*, *Olsenella,* while not highly abundant, showed a marked representation in later CRC stage (4A), intriguingly. *Olsenella* species (*e.g.*, *Osenella uli*, *O. profusa*) are sugar-fermenting and aerotolerant inhabitants of the oral cavity, associated with periodontal and endodontic disease[398] that have had a negative correlation with obesity[384], a CRC risk factor. However, they have had limited impact in CRC microbiome studies for decisive conclusions on its involvement in colorectal pathogenesis.

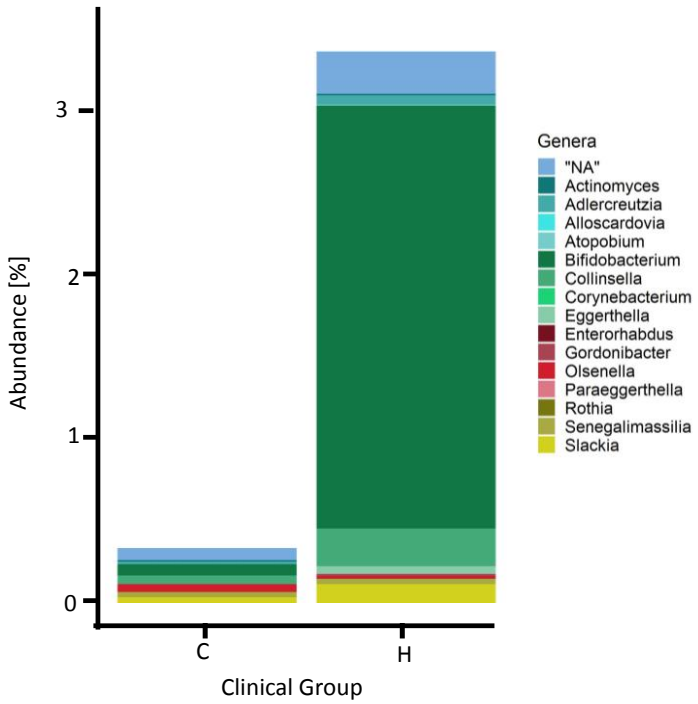

Figure 16: Genera-level (Actinobacteria phylum) relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

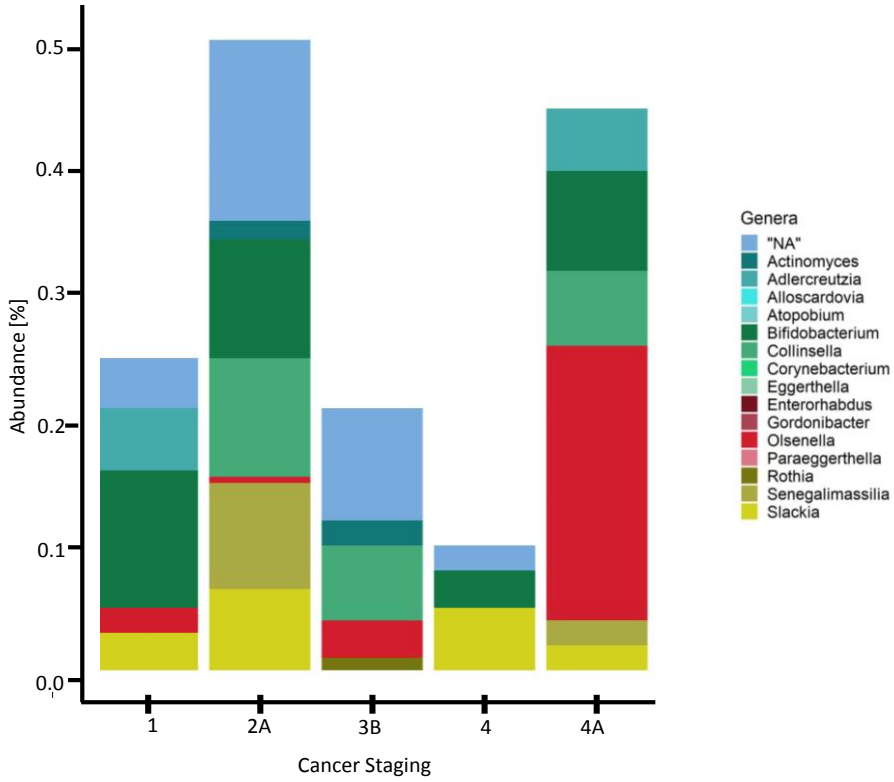

Figure 17: Genera-level (Actinobacteria phylum) relative abundance of bacteria from stool samples CRC-diagnosed (C) individuals, according to diagnosed cancer stage.

Despite its over-representative trend in, not only IBD, but also in numerous early colorectal adenoma and late carcinoma samples[279,280,288], *Fusobacterium* abundance, especially in CRC cases, is quite lower than expected (Figure 18), which could be explained by the limited sample size and heterogenic nature of the stool samples considered in this study. *Fusobacterium* species flourish as invasive pro-inflammatory drivers via the stimulation of inflammatory cytokines (TNF-α), and easily adhere to the intestinal mucosa for facilitated accessibility[279,280]. For instance, *fadA* gene expression, a *Fusobacterium* virulent adhesion factor shown to block anti-tumor immune responses and enhance colorectal oncogenesis signaling, as well as to induce pro-inflammatory responses modulated by the E-Cadherin/ β-Catenin and NF-κB pathways, has been identified and overexpressed in colorectal adenoma and adenocarcinomas [281,399].

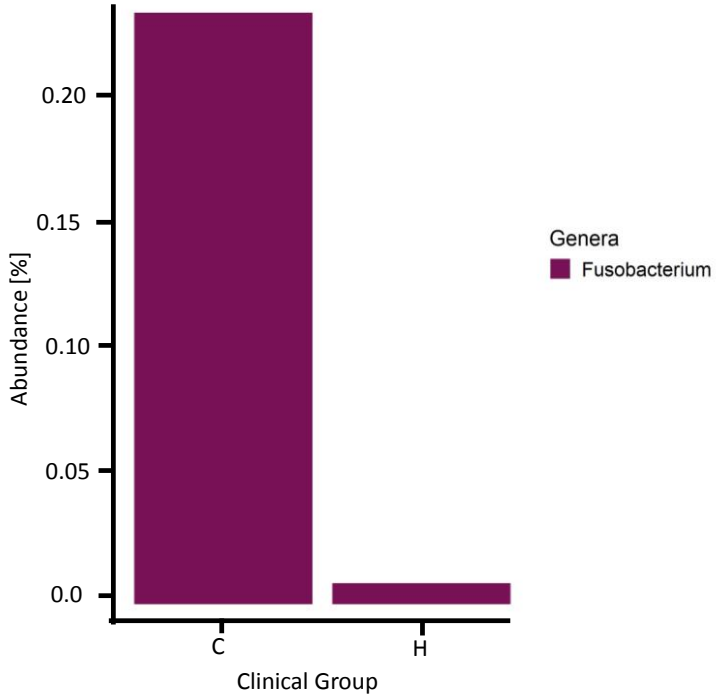

Figure 18: Genera-level (Fusobacteria phylum) relative abundance of bacteria from stool samples of CRC-diagnosed (C) patients and healthy (H) individuals.

4. Conclusion

CRC-associated microbiome showed a loss in alpha bacterial community diversity, and beta diversity demonstrated a clear separation between both clinical groups even at early and late CRC stages, although statistical insignificance was observed for the latter. Some relative differential abundances with respect to phylum, family, and genus were observed to be distinctive to CRC and healthy *status*, although most were homogenous amongst both groups. Higher abundances of *Prevotella* and the exclusive occurrence of *Alloprevotella* have been demonstrated in the CRC group, whereas *Sutterella*, *Desulfovibrio,* and *Olsenella* genera were more representative with advancing CRC stage, thereby potentially serving as microbial based biomarkers for CRC. The variability detected between both clinical groups should be further considered in order to identify significant bacterial species that specifically indicate an accurate CRC disease diagnosis. Notwithstanding, more concrete bacterial associations to CRC *versus* healthy states are yet to be deeply resolved through the application of larger sample sizes.

CHAPTER IV

Final Conclusions

CHAPTER IV: FINAL CONCLUSIONS

Fortunately, CRC is one of the most preventative cancers through diet and lifestyle changes and given the availability of early diagnostic technology, especially because this cancer has a large transition bracket time between benign to malignant[88,148,400]. Therefore, it is not only crucial to prevent the disease by reducing patient behavioral risk commonly implicated with cancer, but also to detect early-on potential benign/malignant tumors that can be surgically resected and increase patients' chances of survival. Even though preventative behaviors are available to decrease CRC risk, they may enclose social stigmas regarding the changes one has to make, besides some uncomfortable diagnostic procedures (*e.g.*, colonoscopy) that patients have to undergo, which have to be addressed and improved in future scientific work[108,146].

With the microbial dysbiosis-CRC link established amongst numerous "omics"-based studies, microbial fecal biomarkers have been of great scientific interest as complementary alternatives to invasive testing for primary screening and early CRC detection. However, the specificity and consistency of most biomarkers have been troublesome, since microbiome dysbiosis has been attributed to other diseases (*e.g.*, IBD), and certain candidate microbial biomarkers have been attributed to other cancers (*e.g.*, *Fusobacterium nucleatum* wide specificity to Gal-GalNAc—displaying tumors such as lung, pancreas adenocarcinomas[401]). Yet, this limitation has to be further improved through the use of more diverse and larger population cohorts, and validation of biomarker reliability, sensitivity, and accurateness for their worldwide clinical application. Matched mucosa-fecal samples for microbial biomarkers could also aid in biomarker validation, since at times fecal samples do not fully characterize nor possess all possible markers for specific CRC identification. A global effort in establishing evident, confident microbial associations to CRC *status* and the mechanisms by which these signatures promote disease have to be considered and targeted to increase overall successful and confident prediction rates. Perhaps, adjuvant/complementary testing using selective multiple biomarkers (*e.g.*, microbial, tumoral) at different molecular levels (*e.g.*, genes, transcripts, proteins, metabolites) using integrative "omics"-based technologies (*e.g.*, genomics, transcriptomics, proteomics, metabolomics) in a single simple test would be

more suggestive of CRC and/or its precursor early stages, as far as it would also more closely address the heterogeneity of this complex disease.

Among the Portuguese population, the microbiome structure was distinguishable between the stool samples of healthy and CRC-affected individuals, the latter being less diverse. Although relative abundance homogeneity at a taxonomic level was witnessed for both clinical groups for the most part, members of Bacteriodetes (*Prevotella, Alloprevotella*), Proteobacteria (*Sutterella*, *Desulfovibrio*), and Actinobacteria (*Olsenella*), could be viewed as microbial indicators for CRC and its progression, scientifically contributing to the global effort for the development of microbial based screening and diagnostic options to be explored clinically in the future.

In conclusion, comparing results collected from multiple sample origins (*e.g.*, stool, mucosa, blood) from patients of diverse ethnic backgrounds, diets, and lifestyles with colorectal neoplasia and malignant lesions could be the unveiling answer in obtaining a well-rounded determination of the microbial biomarkers targeted and sensitive to CRC detection and prevention.

