PhD degree in Systems Medicine (curriculum in Computational Biology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples “Federico II”

Fecal microbiota, serum biomarkers, diet and lifestyle: colorectal cancer risk and prognosis

Settore disciplinare: MED/04

Federica Bellerba

Tutor: Prof. Sara Gandini

European Institute of Oncology, Milan, Italy

PhD Coordinator: Prof. Saverio Minucci

Anno accademico 2022-2023

Abstract

Colorectal cancer (CRC) is the result of a complex interaction between non-modifiable and modifiable risk factors. The modifiable factors include obesity, diet, lifestyle choices, inflammatory markers, and vitamin D (vitD) status. Recent literature also suggests an important link between gut microbiota and CRC prognosis and progression, however assessing whether the relationship is causal is challenging. We conducted a comprehensive investigation of these factors to provide new insights on how their interplay affects CRC. We employed a multi-step approach.

First, we designed a case-control study of CRC patients and healthy individuals. We found that several species, such as *Parvimonas micra*, *Fusobacterium nucleatum* and *Bacteroides fragilis* were significantly more abundant in cases. A poor lifestyle and a high-risk diet were significantly associated with CRC and mediation analysis suggested that the gut microbiota mediated the effect of diet on CRC risk.

Then, we carried out a systematic review of the literature on human studies, to summarize the evidence published so far on the relationship between gut microbiota and vitD. We found that Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria were the most recurrent phyla increasing or decreasing following vitD supplementation and at increasing vitD serum levels or dietary intake.

Finally, we designed a phase II randomized trial involving vitD supplementation or placebo for 1 year and including CRC survivors. Gut microbiota, circulating markers, diet and lifestyle were collected at baseline and at the end of the treatment. We found increased abundances of several probiotic taxa following vitD supplementation, including *Faecalibacterium prausnitzii* and *Holdemanella biformis.* We also found that the microbiota significantly mediated the effect of the supplementation on 25(OH)D levels. In the supplemented group, we found differences by sex/gender in the pathways involved in the biosynthesis of essential amino-acids. The weight status of the participants modulated the effect of the supplementation on both 25(OH)D levels and alpha diversity. In the supplemented group, 25(OH)D levels increased less at increasing Body Mass Index (BMI), while the change in alpha diversity was significantly and positively correlated with the change in 25(OH)D levels only in normal-weight individuals (BMI<25). For a subgroup of patients, we collected the gene expression (GE) profile evaluated on tumour tissue of a panel of 395 immuno-related genes. We identified three clusters of patients based solely on GE. One of the clusters was associated with a higher risk of colorectal and clinical events. Alpha diversity at baseline was also significantly and inversely associated with the risk of colorectal and clinical events, while vitD supplementation and Galactin-9 had a protective effect on both outcomes.

Overall, our findings provide new insights on the complex interconnection between modifiable risk factors of CRC and highlight the importance of personalized strategies.