REFERENCES

REFERENCES

1. Li, F. & Lai, M. Colorectal cancer, one entity or three. *J. Zhejiang Univ. Sci. B* **10,** 219–229 (2009).
2. Brenner, H. *et al.* Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840 149 screening colonoscopies. *Gut* **56,** 1585–1589 (2007).
3. American Cancer Society. *Colorectal Cancer Facts and Figures 2017-2019*. (2017).
4. Stewart, S. L., Wike, J. M., Kato, I., Lewis, D. R. & Michaud, F. A Population-Based Study of Colorectal Cancer Histology in the United States, 1998-2001. *Cancer* **107,** 1128–1141 (2006).
5. Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol* **3,** 153–173 (2012).
6. Ibrahim, K. O., Anjorin, A. S., Afolayan, A. E. & Badmos, K. B. Morphology of colorectal carcinoma among Nigerians : A 30-year review. *Niger. J. Clin. Pract.* **14,** 14–17 (2011).
7. Hamilton, S. R. & Aaltonen, L. A. in *World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of the Digestive System.* 105–136 (IARCPress, 2000).
8. Burke, H. B. Outcome Prediction and the Future of the TNM Staging System. *J. Natl. Cancer Inst.* **96,** 10–11 (2004).
9. Young, J. L., Roffers, S. D., Ries, L. A. G., Fritz, A. G. & Hurlbut, A. A. SEER SUMMARY STAGING MANUAL - 2000 Codes and Coding Instructions. 1–293 (2001).
10. Gao, R., Gao, Z., Huang, L. & Qin, H. Gut microbiota and colorectal cancer. *Eur. J. Clin. Microbiol. Infect. Dis.* **36,** 757–769 (2017).
11. Sun, T. *et al.* Evolutionary biologic changes of gut microbiota in an ‘ adenoma – carcinoma sequence ’ mouse colorectal cancer model induced by 1, 2-Dimethylhydrazine. *Oncotarget* **8,** 444–457 (2017).
12. Britzen-laurent, N. *et al.* GBP-1 acts as a tumor suppressor in colorectal cancer cells. *Carcinogenesis* **34,** 153–162 (2013).
13. Luo, Y. *et al.* RET is a potential tumor suppressor gene in colorectal cancer. *Oncogene* **32,** 2037–2047 (2013).

14. Dai, Y. *et al.* TCF21 functions as a tumor suppressor in colorectal cancer through inactivation of PI3K / AKT signaling. *Onco. Targets. Ther.* **10,** 1603–1611 (2017).
15. Qi, L. & Ding, Y. Screening of Tumor Suppressor Genes in Metastatic Colorectal Cancer. *Biomed Res. Int.* **2017,** 2769140 (2017).
16. Velho, S. *et al.* BRAF, KRAS and PIK3CA mutations in colorectal serrated polyps and cancer: Primary or secondary genetic events in colorectal carcinogenesis? *BMC Cancer* **8,** 225 (2008).
17. Minarikova, P. *et al.* Longitudinal molecular characterization of endoscopic specimens from colorectal lesions. *World J. Gastroenterol.* **22,** 4936–4945 (2016).
18. Oke, S. & Martin, A. Insights into the role of the intestinal microbiota in colon cancer. *Therap. Adv. Gastroenterol.* **10,** 417–428 (2017).
19. Mármol, I. *et al.* Colorectal Carcinoma : A General Overview and Future Perspectives in Colorectal Cancer. *Int. J. Mol. Sci.* **18,** 197 (2017).
20. Lao, V. V. & Grady, W. M. Epigenetics and Colorectal Cancer. *Nat Rev Gastroenterol Hepatol.* **8,** 686–700 (2012).
21. Ormanns, S., Neumann, J., Horst, D., Kirchner, T. & Jung, A. WNT signaling and distant metastasis in colon cancer through transcriptional activity of nuclear β-Catenin depend on active PI3K signaling. *Oncotarget* **5,** 2999–3011 (2014).
22. Calon, A. *et al.* Dependency of colorectal cancer on a TGF-beta-driven programme in stromal cells for metastasis initiation. *Cancer Cell* **22,** 571–584 (2012).
23. Chen, W., Swanson, B. J. & Frankel, W. L. Molecular genetics of microsatellite-unstable colorectal cancer for pathologists. *Diagn. Pathol.* **12,** 24 (2017).
24. Kawakami, H., Zaanan, A. & Sinicrope, F. A. MSI testing and its role in the management of colorectal cancer. *Curr Treat Options* **16,** 30 (2016).
25. Kaemmerer, E., Klaus, C., Jeon, M. K. & Gassler, N. Molecular classification of colorectal carcinomas : The genotype-to-phenotype relation. *World J. Gastroenterol.* **19,** 8163–8167 (2013).
26. Lind, G. E. *et al.* A CpG island hypermethylation profile of primary colorectal carcinomas and colon cancer cell lines. *Mol. Cancer* **3,** 28 (2004).
27. Bray, F. *et al.* Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and

Mortality Worldwide for 36 Cancers in 185 Countries. *CA. Cancer J. Clin.* **68,** 394–424 (2018).

28. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66,** 683–691 (2017).
29. Kang, M. & Martin, A. Microbiome and colorectal cancer : Unraveling host-microbiota interactions in colitis-associated colorectal cancer development. *Semin. Immunol.* **32,** 3–13 (2017).
30. Gal, D. L., Santos, A. & Barros, H. Leisure-time versus full-day energy expenditure: a cross-sectional study of sedentarism in a Portuguese urban population. *BMC Public Health* **5,** 16 (2005).
31. Rodrigues, S., Caraher, M., Trichopoulou, A. & de Almeida, M. Portuguese households' diet quality (adherence to Mediterranean food pattern and compliance with WHO population dietary goals): trends , regional disparities and socioeconomic determinants. *Eur. J. Clin. Nutr.* **62,** 1263–1272 (2008).
32. Miranda, N. *et al. Portugal: Doenças Oncológicas em Números – 2015*. *Direção-Geral da Saúde* (2015).
33. Cotter, J. Colorectal Cancer: Portugal and the World. *Acta Med. Port.* **26,** 485–486 (2013).
34. Bohorquez, M. *et al.* Clinical manifestations of colorectal cancer patients from a large multicenter study in Colombia. *Medicine (Baltimore).* **95,** e4883 (2016).
35. Diloreto, R., Murphy, C. T. & Bement, W. The cell biology of aging. *Mol. Biol. Cell* **25,** 4524–4531 (2015).
36. Amersi, F., Agustin, M. & Ko, C. Y. Colorectal Cancer: Epidemiology, Risk Factors, and Health Services. *Clin Colon Rectal Surg.* **18,** 133–140 (2005).
37. Ollberding, N. J., Nomura, A. M. Y., Wilkens, L. R., Henderson, B. E. & Kolonel, L. N. Racial/Ethnical Differences in Colorectal Cancer Risk: The Multiethnic Cohort Study. *Int J Cancer* **129,** 1899–1906 (2012).
38. Valle, L. Genetic predisposition to colorectal cancer: Where we stand and future perspectives. *World J. Gastroenterol.* **20,** 9828–9849 (2014).
39. Rodrigues, R. V. *et al.* Colorectal cancer surveillance in Portuguese families with

lynch syndrome: a cohort study. *Int. J. Colorectal Dis.* **33,** 695–702 (2018).

40. Vulcan, A., Manjer, J., Ericson, U. & Ohlsson, B. Intake of different types of red meat, poultry, and fish and incident colorectal cancer in women and men: results from the Malmö Diet and Cancer Study. *Food Nutr. Res.* **61,** 1341810 (2017).
41. Bastide, N. M., Pierre, F. H. F. & Corpet, D. E. Heme Iron from Meat and Risk of Colorectal Cancer: A Meta-analysis and a Review of the Mechanisms Involved. *Cancer Prev. Res.* **4,** 177–185 (2011).
42. Zhu, B., Sun, Y., Qi, L., Zhong, R. & Miao, X. Dietary legume consumption reduces risk of colorectal cancer: evidence from a meta-analysis of cohort studies. *Sci. Rep.* **5,** 8797 (2015).
43. Fedirko, V. *et al.* Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann. Oncol.* **22,** 1958–1972 (2011).
44. Wang, Y., Duan, H., Yang, H. & Lin, J. A pooled analysis of alcohol intake and colorectal cancer. *Int J Clin Exp Med* **8,** 6878–6889 (2015).
45. Song, Y. K. *et al.* Alcohol Drinking Increased the Risk of Advanced Colorectal Adenomas. *Intest. Res.* **13,** 74–79 (2015).
46. Limsui, D. *et al.* Cigarette Smoking and Colorectal Cancer Risk by Molecularly Defined Subtypes. *J Natl Cancer Inst.* **102,** 1012–1022 (2010).
47. Stone, T. W. & Darlington, L. G. Microbial carcinogenic toxins and dietary anti-cancer protectants. *Cell. Mol. Life Sci.* **74,** 2627–2643 (2017).
48. Claus, S. P., Guillou, H. & Ellero-Simatos, S. The gut microbiota: a major player in the toxicity of environmental pollutants? *npj Biofilms Microbiomes* **2,** 16003 (2016).
49. Zhang, Y. *et al.* Impacts of Gut Bacteria on Human Health and Diseases. *Int. J. Mol. Sci.* **16,** 7493–7519 (2015).
50. Mcfarland, L. V. Normal flora: diversity and functions. *Microb. Ecol. Health Dis.* **12,** 193–207 (2000).
51. Davis, C. P. in *Medical Microbiology* (ed. Baron, S.) (1996). at <https://www.ncbi.nlm.nih.gov/books/NBK7617/>
52. Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* **65,** 330–339 (2016).

53. Quigley, E. M. M. Gut Bacteria in Health and Disease. *Gastroenterol. Hepatol.* **9,** 560–569 (2013).
54. Conlon, M. A. & Bird, A. R. The Impact of Diet and Lifestyle on Gut Microbiota and Human Health. *Nutrients* **7,** 17–44 (2015).
55. Swidsinski, A. *et al.* Comparative study of the intestinal mucus barrier in normal and inflamed colon. *Gut* **56,** 343–351 (2007).
56. Johansson, M. E. V *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *PNAS* **105,** 15064–15069 (2008).
57. Albenberg, L. *et al.* Correlation Between Intraluminal Oxygen Gradient and Radial Partitioning of Intestinal Microbiota in Humans and Mice. *Gastroenterology* **147,** 1055–1063 (2015).
58. Flint, H. J., Scott, K. P., Louis, P. & Duncan, S. H. The role of the gut microbiota in nutrition and health. *Nat. Rev. Gastroenterol. Hepatol.* **9,** 577–589 (2012).
59. Belzer, C. & Vos, W. M. De. Microbes inside — from diversity to function: the case of *Akkermansia*. *ISME J.* **6,** 1449–1458 (2012).
60. Zoetendal, E. G. *et al.* Mucosa-Associated Bacteria in the Human Gastrointestinal Tract Are Uniformly Distributed along the Colon and Differ from the Community Recovered from Feces. *Appl. Environ. Microbiol.* **68,** 3401–3407 (2002).
61. Zhang, Z. *et al.* Spatial heterogeneity and co-occurrence patterns of human mucosal-associated intestinal microbiota. *ISME J.* **8,** 881–893 (2014).
62. The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486,** 207–214 (2012).
63. Chow, J., Lee, S. M., Shen, Y., Khosravi, A. & Mazmanian, S. K. Host – Bacterial Symbiosis in Health and Disease. *Adv Immunol.* **107,** 243–274 (2011).
64. Ruiz, L. *et al.* How do bifidobacteria counteract environmental challenges? Mechanisms involved and physiological consequences. *Genes Nutr* **6,** 307–318 (2011).
65. Lloyd-Price, J., Abu-Ali, G. & Huttenhower, C. The healthy human microbiome. *Genome Med.* **8,** 51 (2016).
66. Peterson, J. *et al.* The NIH Human Microbiome Project. *Genome Res.* **19,** 2317–

2323 (2009).

67. Goodrich, J. K. *et al.* Human Genetics Shape the Gut Microbiome. *Cell* **159,** 789–799 (2014).

68. Gao, Z., Guo, B., Gao, R., Zhu, Q. & Qin, H. Microbiota disbiosis is associated with colorectal cancer. *Front. Microbiol.* **6,** 20 (2015).

69. Kostic, A. D. *et al.* Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma. *Genome Res.* **22,** 292–298 (2012).

70. Han, Y. *Fusobacterium nucleatum*: a commensal-turned pathogen. *Curr. Opin. Microbiol.* **23,** 141–147 (2016).

71. Toprak, N. U. *et al.* A possible role of *Bacteroides fragilis* enterotoxin in the aetiology of colorectal cancer. *Clin. Microbiol. Infect.* **12,** 782–786 (2006).

72. Coleman, O. I. & Nunes, T. Role of the Microbiota in Colorectal Cancer: Updates on Microbial Associations and Therapeutic Implications. *BioResearch Open Acess* **5,** 279–288 (2016).

73. Mehrabani-Khasraghi, S., Ameli, M. & Khalily, F. Demonstration of Herpes Simplex Virus, Cytomegalovirus, and Epstein-Barr Virus in Colorectal Cancer. *Iran. Biomed. J.* **20,** 302–306 (2016).

74. Picanço-Junior, O. M. *et al.* Association between Human Papillomavirus and Colorectal Adenocarinoma and its Influence on Tumor Staging and Degree of Cell Differentiation. *Arq. Bras. Cir. Dig.* **27,** 172–176 (2014).

75. Mahmoudvand, S., Safaei, A., Erfani, N. & Sarvari, J. Presence of Human Papillomavirus DNA in Colorectal Cancer Tissues in Shiraz , Southwest Iran. *Asian Pacific J. Cancer Prev.* **16,** 7883–7887 (2015).

76. Taherian, H., Tafvizi, F., Fard, Z. T. & Abdirad, A. Lack of association between human papillomavirus infection and colorectal cancer. *Przegląd Gastroenterol.* **9,** 280–284 (2014).

77. Cummins, J. & Tangney, M. Bacteria and tumours: causative agents or opportunistic inhabitants? *Infect. Agent. Cancer* **8,** 1–8 (2013).

78. Flemer, B. *et al.* Tumour-associated and non-tumour-associated microbiota in colorectal cancer. *Gut microbiota* **66,** 633–643 (2017).

79. Liu, L., Zhu, K., Yu, W. & Zhang, S. Detection Rate, Anatomic Sites, and Pathologic Types of Colorectal Cancer During Colonoscopy Procedures. *Surg. Laparosc. Endosc. Percutaneous Tech.* **5,** 394–399 (2017).

80. Dejea, C. M. *et al.* Microbiota organization is a distinct feature of proximal colorectal cancers. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 18321–18326 (2014).

81. Johnson, C. H. *et al.* Metabolism Links Bacterial Biofilms and Colon Carcinogenesis. *Cell Metab.* **21,** 891–897 (2015).

82. Rothschild, B. M., Witzke, B. J. & Hershkovitz, I. Metastatic cancer in the Jurassic. *Lancet* **354,** 398 (1999).

83. Prates, C., Sousa, S., Oliveira, C. & Ikram, S. Prostate metastatic bone cancer in an Egyptian Ptolemaic mummy , a proposed radiological diagnosis. *Int. J. Paleopathol.* **1,** 98–103 (2011).

84. Baxter, N. T., Koumpouras, C. C., Rogers, M. A. M., Iv, M. T. R. & Schloss, P. D. DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. *Microbiome* **4,** 59 (2016).

85. Rollandi, G. A., Biscaldi, E. & Decicco, E. Double contrast barium enema: Technique, indications, results and limitations of a conventional imaging methodology in the MDCT virtual endoscopy era. *Eur. J. Radiol.* **61,** 382–387 (2007).

86. Winawer, S. J. The History of Colorectal Cancer Screening: A Personal Perspective. *Dig. Dis. Sci.* **60,** 596–608 (2015).

87. Wu, C. & Sung, J. J. Colorectal cancer screening: are stool and blood based tests good enough? *Chinese Clin. Oncol.* **2,** 8 (2013).

88. Pickhardt, P. J. Emerging stool-based and blood-based non- invasive DNA tests for colorectal cancer screening: the importance of cancer prevention in addition to cancer detection. *Abdom. Radiol.* **41,** 1441–1444 (2016).

89. Scalise, P., Mantarro, A., Pancrazi, F. & Neri, E. Computed tomography colonography for the practicing radiologist: A review of current recommendations on methodology and clinical indications. *World J. Radiol.* **8,** 472–484 (2016).

90. Relman, D. A. Microbiology: Learning about who we are. *Nature* **486,** 195 (2012).

91. Wolff, W. I. & Shinya, H. Endoscopic Polypectomy. *Cancer* **36,** 683–690 (1975).

92. Wolff, W. I. & Shinya, H. Morphology , Anatomic Distribution and Cancer Potential of Colonic Polyps. *Ann. Surg.* **190,** 679–683 (1979).

93. Wolff, W. I. & Shinya, H. Definitive Treatment of " Malignant " Polyps of the Colon. *Ann. Surg.* **182,** 516–525 (1975).

94. Wolff, W. I. & Shinya, H. Polypectomy via the Fiberoptic Colonoscope. *N. Engl. J. Med.* **288,** 329–332 (1973).

95. Rex, D. K. *et al.* Quality indicators for colonoscopy. *Gastrointest. Endosc.* **63,** S16-28 (2006).

96. Loeve, A. J., Fockens, P. & Breedveld, P. Mechanical analysis of insertion problems and pain during colonoscopy: Why highly skill-dependent colonoscopy routines are necessary in the first place ... and how they may be avoided. *Can. J. Gastroenterol.* **27,** 293–302 (2013).

97. Rubin, M., Lurie, L., Bose, K. & Kim, S. H. Expanding the view of a standard colonoscope with the Third Eye Panoramic cap. *World J. Gastroenterol.* **21,** 10683–10687 (2015).

98. Dik, V. K., Moons, L. M. G. & Siersema, P. D. Endoscopic innovations to increase the adenoma detection rate during colonoscopy. *World J. Gastroenterol.* **20,** 2200–2211 (2014).

99. Gkolfakis, P. *et al.* New endoscopes and add-on devices to improve colonoscopy performance. *World J. Gastroenterol.* **23,** 3784–3796 (2017).

100. Wang, Y. R., Cangemi, R., Loftus, E. V & Picco, M. F. Decreased Risk of Colorectal Cancer after Colonoscopy in Patients 76 – 85 Years Old in the United States. *Digestion* **93,** 132–138 (2016).

101. Sanchez, W., Harewood, G. C., Sc, M. & Petersen, B. T. Evaluation of Polyp Detection in Relation to Procedure Time of Screening or Surveillance Colonoscopy. *Am. J. Gastroenterol.* **99,** 1941–1945 (2004).

102. Rijn, J. C. van *et al.* Polyp Miss Rate Determined by Tandem Colonoscopy: A Systematic Review. *Am. J. Gastroenterol.* **101,** 343–350 (2006).

103. Xirasagar, S. *et al.* Colorectal cancer prevention by an optimized colonoscopy protocol in routine practice. *Int. J. Cancer* **136,** 731–742 (2015).

104. Binefa, G., Rodríguez-Moranta, F., Teule, À. & Medina-Hayas, M. Colorectal cancer : From prevention to personalized medicine. *World J. Gastroenterol.* **20,** 6786–6808 (2014).
105. Day, L. W. & Velayos, F. Colorectal Cancer Screening and Surveillance in the Elderly: Updates and Controversies. *Gut Liver* **9,** 143–151 (2015).
106. Lebwohl, B. *et al.* The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. *Gastrointest. Endosc.* **73,** 1207–1214 (2011).
107. Pohl, H. *et al.* Incomplete Polyp Resection During Colonoscopy — Results of the Complete Adenoma Rescetion (CARE) Study. *Gastroenterology* **144,** 74–80 (2013).
108. Juchems, M. S., Ehmann, J., Brambs, H. & Aschoff, A. J. A Retrospective Evaluation of Patient Acceptance of Computed Tomography Colonography (“ Virtual Colonoscopy ”) in Comparison with Conventional Colonoscopy in an Average Risk Screening Population. *Acta radiol.* **7,** 664–670 (2005).
109. Gilbersten, V. A. Proctosigmoidoscopy and Polypectomy in Reducing the Incidence of Rectal Cancer. *Cancer* **34,** 936–939 (1974).
110. Heo, J., Ryu, C., Jung, E., Paik, J. & Hwang, D. Clinical Significance of Preoperative Virtual Colonoscopy for Evaluation of the Proximal Colon in Patient With Obstructive Colorectal Cancer. *Ann. Coloproctol.* **33,** 130–133 (2017).
111. Schoen, R. E. *et al.* Colorectal-Cancer Incidence and Mortality with Screening Flexible Sigmoidoscopy. *N. Engl. J. Med.* **366,** 2345–2357 (2012).
112. Gatto, N. M. *et al.* Risk of Perforation After Colonoscopy and Sigmoidoscopy: A Population-Based Study. *J. Natl. Cancer Inst.* **95,** 230–236 (2003).
113. Pullens, H. J. M., Leeuwen, M. S. van, Laheij, R. J. F., Vleggaar, F. P. & Siersema, P. D. CT-Colonography After Incomplete Colonoscopy: What is the Diagnostic Yield? *Dis. Colon Rectum* **56,** 593–599 (2013).
114. Stagnitti, A. *et al.* Preoperative staging of colorectal cancer using virtual colonoscopy: correlation with surgical results. *Eur. Rev. Med. Pharmacol. Sci.* **19,** 1645–1651 (2015).
115. Berger, A. Magnetic resonance imaging. *BMJ* **324,** 35 (2002).

116. Dam, C. *et al.* Local staging of sigmoid colon cancer using MRI. *Acta radiol.* **6,** 2058460117720957 (2017).
117. Nerad, E. *et al.* MRI for Local Staging of Colon Cancer: Can MRI Become the Optimal Staging Modality for Patients With Colon Cancer? *Dis. Colon Rectum* **60,** 385–392 (2017).
118. Halappa, V. G. *et al.* Rectal Imaging: Part I, High-Resolution MRI of Carcinoma of the Rectum at 3T. *AJR Am. J. Roentgenol.* **199,** W35–W42 (2013).
119. Gambhir, S. S. Molecular imaging of cancer with positron emission tomography. *Nat Rev Cancer* **2,** 683–693 (2002).
120. Kunawudhi, A., Wong, A. K., Alkasab, T. K. & Mahmood, U. Accuracy of FDG-PET / CT for Detection of Incidental Pre- Malignant and Malignant Colonic Lesions - Correlation with Colonoscopic and Histopathologic Findings. *Asian Pacific J. Cancer Prev.* **17,** 4143–4147 (2016).
121. Squillaci, E. *et al.* Staging of colon cancer: whole-body MRI vs . whole-body PET-CT — initial clinical experience. *Abdom. Imaging* **33,** 676–688 (2008).
122. Yu, X. *et al.* Role of 18 F-Fluorodeoxyglucose Positron Emission Tomography / Computed Tomography in the Diagnosis of Newly Found Suspected Malignant Solitary Pulmonary Lesions in Patients Who Have Received Curative Treatment for Colorectal Cancer. *Gastroenterol. Res. Pract.* **2017,** 3458739 (2017).
123. Martínez-Ares, D. *et al.* The value of abdominal ultrasound in the diagnosis of colon cancer. *Rev. Española Efermedades Dig.* **97,** 877–886 (2005).
124. Rafaelsen, S. R. & Jakobsen, A. Contrast-enhanced ultrasound vs multidetector-computed tomography for detecting liver metastases in colorectal cancer: a prospective, blinded, patient-by-patient analysis. *Color. Dis.* **13,** 420–425 (2011).
125. Ashraf, S. *et al.* A critical appraisal of endorectal ultrasound and transanal endoscopic microsurgery and decision-making in early rectal cancer. *Color. Dis.* **14,** 821–826 (2011).
126. Halligan, S. Commentary on S. Q. Ashraf et al. *Color. Dis.* **14,** 826–827 (2012).
127. Wagner, C. Von *et al.* Patient acceptability of CT colonography compared with double contrast barium enema: results from a multicentre randomised controlled

trial of symptomatic patients. *Eur. Soc. Radiol.* **21,** 2046–2055 (2011).