Table of Contents

Listing of Figures

Acronyms

25(OH)D: 25-hydroxyvitamin D

AICR: American Institute for Cancer Research

alr: additive log-ratio

APC: Adenomatous Polyposis Coli

BH: Benjamini–Hochberg

BIC: Bayesian Information Criterion

BM: Bayesian-multiplicative

BMI: Body Mass Index

CA: Correspondence Analysis

CCA: Canonical Correspondence Analysis

CD: Crohn's Disease

CI: Confidence Interval

clr: centered log-ratio

CoDA: compositional data analysis

CPM: counts per million

CRC: Colorectal Cancer

CV: Cross-validation

DAG: Directed Acyclic Graph

DIABLO: Data Integration Analysis for Biomarker Discovery

DL: Detection Limit

EFS: Event-Free Survival

FAP: Familial Adenomatous Polyposis

FDR: False Discovery Rate

FMT: Fecal Microbiota Transplantation

GE: Gene Expression

GLASSO: Graphical Least Absolute Shrinkage and Selection Operator

HR: Hazard Ratio

hs-CRP: high-sensitivity C-reactive protein

IARC: International Agency for Research on Cancer

IBD: Inflammatory bowel disease

ICH-GCP: Good Clinical Practice

ilr: isometric log-ratio

IU: International Unit

LASSO: Least Absolute Shrinkage and Selection Operator

LDA: Linear Discriminant Analysis

LEfSe: Linear Discriminant Analysis effect size

LS: Lynch Syndrome

MAP: MUTYH-Associated Polyposis

NDE: Natural Direct Effect

NF-κB: Nuclear Factor kappa B

NGS: Next Generation Sequencing

NIE: Natural Indirect Effect

OIRRA: Oncomine Immune Response Research Assay

OR: Odds Ratio

OTU: Operational Taxonomic Unit

PC1: First Principal Component

PC2: Second Principal Component

PCA: Principal Component Analysis

PCR: Polymerase Chain Reaction

PE: Paired End

PEAR: Pair-End Read Merger

PERMANOVA: Permutational Multivariate Analysis of Variance

PLS-DA: Partial Least Squares Discriminant Analysis

RCT: Randomized Clinical Trials

RPM: Reads Per Million

SBP: Sequential Binary Partition

SNP: Single Nucleotide Polymorphism

sPLS-DA: sparse Partial Least Squares Discriminant Analysis

SRR: Summary Relative Risk

TE: Total Effect

UC: Ulcerative Colitis

VDBP: Vitamin D binding protein

VDR: Vitamin D receptor

VITAL: Vitamin D and Omega-3 Trial

vitD: Vitamin D

WCRF: World Cancer Research Fund

1. BACKGROUND ON COLORECTAL CANCER

1.1 Epidemiology, incidence and mortality

Colorectal cancer (CRC) is the third most prevalent cancer among men and the second most prevalent cancer among women, and is the second leading cause of cancer-related mortality[1]. It accounts for approximately 10% of all cancer diagnoses, with more than 1,900,000 new cases in 2020[2]. Males have substantially higher incidence and mortality rates than females, with age-standardized global CRC incidence rates per 100,000 equal to 23.4 for men, and 16.6 for women[1] (Figure 1.1).

Data from GLOBOCAN 2020 revealed significant geographical disparities in both incidence and mortality rates of CRC (Figure 1.1)[1]. Developed regions like Australia/New Zealand, Europe, and North America report the highest incidence, while Africa and South-Central Asia show the lowest. Interestingly, some traditionally low-risk areas, such as Spain and certain countries in Eastern Asia and Eastern Europe, have seen a sharp increase in incidence rates in the last years[3,4]. These geographical variations are often attributed to differences in diet, environmental factors, socioeconomic status, and screening practices[5–7].

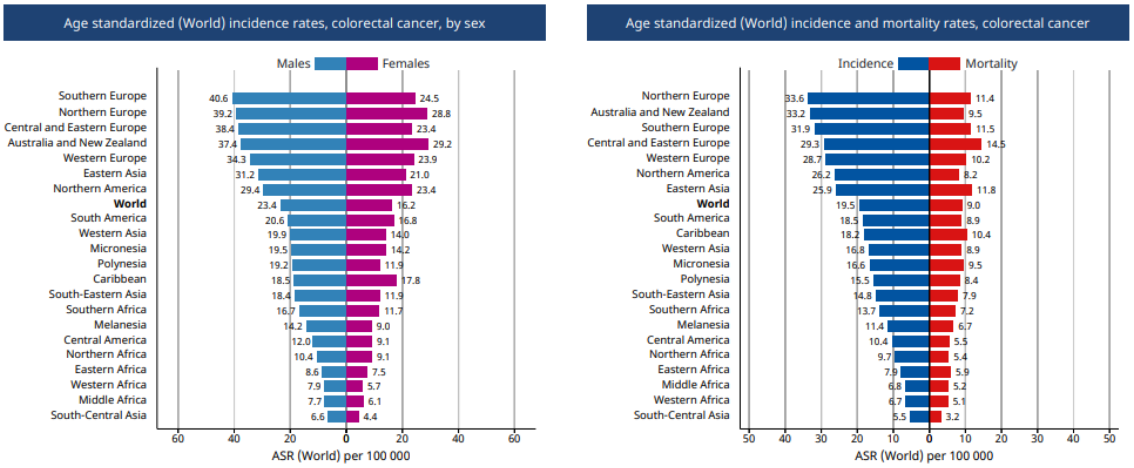


Figure 1.1. In the first panel, age-standardized incidence rates for CRC by region and sex. In the second panel, age-standardized incidence and mortality rates for CRC by region. Data source: GLOBOCAN 2020. Graph production: International Agency for Research on Cancer (IARC) (https://gco.iarc.fr/today). World Health Organization. Adapted from: https://gco.iarc.fr/today/data/factsheets/cancers/10_8_9-Colorectum-fact-sheet.pdf