128. Gillespie, J. S. J. & Kelly, B. E. Double contrast barium enema and colorectal carcinoma: sensitivity and potential role in screening. *Ulster Med. J.* **70,** 15–18 (2001).
129. Tan, K. Y. *et al.* Which colorectal cancers are missed by double contrast barium enema? *Tech Coloproctol* **8,** 169–172 (2004).
130. Toma, J., Paszat, L. F., Gunraj, N. & Rabeneck, L. Rates of New or Missed Colorectal Cancer After Barium Enema and Their Risk Factors: A Population-Based Study. *Am. J. Gastroenterol.* **103,** 3142–3148 (2008).
131. Willyard, C. Screening: Early alert. *Nature* **521,** S4-5 (2015).
132. Imperiale, T. F. *et al.* Multitarget Stool DNA Testing for Colorectal-Cancer Screening. *N. Engl. J. Med.* **370,** 1287–1297 (2014).
133. Hao, Y.-X. *et al.* Effectiveness of circulating tumor DNA for detection of KRAS gene mutations in colorectal cancer patients: a meta-analysis. *Onco. Targets. Ther.* **10,** 945–953 (2017).
134. Yan, S., Liu, Z., Yu, S. & Bao, Y. Diagnostic Value of Methylated Septin9 for Colorectal Cancer Screening: A Meta-Analysis. *Med. Sci. Monit.* **22,** 3409–3418 (2016).
135. Rasmussen, S. L. *et al.* Hypermethylated DNA, a circulating biomarker for colorectal cancer detection. *PLoS One* **12,** e018809 (2017).
136. Johnson, D. A. *et al.* Plasma Septin9 versus Fecal Immunochemical Testing for Colorectal Cancer Screening: A Prospective Multicenter Study. *PLoS One* **9,** e98238 (2014).
137. Kitano, S. *et al.* A Novel Fully Automated Molecular Diagnostic System (AMDS) for Colorectal Cancer Mutation Detection. *PLoS One* **8,** e62989 (2013).
138. Hornbrook, M. C. *et al.* Early Colorectal Cancer Detected by Machine Learning Model Using Gender, Age, and Complete Blood Count Data. *Dig. Dis. Sci.* **62,** 2719–2727 (2017).
139. Zhang, P. *et al.* Identification of differentially expressed circular RNAs in human colorectal cancer. *Tumor Biol.* **39,** 1010428317694546 (2017).

140. Li, X. L. *et al.* Long Noncoding RNA PURPL Suppresses Basal p53 Levels and Promotes Tumorigenicity in Colorectal Cancer. *Cell Rep.* **20,** 2408–2423 (2017).
141. Danese, E., Minicozzi, A. M., Benati, M., Paviati, E. & Gusella, M. Reference miRNAs for colorectal cancer: analysis and verification of current data. *Sci. Rep.* **7,** 8413 (2017).
142. Slattery, M. L. *et al.* Infrequently expressed miRNAs in colorectal cancer tissue and tumor molecular phenotype. *Mod. Pathol.* **30,** 1152–1169 (2017).
143. Niu, Y. *et al.* Identification of reference genes for circulating microRNA analysis in colorectal cancer. *Sci. Rep.* **6,** 35611 (2016).
144. Okugawa, Y. *et al.* Clinical significance of SNORA42 as an oncogene and a prognostic biomarker in colorectal cancer. *Gut* **66,** 107–117 (2015).
145. Zhang, S., Lin, M. & Zhang, H. Diagnostic value of carcinoembryonic antigen and carcinoma antigen 19-9 for colorectal carcinoma. *Int J Clin Exp Pathol* **8,** 9404–9409 (2015).
146. Thomas, D. S. *et al.* Evaluation of serum CEA , CYFRA21-1 and CA125 for the early detection of colorectal cancer using longitudinal preclinical samples. *Br. J. Cancer* **113,** 268–274 (2015).
147. Song, Y. F. *et al.* Serum Cyr61 as a potential biomarker for diagnosis of colorectal cancer. *Clin. Transl. Oncol.* **19,** 519–524 (2017).
148. Fung, K. Y. C. *et al.* Blood-Based Protein Biomarker Panel for the Detection of Colorectal Cancer. *PLoS One* **10,** 1–11 (2015).
149. Wilhelmsen, M. *et al.* Detection of colorectal neoplasia: Combination of eight blood-based, cancer-associated protein biomarkers. *Int. J. Cancer* **140,** 1436–1446 (2017).
150. Zhu, M. M. *et al.* Comparison of immunochemical and guaiac-based fecal occult blood test in screening and surveillance for advanced colorectal neoplasms: A meta-analysis. *J. Dig. Dis.* **11,** 148–160 (2010).
151. Hirai, H. W. *et al.* Systematic review with meta-analysis: faecal occult blood tests show lower colorectal cancer detection rates in the proximal colon in colonoscopy-verified diagnostic studies. *Aliment. Pharmacol. Ther.* **43,** 755–764 (2016).

152. Mousavinezhad, M., Majdzadeh, R., Sari, A. A., Delavari, A. & Mohtasham, F. The effectiveness of FOBT vs. FIT: A meta-analysis on colorectal cancer screening test. *Med. J. Islam. Repub. Iran* **30,** 366 (2016).
153. Scully, A. & Cheung, I. Colorectal Cancer Screening. *Work. Heal. Saf.* **64,** 114–122 (2016).
154. Akram, A. *et al.* Replacing the Guaiac Fecal Occult Blood Test With the Fecal Immunochemical Test Increases Proportion of Individuals Screened in a Large Healthcare Setting. *Clin. Gastroenterol. Hepatol.* **15,** 1265–1270.e1 (2017).
155. Robertson, D. J. *et al.* Recommendations on fecal immunochemical testing to screen for colorectal neoplasia: a consensus statement by the US Multi-Society Task Force on colorectal cancer. *Gastrointest. Endosc.* **85,** 2–21.e3 (2017).
156. Faivre, J. *et al.* Comparison between a guaiac and three immunochemical faecal occult blood tests in screening for colorectal cancer. *Eur. J. Cancer* **48,** 2969–2976 (2012).
157. Brenner, H. & Tao, S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. *Eur. J. Cancer* **49,** 3049–3054 (2013).
158. Brenner, H. & Chen, H. Fecal occult blood versus DNA testing: indirect comparison in a colorectal cancer screening population. *Clin. Epidemiol.* **9,** 377–384 (2017).
159. Jerzak, K. J. *et al.* Targeted metabolomics in colorectal cancer: a strategic approach using standardized laboratory tests of the blood and urine. *Hypoxia* **5,** 61–66 (2017).
160. Qiu, Y. *et al.* A Distinct Metabolic Signature of Human Colorectal Cancer with Prognostic Potential. *Clin. Cancer Res.* **20,** 2136–2147 (2014).
161. Guerrero-Flores, H. *et al.* A non-invasive tool for detecting cervical cancer odor by trained scent dogs. *BMC Cancer* **17,** 79 (2017).
162. Kitiyakara, T. *et al.* The detection of Hepatocellular carcinoma (HCC) from patients' breath using canine scent detection: Proof of concept study. *J. Breath Res.* **11,** 046002 (2017).

163. Sonoda, H. *et al.* Colorectal cancer screening with odour material by canine scent detection. *Gut* **60,** 814–819 (2011).

164. Hirotsu, T. *et al.* A Highly Accurate Inclusive Cancer Screening Test Using *Caenorhabditis elegans* Scent Detection. *PLoS One* **10,** e0118699 (2015).

165. Amal, H. *et al.* Breath testing as potential colorectal cancer screening tool. *Int. J. Cancer* **138,** 229–236 (2016).

166. Arasaradnam, R. P. *et al.* Detection of Colorectal Cancer (CRC) by Urinary Volatile Organic Compound Analysis. *PLoS One* **9,** e108750 (2014).

167. Zonta, G. *et al.* Use of Gas Sensors and FOBT for the Early Detection of Colorectal Cancer. *Proceedings* **1,** 398 (2017).

168. Casen, C. *et al.* Deviations in human gut microbiota: a novel diagnostic test for determining dysbiosis in patients with IBS or IBD. *Aliment. Pharmacol. Ther.* **42,** 71–83 (2015).

169. Weir, T. L. *et al.* Stool Microbiome and Metabolome Differences between Colorectal Cancer Patients and Healthy Adults. *PLoS One* **8,** e70803 (2013).

170. Wong, S. H. *et al.* Quantitation of faecal *Fusobacterium* improves faecal immunochemical test in detecting advanced colorectal neoplasia. *Gut* **66,** 1441–1448 (2016).

171. Liang, Q. *et al.* Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. *Clin. Cancer Res.* **23,** 2061–2071 (2017).

172. Zhang, S., Cao, X. & Huang, H. Sampling Strategies for Three-Dimensional Spatial Community Structures in IBD Microbiota Research. *Front. Cell. Infect. Microbiol.* **7,** 51 (2017).

173. Smith, A. P., Chiu, Y. S. Y. & Lee, N. M. Towards Universal Screening for Colon Cancer: A Cheap, Reliable, Noninvasive Test Using Gene Expression Analysis of Rectal Swabs. *ISRN Gastroenterol.* **2012,** 170210 (2012).

174. McPherson, A. S., Dhungyel, O. P. & Ward, M. P. Comparison of recto-anal mucosal swab and faecal culture for the detection of *Escherichia coli* O157 and identification of super-shedding in a mob of Merino sheep. *Epidemiol. Infect.* **143,** 2733–2742 (2015).

175. Agga, G. E., Arthur, T. M., Schmidt, J. W., Wang, R. & Brichta-Harhay, D. M. Diagnostic Accuracy of Rectoanal Mucosal Swab of Feedlot Cattle for Detection and Enumeration of *Salmonella enterica*. *J. Food Prot.* **79,** 531–537 (2016).
176. Baldwin, D. A., Feldman, M., Alwine, J. C. & Robertson, E. S. Metagenomic Assay for Identification of Microbial Pathogens in Tumor Tissues. *MBio* **5,** e01714-14 (2014).
177. Cummings, J. H., Hill, M. J., Bone, E. S., Branch, W. J. & Jenkins, D. J. A. The effect of meat protein and dietary fiber on colonic function and metabolism II. Bacterial metabolites in feces and urine. *Am. J. Clin. Nutr.* **32,** 2094–2101 (1979).
178. Qiu, Y. *et al.* Urinary Metabonomic Study on Colorectal Cancer. *J. Proteome Res.* **9,** 1627–1634 (2010).
179. Flemer, B. *et al.* The oral microbiota in colorectal cancer is distinctive and predictive. *Gut microbiota* **67,** 1454–1463 (2018).
180. Flynn, K. J., Baxter, N. T. & Schloss, P. D. Metabolic and Community Synergy of Oral Bacteria in Colorectal Cancer. *Msph. Host-Microbe Biol.* **1,** e00102-16 (2016).
181. Nambiar, P. R., Gupta, R. R. & Misra, V. An " Omics " based survey of human colon cancer. *Mutat. Res.* **693,** 3–18 (2010).
182. Henry, V. J., Bandrowski, A. E., Pepin, A.-S., Gonzalez, B. J. & Desfeux, A. OMICtools: an informative directory for multi-omic data analysis. *Database (Oxford)* **2014,** bau069 (2014).
183. Zhu, J., Shi, Z., Wang, J. & Zhang, B. Empowering biologists with multi-omics data: colorectal cancer as a paradigm. *Bioinformatics* **31,** 1436–1443 (2015).
184. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. *Genome Biol.* **18,** 83 (2017).
185. Ligthart, S. *et al.* DNA methylation signatures of chronic low-grade inflammation are associated with complex diseases. *Genome Biol.* **17,** 255 (2016).
186. Clarke, T.-K. *et al.* Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N = 112 117). *Mol. Psychiatry* **22,** 1376–1384 (2017).
187. Rokas, A., Gibbons, J. G., Zhou, X., Beauvais, A. & Latge, J.-P. The diverse applications of RNA-seq for functional genomic studies in Aspergillus fumigatus.

Ann. N. Y. Acad. Sci. **1273,** 25–34 (2013).

188. Geyer, P. E. *et al.* Plasma Proteome Profiling to Assess Human Health and Disease. *Cell Syst.* **2,** 185–195 (2016).
189. Tzoulaki, I., Ebbels, T. M. D., Valdes, A., Elliott, P. & Ioannidis, J. P. A. Design and Analysis of Metabolomics Studies in Epidemiologic Research: A Primer on -Omic Technologies. *Am. J. Epidemiol.* **180,** 129–139 (2014).
190. Guo, L. *et al.* Plasma metabolomic profiles enhance precision medicine for volunteers of normal health. *PNAS* **112,** E4901–E4910 (2015).
191. Sensen, S. *et al.* A plasma lipidomics strategy reveals perturbed lipid metabolic pathways and potential lipid biomarkers of human colorectal cancer. *J. Chromatogr. B* **1068–1069,** 41–48 (2017).
192. Snyder, C. M. *et al.* Complementary Glycomic Analyses of Sera Derived from Colorectal Cancer Patients by MALDI-TOF-MS and Microchip Electrophoresis. *Anal. Chem.* **88,** 9597–9605 (2016).
193. Segata, N. *et al.* Computational meta'omics for microbial community studies. *Mol. Syst. Biol.* **9,** 666 (2013).
194. Horgan, R. P. & Kenny, L. C. ' Omic ' technologies: genomics, transcriptomics, proteomics and metabolomics. *Obs. Gynaecol.* **13,** 189–195 (2011).
195. Sethi, M. K. & Fanayan, S. Mass Spectrometry-Based N -Glycomics of Colorectal Cancer. *Int. J. Mol. Sci.* **16,** 29278–29304 (2015).
196. Perrotti, F. *et al.* Advances in Lipidomics for Cancer Biomarkers Discovery. *Int. J. Mol. Sci.* **17,** 1992 (2016).
197. Ferguson, J. F. Omics Gets Personal: Integrative Profiling of Health and Disease. *Circ. Cardiovasc. Genet.* **5,** 381–382 (2012).
198. Ferlay, J. *et al.* Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. *GLOBOCAN 2012 v1.0* (2013).
199. Zackular, J. P. *et al.* The Gut Microbiome Modulates Colon Tumorigenesis. *MBio* **4,** e00692-13 (2013).
200. Feng, Q. *et al.* Gut microbiome development along the colorectal adenoma – carcinoma sequence. *Nat. Commun.* **6,** 6528 (2015).

201. Hibberd, A. A. *et al.* Intestinal microbiota is altered in patients with colon cancer and modified by probiotic intervention. *BMJ Open Gastroenterol.* **4,** e000145 (2017).
202. Shen, X. J. *et al.* Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. *Gut Microbes* **1,** 138–147 (2010).
203. Vogtmann, E. *et al.* Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole- Genome Shotgun Sequencing. *PLoS One* **11,** e0155362 (2016).
204. Purcell, R. V, Visnovska, M., Biggs, P. J., Schmeier, S. & Frizelle, F. A. Distinct gut microbiome patterns associate with consensus molecular subtypes of colorectal cancer. *Sci. Rep.* **7,** 11590 (2017).
205. Boleij, A. & Tjalsma, H. The itinerary of *Streptococcus gallolyticus* infection in patients with colonic malignant disease. *Lancet Infect. Dis.* **13,** 719–724 (2013).
206. Sobhani, I. *et al.* Microbial dysbiosis and colon carcinogenesis: could colon cancer be considered a bacteria-related disease ? *Therap. Adv. Gastroenterol.* **6,** 215–229 (2013).
207. Nakatsu, G. *et al.* Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat. Commun.* **6,** 8727 (2015).
208. Sears, C. L. & Garrett, W. S. Microbes, Microbiota and Colon Cancer. *Cell Host Microbe* **15,** 317–328 (2015).
209. Dai, Z. *et al.* Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* **6,** 70 (2018).
210. 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).
211. Lapointe, L. C. *et al.* Discovery and Validation of Molecular Biomarkers for Colorectal Adenomas and Cancer with Application to Blood Testing. *PLoS One* **7,** e29059 (2012).
212. Arthur, J. C. *et al.* Intestinal Inflammation Targets Cancer-Inducing Activity of the Microbiota. *Science (80-.).* **338,** 120–123 (2013).

213. Sole, X. *et al.* Discovery and Validation of New Potential Biomarkers for Early Detection of Colon Cancer. *PLoS One* **9,** e106748 (2014).
214. Zeller, G. *et al.* Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10,** 766 (2014).
215. Hale, V. L. *et al.* Shifts in the fecal microbiota associated with adenomatous polyps. *Cancer Epidemiol. Biomarkers Prev.* **26,** 85–94 (2018).
216. Jung, Y. *et al.* Clinical Validation of Colorectal Cancer Biomarkers Identified from Bioinformatics Analysis of Public Expression Data. *Clin. Cancer Res.* **17,** 700–710 (2011).
217. Parker, C. E. & Borchers, C. H. Mass spectrometry based biomarker discovery, verification, and validation - Quality assurance and control of protein biomarker assays. *Mol. Oncol.* **8,** 840–858 (2014).
218. Kuppusamy, P., Govindan, N., Yusoff, M. M. & Ichwan, S. J. A. Proteins are potent biomarkers to detect colon cancer progression. *Saudi J. Biol. Sci.* **24,** 1212–1221 (2017).
219. Lanschot, M. C. J. Van *et al.* Molecular stool testing as an alternative for surveillance colonoscopy: a cross-sectional cohort study. *BMC Cancer* **17,** 116 (2017).
220. Shiromizu, T. *et al.* Quantitation of putative colorectal cancer biomarker candidates in serum extracellular vesicles by targeted proteomics. *Sci. Rep.* **7,** 12782 (2017).
221. Marchesi, J. R. *et al.* Towards the Human Colorectal Cancer Microbiome. *PLoS One* **6,** e20447 (2011).
222. Viljoen, K. S., Dakshinamurthy, A., Goldberg, P. & Blackburn, J. M. Quantitative Profiling of Colorectal Cancer- Associated Bacteria Reveals Associations between *Fusobacterium* spp ., Enterotoxigenic *Bacteroides fragilis* (ETBF) and Clinicopathological Features of Colorectal Cancer. *PLoS One* **10,** e0119462 (2015).
223. Goedert, J. J. *et al.* Fecal metabolomics: assay performance and association with colorectal cancer. *Carcinogenesis* **35,** 2089–2096 (2014).
224. Gagnière, J. *et al.* Gut microbiota imbalance and colorectal cancer. *World J. Gastroenterol.* **22,** 501–518 (2016).

225. Sly, J. R., Edwards, T., Shelton, R. C. & Jandorf, L. Identifying Barriers to Colonoscopy Screening for Nonadherent African American Participants in a Patient Navigation Intervention. *Heal. Educ. Behav.* **40,** 449–457 (2013).
226. Talaat, N. & Harb, W. Reluctance to Screening Colonoscopy in Arab Americans: A Community Based Observational Study. *J. Community Health* **38,** 619–625 (2013).
227. Wang, W.-L. *et al.* Application of metagenomics in the human gut microbiome. *World J. Gastroenterol.* **21,** 803–814 (2015).
228. Gill, S. R. *et al.* Metagenomic Analysis of the Human Distal Gut Microbiome. *Science.* **312,** 1355–1359 (2006).
229. Rosselli, R. *et al.* Direct 16S rRNA-seq from bacterial communities: a PCR-independent approach to simultaneously assess microbial diversity and functional activity potential of each taxon. *Sci. Rep.* **6,** 32165 (2016).
230. Imaeda, H. *et al.* Terminal-Restriction Fragment Length Polymorphism (T-RFLP) Analysis for Changes in the Gut Microbiota Profiles of Indomethacin- and Rebamipide-Treated Mice. *Digestion* **86,** 250–257 (2012).
231. Liu, T. *et al.* Analysis of microbial diversity in Shenqu with different fermentation times by PCR-DGGE. *Brazilian J. Microbiol.* **48,** 246–250 (2017).
232. Degnan, P. H. & Ochman, H. Illumina-based analysis of microbial community diversity. *ISME J.* **6,** 183–194 (2012).
233. Wang, Y. *et al.* Profiling of Oral Microbiota in Early Childhood Caries Using Sequencing. *Front. Microbiol.* **8,** 2244 (2017).
234. Zhang, J. *et al.* 454 pyrosequencing reveals changes in the faecal microbiota of adults consuming *Lactobacillus casei* Zhang. *FEMS Microbiol. Ecol.* **88,** 612–622 (2014).
235. Frankel, A. E. *et al.* Metagenomic Shotgun Sequencing and Unbiased Metabolomic Profiling Identify Specific Human Gut Microbiota and Metabolites Associated with Immune Checkpoint Therapy Efficacy in Melanoma Patients. *Neoplasia* **19,** 848–855 (2017).
236. Ji, B. & Nielsen, J. From next-generation sequencing to systematic modeling of the gut microbiome. *Front. Genet.* **6,** 219 (2015).

237. Manor, O. *et al.* Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. *Sci. Rep.* **6,** 22493 (2016).
238. Moustafa, A. *et al.* Genetic risk, dysbiosis, and treatment stratification using host genome and gut microbiome in inflammatory bowel disease. *Clin. Transl. Gastroenterol.* **9,** e132 (2018).
239. Verma, M. K. *et al.* Functional metagenomics identifies novel genes ABCTPP, TMSRP1 and TLSRP1 among human gut enterotypes. *Sci. Rep.* **8,** 1397 (2018).
240. Lakhdari, O. *et al.* Identification of NF- κ B Modulation Capabilities within Human Intestinal Commensal Bacteria. *J. Biomed. Biotechnol.* **2011,** 282356 (2011).
241. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464,** 59–65 (2010).
242. Argenio, V. D' & Salvatore, F. The role of the gut microbiome in the healthy adult status. *Clin. Chim. Acta* **451,** 97–102 (2015).
243. Ventura, M. *et al.* Genome-scale analyses of health- promoting bacteria: probiogenomics. *Nat. Rev. Microbiol.* **7,** 61–71 (2009).
244. Kurokawa, K. *et al.* Comparative Metagenomics Revealed Commonly Enriched Gene Sets in Human Gut Microbiomes. *DNA Res.* **14,** 169–181 (2007).
245. Zackular, J. P., Rogers, M. A. M., IV, M. T. R. & Schloss, P. D. The Human Gut Microbiome as a Screening Tool for Colorectal Cancer. *Cancer Prev. Res.* **7,** 1112–1121 (2015).
246. Ijssennagger, N. *et al.* Gut microbiota facilitates dietary heme-induced epithelial hyperproliferation by opening the mucus barrier in colon. *PNAS* **112,** 10038–10043 (2015).
247. Brim, H. *et al.* Microbiome Analysis of Stool Samples from African Americans with Colon Polyps. *PLoS One* **8,** e81352 (2013).
248. Brim, H. *et al.* Colonic Lesions Reveals *Streptococcus* sp . VT162 as a Marker of Neoplastic Transformation. *Genes (Basel).* **8,** 314 (2017).
249. Hester, C. M. *et al.* Fecal microbes , short chain fatty acids , and colorectal cancer across racial / ethnic groups. *World J. Gastroenterol.* **21,** 2759–2769 (2015).