Early detection is crucial for effective treatment. Survival rates for CRC are highly dependent on the stage at which the disease is diagnosed. Early-stage cases have a 90% five-year survival rate, which drops to 13% for late-stage diagnoses. The cumulative risk of dying from CRC between ages 0 and 74 is 0.65% for men and 0.45% for women[8,9]. Advances in screening methods, including colonoscopies and various fecal tests, have contributed to improved survival rates, despite an increase in incidence[10,11].

By 2030, it is projected that the global burden of CRC will surge by 60%, resulting in over 2.2 million new cases and 1.1 million deaths[10]. This increase is not solely due to genetic factors but is also strongly influenced by diet, lifestyle choices and environmental factors. It is expected to be the result of economic development and the transition of developing countries towards a "western lifestyle", characterized by high consumption of processed foods, red meat, and alcohol, as well as sedentary behavior and obesity[10].

Recent research has also highlighted the role of the gut microbiome in the prognosis and progression of CRC[12,13]. A diverse and balanced gut microbiome is essential for maintaining intestinal homeostasis, and disruptions in this microbial community, known as dysbiosis, have been found to be implicated in the initiation and progression of CRC[14]. Certain bacterial species, like *Fusobacterium nucleatum*, are found in higher abundances in CRC tumors and are believed to promote cancer through various mechanisms, including chronic inflammation, alteration of host metabolism, and direct interaction with cancer cells[15]. Moreover, the gut microbiome and the fecal microbiota transplantation (FMT) seem to influence the effectiveness of cancer treatments, including chemotherapy and immunotherapy[16,17]. Therefore, the gut microbiome may not only serve as a potential diagnostic and prognostic marker but it could also offer promising opportunities for targeted therapeutic interventions to improve CRC outcomes.

Making a comprehensive understanding of the epidemiological landscape and etiology of CRC is therefore essential for the development of effective strategies aimed at the prevention, early detection, and treatment of this increasingly prevalent disease.

1.2 Risk factors

Several factors contribute to the development of CRC, categorized into non-modifiable and modifiable determinants[11,18,19]. Non-modifiable risk factors, such as genetic predispositions and family history, are inherent and cannot be altered, whereas modifiable risk factors are predominantly associated with environmental influences and lifestyle behaviors.

1.2.1 Non-Modifiable risk factors

- *Age*

Age is one of the most important risk factor of sporadic CRC, with the risk of CRC significantly increasing after the age of 50. It was estimated that individuals over 65 have about three times higher risk of CRC than those aged 50-64 years and about 30 times the risk of those aged 25–49[20]. However, data from the United States shows that the incidence of CRC in individuals younger than 50 has been increasing, and this increase does not appear to be attributable to the enhancement of screening programs, as the disease is being diagnosed at later stages[21–23].

The reasons for this increase, though, are not clear and are probably multifactorial, due to a mix of genetic predisposition and changes in environmental and lifestyle factors[24].

- *Inflammatory Bowel Disease (IBD)*

Inflammatory bowel disease (IBD) is a group of autoimmune disorders that lead to chronic inflammation in the gastrointestinal tract and is associated with a higher risk of CRC[25,26]. The two most common forms of IBD are Crohn's disease (CD) and ulcerative colitis (UC).

Patients with IBD have a higher risk of CRC compared to the general population[26–29], with this risk increasing with increasing duration and severity of IBD condition[30–32]. Although IBD-associated CRCs are only 1-2% of total cases of CRC in the general population[33], mortality rates are higher[34,35] for both CD and UC.

The etiology of IBD is complex, involving a combination of genetic, environmental, and microbiological factors that trigger an overactive immune response, leading to persistent inflammation[36]. Several pathways have been proposed to explain the progression from inflammation to tumorigenesis in IDB-associated CRCs, which are distinct from those implicated in sporadic CRC. These pathways involve genetics and epigenetics alterations, crosstalk between immune system cells which leads to cytokine dysregulation and activation of specific signaling pathways, such as the nuclear factor kappa B (NF-κB) pathway, and environmental factors such as gut microbiota[37]. Growing evidence, in fact, shows the association between IBD and gut dysbiosis, with higher abundances of specific bacterial species observed not only in IBD patients but also in patients with CRC[38–40].