250. Lieberman, D. A. *et al.* Race, Ethnicity, and Sex Affect Risk for Polyps Greater than 9mm in Average-risk Individuals. *Gastroenterology* **147,** 351–358 (2015).
251. Wang, H. *et al.* Novel colon cancer susceptibility variants identified from a genomewide association study in African Americans. *Int. J. Cancer* **140,** 2728–2733 (2017).
252. Ahn, J. *et al.* Human Gut Microbiome and Risk for Colorectal Cancer. *J. Natl. Cancer Inst.* **105,** 1907–1911 (2013).
253. Eklof, V. *et al.* Cancer-associated fecal microbial markers in colorectal cancer. *Int. J. Cancer* **141,** 2528–2536 (2017).
254. Sobhani, I. *et al.* Microbial Dysbiosis in Colorectal Cancer (CRC) Patients. *PLoS One* **6,** e16393 (2011).
255. Arthur, J. C. *et al.* Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. *Nat. Commun.* **5,** 4724 (2014).
256. Scherl-Mostageer, M. *et al.* Identification of a novel gene, CDCP1, overexpressed in human colorectal cancer. *Oncogene* **20,** 4402–4408 (2001).
257. Shimpoh, T. *et al.* Prevalence of pks-positive *Escherichia coli* in Japanese patients with or without colorectal cancer. *Gut Pathog.* **9,** 35 (2017).
258. Han, A., Bennett, N., Ahmed, B., Whelan, J. & Donohoe, D. R. Butyrate decreases its own oxidation in colorectal cancer cells through inhibition of histone deacetylases. *Oncotarget* **9,** 27280–27292 (2018).
259. Zhou, Y. N. *et al.* Isolation and Characterization of RNA Polymerase rpoB Mutations That Alter Transcription Slippage during Elongation in *Escherichia coli*. *J. Biol. Chem.* **288,** 2700–2710 (2013).
260. Bashiardes, S., Zilberman-Schapira, G. & Elinav, E. Use of Metatranscriptomics in Microbiome Research. *Bioinform. Biol. Insights* **10,** 19–25 (2016).
261. Aidy, S. El *et al.* Gut bacteria – host metabolic interplay during conventionalisation of the mouse germfree colon. *ISME J.* **7,** 743–755 (2013).
262. Pfalzer, A. C. *et al.* Interactions between the colonic transcriptome, metabolome, and microbiome in mouse models of obesity-induced intestinal cancer. *Physiol. Genomics* **48,** 545–553 (2016).

263. Richards, A. L. *et al.* Genetic and Transcriptional Analysis of Human Host Response to Healthy Gut Microbiota. *mSystems* **1,** e00067-16 (2016).
264. Häsler, R. *et al.* Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease. *Gut* **66,** 2087–2097 (2017).
265. Wu, C.-H., Chung, F.-Y., Chang, J.-Y. & Wang, J.-Y. Rapid detection of gene expression by a colorectal cancer Enzymatic Gene Chip Detection Kit. *Biomarkers Genomic Med.* **5,** 87–91 (2013).
266. Zeka, F. *et al.* Straightforward and sensitive RT-qPCR based gene expression analysis of FFPE samples. *Sci. Rep.* **6,** 21418 (2016).
267. Xu, X. *et al.* Parallel comparison of Illumina RNA-Seq and Affymetrix microarray platforms on transcriptomic profiles generated from 5-aza-deoxy-cytidine treated HT-29 colon cancer cells and simulated datasets. *BMC Bioinformatics* **14,** S1 (2013).
268. Lee, J.-R. *et al.* Transcriptome analysis of paired primary colorectal carcinoma and liver metastases reveals fusion transcripts and similar gene expression profiles in primary carcinoma and liver metastases. *BMC Cancer* **16,** 539 (2016).
269. Conway, T. & Schoolnik, G. K. Microarray expression profiling: capturing a genome-wide portrait of the transcriptome. *Mol. Microbiol.* **47,** 879–889 (2003).
270. Ericsson, A. C. *et al.* Noninvasive Detection of Inflammation-Associated Colon Cancer in a Mouse Model. *Neoplasia* **12,** 1054–1065 (2010).
271. Vliet, A. H. M. Van. Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS Microbiol. Lett.* **302,** 1–7 (2010).
272. Hou, Z. *et al.* A cost-effectivness RNA sequencing protocol for large-scale gene expression studies. *Sci. Rep.* **5,** 9570 (2015).
273. Chen, L., Zhang, Y.-H., Huang, T. & Cai, Y.-D. Gene expression profiling gut microbiota in different races of humans. *Sci. Rep.* **6,** 23075 (2016).
274. Franzosa, E. A. *et al.* Relating the metatranscriptome and metagenome of the human gut. *PNAS* **22,** E2329–E2338 (2014).
275. Booijink, C. C. G. M. *et al.* Metatranscriptome Analysis of the Human Fecal Microbiota Reveals Subject-Specific Expression Profiles, with Genes Encoding

Proteins Involved in Carbohydrate Metabolism Being Dominantly Expressed. *Appl. Environ. Microbiol.* **76,** 5533–5540 (2010).

276. Gosalbes, M. J. *et al.* Metatranscriptomic Approach to Analyze the Functional Human Gut Microbiota. *PLoS One* **6,** e17447 (2011).
277. Heintz-Buschart, A. *et al.* Integrated multi-omics of the human gut microbiome in a case study of familial type 1 diabetes. *Nat. Microbiol.* **2,** 16180 (2016).
278. Tap, J. *et al.* Early Phenotypical Changes Induced By Transfer of Human Fecal Microbiota from Colorectal Cancer Patients to Germ-Free Mice. *Int. J. Cancer Oncol.* **3,** 1–11 (2016).
279. Castellarin, M. *et al. Fusobacterium nucleatum* infection is prevalent in human colorectal carcinoma. *Genome Res.* **22,** 299–306 (2012).
280. McCoy, A. N. *et al. Fusobacterium* Is Associated with Colorectal Adenomas. *PLoS One* **8,** e53653 (2013).
281. Kostic, A. D. *et al. Fusobacterium nucleatum* potentiates intestinal tumorigenesis and modulates the tumor immune microenvironment. *Cell Host Microbe* **14,** 207–215 (2013).
282. Matusiak, D. & Benya, R. V. CYP27A1 and CYP24 Expression as a Function of Malignant Transformation in the Colon. *J. Histochem. Cytochem.* **55,** 1257–1264 (2007).
283. Ji, Y.-C. *et al.* Intestinal bacterium-derived cyp27a1 prevents colon cancer cell apoptosis. *Am. J. Transl. Res.* **8,** 4434–4439 (2016).
284. Xin, J., Ren, X., Chen, L. & Wang, Y. Identifying network biomarkers based on protein- protein interactions and expression data. *BMC Med. Genomics* **8,** S11 (2015).
285. Durand, S. *et al.* A transcriptome-based protein network that identifies new therapeutic targets in colorectal cancer. *BMC Genomics* **18,** 758 (2017).
286. Roumeliotis, T. I. *et al.* Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. *Cell Rep.* **20,** 2201–2214 (2017).
287. Tonry, C. L. *et al.* The Role of Proteomics in Biomarker Development for Improved Patient Diagnosis and Clinical Decision Making in Prostate Cancer. *Diagnostics* **6,** 27

(2016).

288. Zanzoni, A., Spinelli, L., Braham, S. & Brun, C. Perturbed human sub-networks by *Fusobacterium nucleatum* candidate virulence proteins. *Microbiome* **5,** 89 (2017).
289. Wang, H. *et al.* Evaluation of specific fecal protein biochips for the diagnosis of colorectal cancer. *World J. Gastroenterol.* **20,** 1332–1339 (2014).
290. Petriz, B. A. & Franco, O. L. Metaproteomics as a Complementary Approach to Gut Microbiota in Health and Disease. *Front. Chem.* **5,** 4 (2017).
291. Juste, C. *et al.* Bacterial protein signals are associated with Crohn ' s disease. *Gut* **63,** 1566–1577 (2014).
292. Presley, L. L. *et al.* Host-Microbe Relationships in Inflammatory Bowel Disease Detected by Bacterial and Metaproteomic Analysis of the Mucosal-Luminal Interface. *Inflamm. Bowel Dis.* **18,** 409–417 (2013).
293. Erickson, A. R. *et al.* Integrated Metagenomics / Metaproteomics Reveals Human Host-Microbiota Signatures of Crohn ' s Disease. *PLoS One* **7,** e49138 (2012).
294. Xu, M. *et al.* FadA from *Fusobacterium nucleatum* Utilizes both Secreted and Nonsecreted Forms for Functional Oligomerization for Attachment and Invasion of Host Cells. *J. Biol. Chem.* **282,** 25000–25009 (2007).
295. Lu, R., Bosland, M., Xia, Y., Zhang, Y.-G. & Kato, I. Presence of *Salmonella* AvrA in colorectal tumor and its precursor lesions in mouse intestine and human specimens. *Oncotarget* **8,** 55104–55115 (2017).
296. Wang, H.-F. *et al.* Evaluation of antibody level against *Fusobacterium nucleatum* in the serological diagnosis of colorectal cancer. *Sci. Rep.* **6,** 33440 (2016).
297. Wu, J. *et al.* Development of an Integrated Pipeline for Profiling Microbial Proteins from Mouse Fecal Samples by LC − MS/MS. *J. Proteome Res.* **15,** 3625–3642 (2016).
298. Chung, L. *et al. Bacteroides fragilis* Toxin Coordinates a Pro- carcinogenic Inflammatory Cascade via Targeting of Colonic Epithelial Cells. *Cell Host Microbe* **23,** 1–12 (2018).
299. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45,** 158–169 (2017).
300. Kriventseva, E. V, Servant, F. & Apweiler, R. Improvements to CluSTr: the database

of SWISS-PROT + TrEMBL protein clusters. *Nucleic Acids Res.* **31,** 388–389 (2003).

301. Goll, J. *et al.* MPIDB: the microbial protein interaction database. *Bioinformatics* **24,** 1743–1744 (2008).
302. Wiwanitkit, V. & House, W. Utilization of multiple 'omics' studies in microbial pathogeny for microbiology insights. *Asian Pac. J. Trop. Biomed.* **3,** 330–333 (2013).
303. Blackburn, J. M. & Martens, L. The challenge of metaproteomic analysis in human samples. *Expert Rev. Proteomics* **13,** 135–138 (2016).
304. Lee, P. Y., Chin, S.-F., Neoh, H.-M. & Jamal, R. Metaproteomic analysis of human gut microbiota: where are we heading? *J. Biomed. Sci.* **24,** 36 (2017).
305. Kilpatrick, L. E. & Kilpatrick, E. L. Optimizing High Resolution Mass Spectrometry for the Identification of Low Abundance Post-Translational Modifications of Intact Proteins. *J. Proteome Res.* **16,** 3255–3265 (2017).
306. Siggins, A., Gunnigle, E. & Abram, F. Exploring mixed microbial community functioning: recent advances in metaproteomics. *FEMS Microbiol. Ecol.* **80,** 265–280 (2012).
307. Muth, T., Benndorf, D., Reichl, U., Rapp, E. & Martens, L. Searching for a needle in a stack of needles: challenges in metaproteomic data analysis. *Mol. Biosyst.* **9,** 578–585 (2013).
308. Jin, P., Wang, K., Huang, C. & Nice, E. C. Mining the Fecal Proteome: From Biomarkers to Personalised Medicine. *Expert Rev. Proteomics* **14,** 445–459 (2017).
309. Dick, J. M. Proteomic indicators of oxidation and hydration state in colorectal cancer. *PeerJ* **4,** e2238 (2016).
310. Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E. & Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat. Biotechnol.* **36,** 316–320 (2018).
311. Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17,** 451–459 (2017).
312. Patti, G. J., Yanes, O. & Siuzdak, G. Metabolomics: the apogee of the omic triology. *Nat. Rev. Mol. Cell Biol.* **13,** 263–269 (2013).
313. Lee, D. Y., Bowen, B. P. & Northen, T. R. Mass spectrometry-based metabolomic,

analysis of metabolite-protein interactions, and imaging. *Biotechniques* **49,** 557–565 (2011).

314. Ramsden, J. J. in *Bioinformatics. Computational Biology* 1–271 (Springer, 2009). doi:10.1007/978-1-84800-257-9 1
315. Nicholson, J. K. & Lindon, J. C. Metabonomics. *Nature* **455,** 1054–1056 (2008).
316. Ryan, D. & Robards, K. Metabolomics: The Greatest Omics of Them All ? *Anal. Chem.* **78,** 7954–7958 (2006).
317. Armitage, E. G. & Barbas, C. Metabolomics in cancer biomarker discovery: Current trends and future perspectives. *J. Pharm. Biomed. Anal.* **87,** 1–11 (2014).
318. Deng, L. *et al.* Development and Validation of a High-Throughput Mass Spectrometry Based Urine Metabolomic Test for the Detection of Colonic Adenomatous Polyps. *Metabolites* **7,** 32 (2017).
319. Liang, Q., Liu, H., Xie, L.-X., Li, X. & Zhang, A.-H. High-throughput metabolomics enables biomarker discovery in prostate cancer. *RSC Adv.* **7,** 2587–2593 (2016).
320. Roig, B. *et al.* Metabolomics reveals novel blood plasma biomarkers associated to the BRCA1-mutated phenotype of human breast cancer. *Sci. Rep.* **7,** 17831 (2017).
321. Farshidfar, F. *et al.* A validated metabolomic signature for colorectal cancer: exploration of the clinical value of metabolomics. *Br. J. Cancer* **115,** 848–857 (2016).
322. Veenstra, T. D. Metabolomics: the final frontier? *Genome Med.* **4,** 40 (2012).
323. Tamang, J. P., Shin, D.-H., Jung, S.-J. & Chae, S.-W. Functional Properties of Microorganisms in Fermented Foods. *Front. Microbiol.* **7,** 578 (2016).
324. Hamilton, M. K., Boudry, G., Lemay, D. G. & Raybould, H. E. Changes in intestinal barrier function and gut microbiota in high-fat diet-fed rats are dynamic and region dependent. *Am. J. Physiol. Liver Physiol.* **308,** 840–851 (2015).
325. Woods, J. A. *et al.* Exercise alters the gut microbiome and microbial metabolites: Implications for colorectal cancer and inflammatory bowel disease. *Brain. Behav. Immun.* **49,** e7 (2015).
326. Jiang, W. *et al.* Dysbiosis gut microbiota associated with inflammation and impaired mucosal immune function in intestine of humans with non-alcoholic fatty liver

disease. *Sci. Rep.* **5,** 8096 (2015).

327. Chu, H. *et al.* Gene-microbiota interactions contribute to the pathogenesis of inflammatory bowel disease. *Science (80-.).* **352,** 1116–1120 (2016).
328. Phua, L. C. *et al.* Non-invasive fecal metabonomic detection of colorectal cancer. *Cancer Biol. Ther.* **15,** 389–397 (2014).
329. 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.* **13,** 2848–2854 (2017).
330. Sinha, R. *et al.* Fecal Microbiota, Fecal Metabolome, and Colorectal Cancer Interrelations. *PLoS One* **11,** e0152126 (2016).
331. Pfalzer, A. C. *et al.* Diet- and Genetically-Induced Obesity Differentially Affect the Fecal Microbiome and Metabolome in Apc 1638N Mice. *PLoS One* **10,** e0135758 (2015).
332. Ríos-Covián, D. *et al.* Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. *Front. Microbiol.* **7,** 185 (2016).
333. Morrison, D. J. & Preston, T. Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **7,** 189–200 (2016).
334. Chen, H.-M. *et al.* Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma. *Am. J. Clin. Nutr.* **97,** 1044–1052 (2013).
335. Ou, J. *et al.* Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. *Am. J. Clin. Nutr.* **98,** 111–120 (2013).
336. Lin, Y. *et al.* NMR-based fecal metabolomics fingerprinting as predictors of earlier diagnosis in patients with colorectal cancer. *Oncotarget* **7,** 29454–29464 (2016).
337. Martin, H. M. *et al.* Enhanced *Escherichia coli* Adherence and Invasion in Crohn's Disease and Colon Cancer. *Gastroenterology* **127,** 80–93 (2004).
338. Scanlan, P. D. *et al.* Culture-independent analysis of the gut microbiota in colorectal cancer and polyposis. *Environ. Microbiol.* **10,** 789–798 (2008).
339. Huipeng, W., Lifeng, G., Chuang, G., Jiaying, Z. & Yuankun, C. The Differences in

Colonic Mucosal Microbiota Between Normal Individual and Colon Cancer Patients by Polymerase Chain Reaction-denaturing Gradient Gel Electrophoresis. *J. Clin. Gastroenterol.* **48,** 138–144 (2014).

340. Huycke, M. M., Abrams, V. & Moore, D. R. Enterococcus faecalis produces extracellular superoxide and hydrogen peroxide that damages colonic epithelial cell DNA. *Carcinogenesis* **23,** 529–536 (2002).
341. Malvezzi, M. *et al.* European cancer mortality predictions for the year 2018 with focus on colorectal cancer. *Ann. Oncol.* **29,** 1016–1022 (2018).
342. Yazici, C. *et al.* Race-dependent association of sulfidogenic bacteria with colorectal cancer. *Gut* **66,** 1983–1994 (2017).
343. Zhou, Y. *et al.* Association of oncogenic bacteria with colorectal cancer in South China. *Oncotarget* **7,** 80794–80802 (2016).
344. Allali, I. *et al.* Gut microbiome of Moroccan colorectal cancer patients. *Med. Microbiol. Immunol.* **207,** 211–225 (2018).
345. Allali, I. *et al.* Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. *Gut Microbes* **6,** 161–172 (2015).
346. Martins, S. F. *et al.* A Hospital Based Cohort Study of Colorectal Cancer Cases Treated at Braga Hospital, Northern Portugal. *Gastrointest. Dig. Syst.* **3,** 1000146 (2013).
347. Ferreira, R. M. *et al.* Gastric microbial community profiling reveals a dysbiotic cancer-associated microbiota. *Gut* **67,** 226–236 (2018).
348. Ravasco, P., Monteiro-Grillo, I., Vidal, P. M. & Camilo, M. E. Nutritional risks and colorectal cancer in a portuguese population. *Nutr. Hosp.* **20,** 165–172 (2005).
349. Amplicon, P. C. R., Clean-Up, P. C. R. & Index, P. C. R. 16S Metagenomic Sequencing Library Preparation. 1–28 (2013). at <https://www.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf>
350. Edgar, R. C. USEARCH Ultra-fast sequencing analysis. at

<https://www.drive5.com/usearch/>

351. Cole, J. R. *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **42,** 633–642 (2014).
352. RStudio Team. RStudio: Integrated Development for R. RStudio, Inc. (2016). at <http://www.rstudio.com/>
353. Desjardins, P. & Conklin, D. NanoDrop Microvolume Quantitation of Nucleic Acids. *J. Vis. Exp.* **45,** e2565 (2010).
354. Panek, M. *et al.* Methodology challenges in studying human gut microbiota – effects of collection, storage, DNA extraction and next generation sequencing technologies. *Sci. Rep.* **8,** 5143 (2018).
355. Chao, A., Chazdon, R. L., Colwell, R. K. & Shen, T.-J. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* **8,** 148–159 (2005).
356. Faith, D. P. & Baker, A. M. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evol. Bioinform. Online* **2,** 121–128 (2006).
357. DeJong, T. M. A comparison of three diversity indices based on their components of richness and evenness. *Oikos* **26,** 222–227 (1975).
358. Morgan, X. C. & Huttenhower, C. Chapter 12 : Human Microbiome Analysis. *PLos Comput. Biol.* **8,** e1002808 (2012).
359. Russo, E. *et al.* Preliminary Comparison of Oral and Intestinal Human Microbiota in Patients with Colorectal Cancer: A Pilot Study. *Front. Microbiol.* **8,** 2699 (2018).
360. Peters, B. A. *et al.* The gut microbiota in conventional and serrated precursors of colorectal cancer. *Microbiome* **4,** 69 (2016).
361. Goedert, J. J. *et al.* Fecal Microbiota Characteristics of Patients with Colorectal Adenoma Detected by Screening: A Population-based Study. *EBioMedicine* **2,** 597–603 (2015).
362. Bamola, V. D. *et al.* Gut microbial diversity in health and disease: experience of healthy Indian subjects, and colon carcinoma and inflammatory bowel disease patients. *Microb. Ecol. Health Dis.* **28,** 1322447 (2017).

363. Lu, Y. *et al.* Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. *Sci. Rep.* **6,** 26337 (2016).

364. Kinross, J. *et al.* A prospective analysis of mucosal interactions in colorectal cancer using a combined MAS ^{1}HNMR and metataxonomic strategy. *Sci. Rep.* **7,** 8979 (2017).

365. Binda, C. *et al.* Actinobacteria: a relevant minority for the maintenance of gut homeostasis. *Dig. Liver Dis.* **50,** 421–428 (2018).

366. 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).

367. Geirnaert, A. *et al.* Butyrate-producing bacteria supplemented in vitro to Crohn's disease patient microbiota increased butyrate production and enhanced intestinal epithelial barrier integrity. *Sci. Rep.* **7,** 11450 (2017).

368. Mottawea, W. *et al.* Altered intestinal microbiota-host mitochondria crosstalk in new onset Chron's disease. *Nat. Commun.* **7,** 13419 (2016).

369. Dabard, J. *et al.* Ruminococcin A, a New Lantibiotic Produced by a *Ruminococcus gnavus* Strain Isolated from Human Feces. *Appl. Environ. Microbiol.* **67,** 4111–4118 (2001).

370. Crost, E. H. *et al.* Ruminococcin C, a new anti-*Clostridium perfringens* bacteriocin produced in the gut by the commensal bacterium *Ruminococcus gnavus* E1. *Biochimie* **93,** 1487–1494 (2011).

371. Jessamy, K. *et al. Clostridium septicum*: An Unusual Link to a Lower Gastrointestinal Bleed. *Case Rep. Gastroenterol.* **10,** 489–493 (2016).

372. Kuugbee, E. D. *et al.* Structural Change in Microbiota by a Probiotic Cocktail Enhances the Gut Barrier and Reduces Cancer via TLR2 Signaling in a Rat Model of Colon Cancer. *Dig. Dis. Sci.* **61,** 2908–2920 (2016).

373. Alomair, A. O. *et al.* Colonic Mucosal Microbiota in Colorectal Cancer: A Single-Center Metagenomic Study in Saudi Arabia. *Gastroenterol. Res. Pract.* **2018,** 5284754 (2018).

374. Purcell, R. V *et al.* Colonization with enterotoxigenic *Bacteroides fragilis* is

associated with early-stage colorectal neoplasia. *PLoS One* **12,** e0171602 (2017).

375. Candela, M. *et al.* Inflammation and colorectal cancer, when microbiota-host mutualism breaks. *World J. Gastroenterol.* **20,** 908–922 (2014).
376. Cai, W.-J., Wang, M.-J., Ju, L.-H., Wang, C. & Zhu, Y.-C. Hydrogen sulfide induces human colon cancer cell proliferation: role of Akt , ERK and p21. *Cell Biol. Int.* **34,** 565–572 (2010).
377. Fukugaiti, M. H. *et al.* High occurrence of *Fusobacterium nucleatum* and *Clostridium difficile* in the intestinal microbiota of colorectal carcinoma patients. *Brazilian J. Microbiol.* **46,** 1135–1140 (2015).
378. Wong, S. H. *et al.* Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice. *Gastroenterology* **153,** 1621–1633.e6 (2017).
379. Abbeele, P. Van Den *et al.* Butyrate-producing *Clostridium* cluster XIVa species specifically colonize mucins in an in vitro gut model. *ISME J.* **7,** 949–961 (2012).
380. Ohigashi, S., Sudo, K. & Kobayashi, D. Changes of the Intestinal Microbiota, Short Chain Fatty Acids, and Fecal pH in Patients with Colorectal Cancer. *Dig. Dis. Sci.* **58,** 1717–1726 (2013).
381. Martín, R. *et al.* Functional Characterization of Novel *Faecalibacterium prausnitzii* Strains Isolated from Healthy Volunteers: A Step Forward in the Use of *F . prausnitzii* as a Next-Generation Probiotic. *Front. Microbiol.* **8,** 1226 (2017).
382. Chang, Y.-J. *et al.* Complete genome sequence of *Acidaminococcus fermentans* type strain (VR4T). *Stand. Genomic Sci.* **3,** 1–14 (2010).
383. Bener, A., Dogan, M., Azab, I. A., Rashed, A. & Siddiqui, M. Short communication: Amino acid profiles among colorectal cancer patients. *Biomed. Res.* **17,** 149–154 (2006).
384. Andoh, A. *et al.* Comparison of the gut microbial community between obese and lean peoples using 16S gene sequencing in a Japanese population. *J. Clin. Biochem. Nutr.* **59,** 65–70 (2016).
385. Larsen, J. M. The immune response to *Prevotella* bacteria in chronic inflammatory disease. *Immunology* **151,** 363–374 (2017).