- *Personal and family history of colorectal cancer or colorectal polyps*

Individuals with a personal history of CRC are more likely to develop it in the future, especially if they were first diagnosed at a young age. A history of adenomatous polyps also increases the risk of CRC, particularly if the polyps are large (>1 cm) or multiple, or if they show high-grade dysplasia[41].

Family history is another important risk factor for CRC, with nearly one-third of CRC patients having a case of CRC in the family. Individuals with a first-degree relative (parent, sibling, or child) diagnosed with CRC have a 2 to 4 times higher risk of having the disease compared to the general population, and this risk is even higher if the relative was diagnosed before the age of 50 or if multiple family members have been affected by the disease[42,43]. Additionally, a higher risk of CRC has been observed also in case of family history in distant relatives or in case of family history of adenomas[42,44]. These clusters of CRC in families have been attributed not only to genetic predisposition but also to shared lifestyle factors and to a combination of both.

- *Hereditary syndromes*

While the majority of CRC cases are sporadic, it is estimated that approximately 5-10% are hereditary[45], meaning they are directly linked to inherited gene mutations.

Lynch syndrome (LS) is the most common hereditary CRC syndrome, accounting for about 3% of all CRC cases[45]. It is an autosomal dominant disorder caused by mutations in mismatch repair genes like *MLH1*, *MSH2*, *MSH6*, and *PMS2*. Individuals with LS have an elevated risk of developing CRC, as well as other cancers such as endometrial, ovarian, and gastric cancers[46]. Depending on the affected gene, people with LS have a lifetime risk of CRC up to 50%, a predominance of right-sided lesions and a younger age of onset, with the average age of first diagnosis being 48 years[47,48]. Through a meta-analysis of published literature, we investigated the relationship between obesity and CRC risk in LS patients. We found a sex/gender-specific association between obesity and CRC in patients with LS, with obese men having double the odds of developing CRC compared to non-obese men (Summary Relative Risk (SRR)=2.09; 95%CI: 1.23–3.55, I^2= 33%), while no differences were observed between obese and non-obese women. Moreover, patients with LS carrying *MLH1* mutation had a significant higher risk of CRC at increasing body-mass index (BMI) levels (SRR for an increase in 5 kg/m^2 BMI: 1.49; 95% CI: 1.11–1.99, I^2= 0%), while no differences according to BMI were observed in those carrying the *MSH2* mutation[49].

Familial Adenomatous Polyposis (FAP) is another autosomal dominant syndrome characterized by the development of hundreds to thousands of adenomatous polyps in the colon and rectum. FAP accounts for approximately 1% of CRC cases[45] and is caused by inherited mutations in the adenomatous polyposis coli (*APC*) gene[50]. Without colectomy, individuals with FAP have a high lifetime risk of CRC, with 87% of untreated FAP individuals developing CRC by age 45. Attenuated FAP (AFAP) also carries a high risk of CRC, but it is characterized by fewer adenomas and a later age of CRC onset than FAP[51].

Unlike FAP and Lynch syndrome, *MUTYH*-Associated Polyposis (MAP) is an autosomal recessive condition. It is caused by mutations in the *MUTYH* gene and is characterized by a predisposition to CRC and multiple adenomatous polyps, although fewer than in FAP (typically fewer than 500 adenomas)[50].

Other rare hereditary syndromes related to CRC are Peutz-Jeghers syndrome (characterized by mutations in the *LKB1/STK11* gene), juvenile polyposis syndrome, and serrated polyposis syndrome[52].

1.2.2 Modifiable risk factors

- *Overweight and obesity*

Obesity is a significant risk factor for several types of cancer, including CRC, with several studies highlighting its relationship with both the incidence and mortality of the disease[53]. This association is particularly strong for colon cancer compared to rectal cancer and appears to be gender specific[54]. Generally, an individual is considered obese if his/her Body Mass index (BMI) is greater than 30.

Compared with normal weight men, obese men have a significantly higher risk of CRC, with the association with colon being stronger than that with rectum. In women, the association between obesity and colon cancer is weaker than in men and is not significant with rectal cancer[54].