386. Su, T. *et al.* Altered Intestinal Microbiota with Increased Abundance of *Prevotella* Is Associated with High Risk of Diarrhea-Predominant Irritable Bowel Syndrome. *Gastroenterol. Rep.* **2018,** 6961783 (2018).

387. Wright, D. P., Rosendale, D. I. & Roberton, A. M. *Prevotella* enzymes involved in mucin oligosaccharide degradation and evidence for a small operon of genes expressed during growth on mucin. *FEMS Microbiol. Lett.* **190,** 73–79 (2000).

388. Chen, C. *et al.* Oral microbiota of periodontal health and disease and their changes after nonsurgical periodontal therapy. *ISME J.* **12,** 1210–1224 (2018).

389. O'Hara, C. M., Brenner, F. W. & Miller, J. M. Classification, Identification, and Clinical Significance of *Proteus*, *Providencia*, and *Morganella*. *Clin. Microbiol. Rev.* **13,** 534–546 (2000).

390. Dzutsev, A. & Trinchieri, G. *Proteus mirabilis*:The Enemy Within. *Immunity* **42,** 602–604 (2015).

391. Hamilton, A. L., Kamm, M. A., Ng, S. C. & Morrison, M. *Proteus* spp. as Putative Gastrointestinal Pathogens. *Clin. Microbiol. Rev.* **31,** e00085-17 (2018).

392. Balamurugan, R., Rajendiran, E., George, S., Samuel, G. V. & Ramakrishna, B. S. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria , *Desulfovibrio* and *Enterococcus faecalis* in the feces of patients with colorectal cancer. *Gastroenterology* **23,** 1298–1303 (2008).

393. Loubinoux, J., Bronowicki, J.-P., Pereira, I. A. C., Mougenel, J.-L. & Faou, A. E. Le. Sulfate-reducing bacteria in human feces and their association with inflammatory bowel diseases. *FEMS Microbiol. Ecol.* **40,** 107–112 (2002).

394. Hiippala, K., Kainulainen, V., Kalliomäki, M., Arkkila, P. & Reetta Satokari. Mucosal Prevalence and Interactions with the Epithelium Indicate Commensalism of Sutterella spp. *Front. Microbiol.* **7,** 1706 (2016).

395. Lee, D. K. *et al.* Anti-proliferative effects of *Bifidobacterium adolescentis* SPM0212 extract on human colon cancer cell lines. *BMC Cancer* **8,** 310 (2008).

396. Wickramasinghe, S., Pacheco, A. R., Lemay, D. G. & Mills, D. A. *Bifidobacteria* grown on human milk oligosaccharides downregulate the expression of inflammation-related genes in Caco-2 cells. *BMC Microbiol.* **15,** 172 (2015).

397. Ling, X., Linglong, P., Weixia, D. & Hong, W. Protective Effects of *Bifidobacterium* on Intestinal Barrier Function in LPS-Induced Enterocyte Barrier Injury of Caco-2 Monolayers and in a Rat NEC Model. *PLoS One* **11,** e0161635 (2016).

398. Kraatz, M., Wallace, R. J. & Svensson, L. *Olsenella umbonata* sp . nov ., a microaerotolerant anaerobic lactic acid bacterium from the sheep rumen and pig jejunum, and emended descriptions of *Olsenella*, *Olsenella uli* and *Olsenella profusa*. *Int. J. Syst. Evol. Microbiol.* **61,** 795–803 (2011).

399. Rubinstein, M. R. *et al. Fusobacterium nucleatum* Promotes Colorectal Carcinogenesis by Modulating E-Cadherin / B -Catenin Signaling via its FadA Adhesin. *Cell Host Microbe* **14,** 195–206 (2013).

400. Clercq, C. M. C. *et al.* Postcolonoscopy colorectal cancers are preventable: a population-based study. *Gut* **63,** 957–963 (2014).

401. Abed, J. *et al.* Tumor Targeting by *Fusobacterium nucleatum*: A Pilot Study and Future Perspectives. *Front. Cell. Infect. Microbiol.* **7,** 295 (2017).

Annexes

Diversity of the Microbial Community (microbiome) in Stool Samples of Healthy Individuals *versus* Colorectal Cancer Patients

Carla Filipa Norte Gomes, Amadeu M. V. M. Soares, Catarina R. Marques

Department of Biology and CESAM, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal
*mail to: crmarques@ua.pt; cgomes2@ua.pt

cesam universidade de aveiro centro de estudos do ambiente e do mar | applEE applied Ecology and Ecotoxicology Research Group | dbio universidade de aveiro departamento de biologia | CENTRO HOSPITALAR BAIXO VOUGA | IV POSTGRAD SYMPOSIUM IN BIOMEDICINE

Introduction

- Colorectal Cancer (CRC) is the 3rd most frequently diagnosed cancer solely in the United States, with an alarming increase to approximately 1.1 million deaths by 2030[1–4].
- Changes in the delicate host-microbiome relationship (dysbiosis) has recently been targeted as an influencing factor of carcinogenesis, due to shifts in the microbial community sustaining different molecular/biochemical (*e.g.*, toxins, virulence factors) and functional (*e.g.*, modulation of host immune system) traits[5,6].
- The Portuguese population has been barely considered in gut microbiome studies particularly related with CRC, despite the relevant influence of the geographic area.
- **Aim:** to compare the microbiome profiles of fecal samples of healthy (H) individuals *versus* CRC-affected (C) individuals.

Material and Methods

- Study approved by the Ethics Committee of Centro Hospitalar Baixo Vouga, E. P. E., Aveiro, Portugal, and performed in compliance with Helsinki Declaration. All individuals signed an Informed Consent Form.
- **Sample Collection**: Fecal samples obtained from 6 healthy individuals and 6 patients diagnosed with CRC.
- **DNA extraction** (QIAamp® Powerfecal® DNA Kit) and quality analysis (Nanodrop™)
- Preparation of 16S library and sequencing on an Illumina MiSeq (Helmholtz Centre Munich)
- **Data analysis**: richness, relative abundance (%) of phyla and families

Results and Discussion

Table 1: Quality of microbial gDNA extracted from stool samples of CRC-diagnosed (C) and healthy (H) individuals.

Sample	[DNA] ng/µL	A_{260}/A_{280}	Sample	[DNA] ng/µL	A_{260}/A_{280}
C1	76.8	1.82	**H1**	134.7	1.80
C2	123.3	1.82	**H2**	126.8	1.75
C3	103.2	1.83	**H3**	74.7	1.74
C4	100.9	1.80	**H4**	104.1	1.64
C5	68.0	1.79	**H5**	149.2	1.78
C6	62.4	1.81	**H6**	170.8	1.83

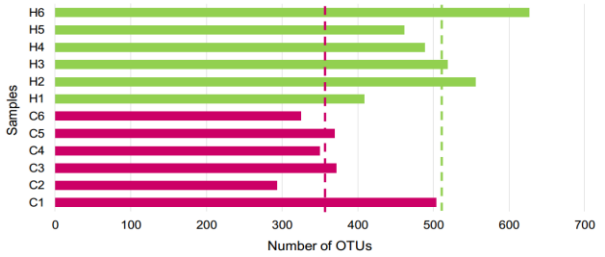

Figure 1: Number of Operational Taxonomic Unit (OTUs) of bacteria from individual stool samples of healthy (H) and CRC-diagnosed (C) individuals. Dashed lines represent the average number of OTUs within each clinical group.

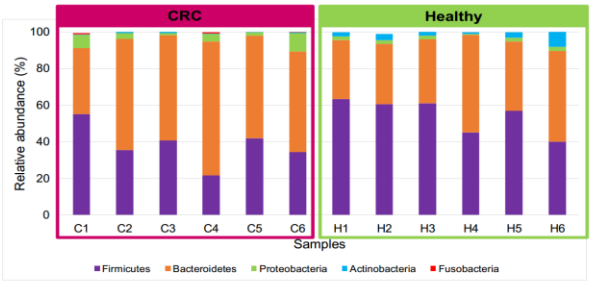

Figure 2: Phyla-level microbial relative abundance of bacteria from individual stool samples of healthy (H) and CRC-diagnosed (C) individuals.

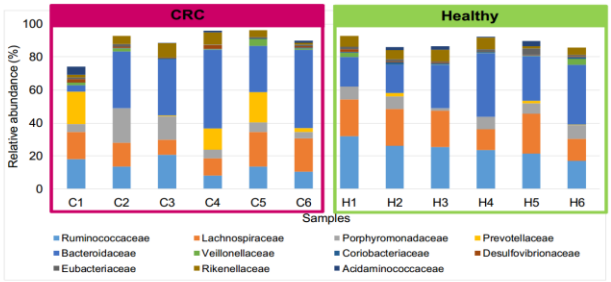

Figure 3: Family-level microbial relative abundance of bacteria from individual stool samples of healthy (H) and CRC-diagnosed (C) individuals.

- Suitable DNA concentration ranging from 62.4-170.8 ng μL^{-1} with relatively good quality (~1.8) (Table 1).
- H samples had generally greater bacterial diversity (> no. OTUs) comparatively to C samples, demonstrating a loss of bacterial diversity (Fig. 1).
- The microbiome between and within H and C samples are constituted by similar phyla, being Firmicutes and Bacteroidetes the most represented (Fig. 2).
- Although Fusobacteria has been linked to CRC and Actinobacteria with healthy states, the former was underrepresented in CRC patients (Fig. 2).
- The most dominant families observed are Ruminococcaceae (H>C), Lachnospiraceae (H>C), Porphyromondaceae (H<C), Prevotellaceae (H<C), and Bacteroidaceae(H≈C) (Fig. 3).
- Some heterogeneity is observed between the clinical groups and within C individuals, whereas relative homogeneity was detected in the H group (Fig. 3).

Conclusions

- Concrete bacterial associations to CRC *versus* healthy states are yet to be deeply resolved through the application of larger sample sizes.
- CRC-associated microbiome showed a loss in bacterial community diversity, but relatively overall homogenous and heterogeneous abundances with respect to phyla and family, respectively.
- The variability detected between both clinical groups should be further studied in order to identify significant bacterial species with biomedical relevance, namely for disease diagnosis.

References

1. Arnold, M. *et al.* Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **66**, 683–691 (2017).
2. Kang, M. & Martin, A. Microbiome and colorectal cancer : Unraveling host-microbiota interactions in colitis-associated colorectal cancer development. *Semin. Immunol.* **32**, 3-13 (2017).
3. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer.* **136**, 359-386 (2015).
4. Cotter, J. Colorectal Cancer: Portugal and the World. *Acta Med. Port.* **26**, 485–486 (2013).
5. Marchesi, J. R. *et al.* The gut microbiota and host health : a new clinical frontier. *Gut* **65**, 330–339 (2016).
6. Gao, Z., *et al.* Microbiota dysbiosis is associated with colorectal cancer. *Front. Microbiol.* **6**, 1–9 (2015).

Acknowledgments

We acknowledge the volunteers participating in this study, as well as the Doctors, Nurses and Administrative stuff for their excellent collaboration and kind availability, specially those from the services of Cirurgia e Medicina do Trabalho. Financial support to CESAM (UID/AMB/50017), to FCT/MEC through national funds, and the co-funding by the FEDER (POCI-01-0145-FEDER-007638), within the PT2020 Partnership Agreement and Compete 2020.

PORTUGAL 2020 | COMPETE 2020 | FCT Fundação para a Ciência e a Tecnologia MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Figure 19: Submitted Poster for the IV PostGrad Symposium in Biomedicine, ibiMed University of Aveiro, 2018