Moreover, obesity is associated with later-stage CRC and with lymph nodal metastases[55], while obesity before diagnosis is associated with a 22% higher risk of CRC-specific mortality and a 25% higher risk of all-causes mortality compared to non-obese[56].

Being overweight (BMI>25), not just obese, also contributes to CRC risk, even among those who are physically active[57,58]. Abdominal fat, as measured by waist circumference or waist-to-hip ratio, is strongly correlated with CRC risk, independent of overall body weight[59–61]. Timing of weight gain also matters; excess weight during adolescence and young adulthood appears to be an important risk factor for CRC in women, while for men, the risk increases later in life[62,63].

The biological mechanisms leading from obesity to carcinogenesis are very complex and not yet fully understood. One hypothesis is that the aberrant secretion of adipokines by adipose tissue may foster a pro-inflammatory environment[64]. Two such adipokines, leptin and adiponectin, have been the focus of several studies due to their contrasting roles in metabolic regulation and their potential impact on cancer development and progression.

Adiponectin is recognized for its anti-inflammatory and antiproliferative properties, playing a crucial role in glucose regulation and fatty acid breakdown[65]. Conversely, leptin acts as a pro-inflammatory, proliferative, and anti-apoptotic agent[62,66]. In the context of obesity, it is usually observed an increase in leptin levels coupled with a decrease in adiponectin levels[67,68]. The dysregulated secretion of these adipokines contributes to a pro-inflammatory environment, with overweight individuals characterized by elevated levels of inflammatory markers such as IL-6, TNFα, and C-reactive protein[69]. This state is associated with a higher prevalence of chronic inflammatory diseases, which could lead to DNA damage and, consequently, to carcinogenesis.

Obesity is also associated to metabolic syndrome, characterized by insulin resistance, elevated glucose and lipid levels, and hypertension. Both adiponectin and leptin have roles in these metabolic processes, as adiponectin improves insulin sensitivity, while leptin, in excess, can exacerbate insulin resistance[70].

- *Not being physically active*

Several studies have shown that physical activity is inversely associated with CRC risk[71–76] and with overall and CRC-specific mortality[77]. Specifically, the most physically active individuals have a 27% reduced risk of proximal CRC and a 26% reduced risk of distal CRC compared to the least active ones[75]. However, even transitioning from a sedentary life to an active one later in life showed a reduction in CRC risk[78].

- *Diet*

The role of diet on CRC risk has been widely investigated in the literature, with evidence suggesting a key role of dietary habits on the development of CRC[79–81]. This relationship could be indirect, through high calorie intakes leading to obesity and inflammation, or direct, through the effect of specific dietary components. Recent research suggests that the relationship between diet and CRC may also be mediated by the modulation of gut microbiota, which is significantly affected by dietary patterns and was found to be associated with both CRC prognosis and progression[82,83].

Diets rich in consumption of red meats, refined carbohydrates and processed sugar are correlated with inflammation and appears to be associated with increased risk of CRC, whereas diets rich in vegetables and fruits intake seem to have a protective effect[84]. However, studying the actual direct effect of diet and specific food components on the onset of CRC is very challenging.

- *High red and processed meat consumption*

In 2015, the International Agency for Research on Cancer (IARC) classified processed meat as "carcinogenic to humans" and red meat as "probably carcinogenic for humans", primarily based on evidence related to CRC risk[85]. Similarly, in 2018 the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) concluded that the evidence on consumption of processed meat was convincing, whereas the evidence for consumption of unprocessed red meat was classified as probable. They estimated an 18% increase in CRC risk for every 50 gr/day of processed meat and a 12% increase for every 100 grams/day of red meat. The risk is notably higher for colon cancer compared to rectal cancer and is more pronounced for processed meat than for red meat[86].

Various mechanisms have been proposed to explain this association, including an effect of specific constituents of meat and/or the formation of carcinogens during high-temperature[87].

Although this evidence is not entirely consistent across all studies and is based on observational studies, the prevailing scientific consensus supports limiting the consumption of red and processed meats as a preventive measure against CRC.

- *Low fruit and vegetables consumption*

The relationship between the consumption of fruits and vegetables and the risk of CRC has been a topic of extensive research, but the findings have been somewhat inconsistent. While many epidemiological studies have indicated a protective effect of a diet rich in fruits and vegetables against CRC[79–81], others have challenged this association.

A large prospective cohort study found no significant association between fruit and vegetable consumption and the incidence of either colon or rectal cancer[88]. However, a pooled analysis of 14 cohort studies found that consuming more than 800 gr of fruits and vegetables daily decreases the risk of distal colon cancer, although it had no effect on proximal colon cancer[89]. Another meta-

analysis of 19 cohort studies supported this, revealing a modest protective effect, particularly for distal colon cancers, when consumption exceeded 100 grams per day[90].

A vegetarian diet was also found associated with a reduced CRC risk compared to non-vegetarians[91,92]. However, two meta-analyses found no significant relationship for fruit consumption and a weak association for vegetable consumption when comparing the highest and lowest levels of intake[86,93].

- *Low whole grains/fiber consumption*

Dietary fiber is believed to reduce CRC risk due to factors like increased stool volume and faster transit time, which reduce exposure to carcinogens. This relationship appears to be stronger with whole grains[94], with two meta-analyses showing a reduction of 5% CRC risk for every 30 gr/day of whole-grain intake[86,93]. Additionally, a case-control nested within the European Prospective Investigation into Cancer and nutrition (EPIC) study showed an association between high consumption of whole grain and reduced rate of distal colon cancer, although no association with overall CRC risk[95].

Other studies found a protective effect of fiber intake on colonic adenomas[96–100], while others found none[101–105]. This inconsistency in results is most likely dependent on the type of fiber[106,107] and on the processed form[103] investigated in each study.

Another research suggested an interaction between the gut microbiota and fiber consumption, with a fiber-rich diet possibly mediating the effect of *Fusobacterium nucleatum* – a species known to promote tumorigenesis – on the risk of CRC.

- *Vitamin D*

Vitamin D (vitD) is a fat-soluble nutrient that plays a pivotal role in multiple physiological functions, including the regulation of calcium and phosphate, bone health, and immune system support.

It exists in two primary forms: vitamin D2 (ergocalciferol), which is derived from plant-based sources, and vitamin D3 (cholecalciferol), usually derived from animal sources.

Although vitD is primarily synthesized by exposing the skin to ultraviolet B rays from the sun, both forms can also be acquired through dietary sources that include fatty fish such as salmon, mackerel, and sardines, as well as cod liver oil, fortified foods like milk and cereals, beef liver, and egg yolks. Upon ingestion or synthesis, vitD is metabolized in the liver to 25-hydroxyvitamin D (25(OH)D), which is the primary circulating form of vitD and the best indicator of vitD status[108].

The role of vitD in cancer prevention and treatment has been a subject of increasing scientific interest. VitD regulates cell growth and differentiation, inhibits the uncontrolled proliferation of cancer cells, and promotes apoptosis, a form of programmed cell death that is often disrupted in cancer cells[109]. It also has anti-inflammatory properties, which are significant since chronic inflammation is a known risk factor for several types of cancer[110]. Moreover, vitD modulates the

immune system, enhancing the effects of monocytes and macrophages that fight pathogens and reducing the proliferation of pro-inflammatory cytokines[111].

According to the World Health Organization, CRC is the type of cancer with the greatest risk associated with poor vitD status[112]. Numerous studies have suggested an inverse relationship between vitD levels and CRC outcomes, although the evidence establishing causality remains inconclusive.

Several epidemiological studies and meta-analysis of prospective cohort studies, which are less prone to reverse causation bias, showed that low serum concentrations of 25(OH)D were associated with increased total cancer mortality[113–116]. Moreover, a meta-analysis of observational studies showed a significant inverse relationship between CRC risk and serum 25(OH)D[117], whereas meta-analyses of randomized trials have shown that vitD supplementation reduces overall mortality[118] and total cancer mortality[119–122].

A recent study including data from the UK Biobank cohort, involving 411,436 participants aged 40–69, revealed that during a median follow-up of 12.7 years, individuals with vitD deficiency (25(OH)D <30 nmol/L) had significantly higher overall and CRC-specific mortality compared to those with sufficient levels (25(OH)D ≥50 nmol/L). VitD insufficiency (30-50 nmol/L) was also associated with a 14% increase in CRC mortality, whereas vitD supplementation was associated with a 15% reduction in overall cancer mortality[123].

However, randomized clinical trials (RCTs) examining vitD supplementation for CRC prevention have yielded inconsistent results[124–126], likely due to methodological limitations such as sample size, dosing regimens, and follow-up duration.

Most of these studies, though, primarily focused on older populations, despite a concerning rise in CRC incidence among individuals under 50[127] - an age group also recently displaying higher rates of vitD deficiency (<20 ng/mL) compared to older individuals[128]. This evidence highlighted the necessity of exploring the role of vitD in early-onset CRC, particularly in the context of vitD deficiency status.

The prospective Nurses' Health Study II found that in young women vitD intake was associated with a lower incidence of early-onset CRC in a 24 years follow-up. This inverse relationship was significant for both dietary and supplemental sources of vitD and extended to the onset of adenomas and serrated polyps[129].

Another prospective study involving a Korean adult cohort found an inverse dose-response relationship between serum 25(OH)D levels and CRC risk in individuals younger than 50. In contrast, this association was less pronounced and occasionally non-significant in older individuals, suggesting that vitD deficiency may be a stronger risk factor for early-onset CRC[130].

The Vitamin D and Omega-3 Trial (VITAL), enrolling more than 25,000 participants, found no significant effect of vitD3 supplementation on CRC incidence. However, the majority of participants were not deficient in vitD at the study baseline.

Interestingly, they identified an interaction effect between vitD supplementation and body weight on both 25(OH)D levels and CRC outcomes[131]. At baseline, levels of 25(OH)D were inversely correlated with BMI. Furthermore, in the supplemented group, increases in 25(OH)D were lower at higher BMI. In the trial they also observed that supplementation was associated with a significant 24% lower incidence of cancer[132] and 42% lower mortality from cancer[133] among participants with normal body weight, but no reductions in overweight or obese individuals.
These findings suggest that the efficacy of vitD supplementation may vary based on individual characteristics, such as adiposity, and highlight the need of further investigation into personalized supplementation strategies for those at risk.

- *Smoking*

In 2009, the IARC stated that there is sufficient evidence to conclude that tobacco smoking is a causative factor for CRC[134]. A large number of studies have shown that cigarette smokers have a significantly higher risk of developing CRC[135], as well as all types of colonic polyps, particularly more advanced adenomas[136], but also hyperplastic polyps and those with dysplasia[137,138].
Moreover, smoking is significantly associated with a lower CRC-specific survival, particularly among current smokers[139,140]. We carried out a systematic review of the literature[141] which showed that quitting smoking at or around the time of diagnosis has a beneficial effect on the disease-specific survival of patients with gastrointestinal cancers, including CRC, compared to those who continue to smoke.

- *Alcohol*

Alcohol consumption has been identified as a significant risk factor for the development of many diseases, including CRC. Several studies and meta-analyses have consistently shown that moderate to heavy drinking significantly increases the risk of CRC[142–144]. Specifically, individuals who consume two to three drinks per day face a 21% higher risk, while those consuming four or more drinks daily see their risk escalate by 52%, compared to non-drinkers. However, light drinkers (up to one drink per day) did not show a significant increase in risk.

- *Gut microbiota*

The gut microbiota is a complex community of microorganisms, including bacteria, viruses, fungi, and other microbes, that inhabit the gastrointestinal tract. These microorganisms play a crucial role in various physiological processes such as digestion, nutrient absorption, vitamin synthesis, and immune system modulation. The gut microbiota of an individual is a dynamic entity that can be influenced by various factors like diet, lifestyle, and medication, and it plays a significant role in maintaining homeostasis in the body.

Conversely, the term "microbiome" denotes the aggregate genetic content of all microorganisms in a specific habitat, including the gut. This term extends beyond the microbiota to encompass their genetic material, genomes, and interactions with the environment, offering a holistic perspective on the functional potential of the microbial community.

The disruption of gut microbiota, known as *dysbiosis*, is associated with a number of diseases, including CRC[145]. Despite inter-individual variations in gut microbiota composition, two meta-analyses including case-control studies identified reproducible microbiome signatures in CRC, setting the basis for future diagnostic applications[146,147]. Among these, *Fusobacterium nucleatum* is known to be significantly enriched in patients with CRC. It also appears to contribute not only to CRC pathogenesis but also on the modulation of responses to therapeutic interventions like chemotherapy and immune checkpoint inhibitors[148]. Furthermore, intratumoral colonization by specific bacteria, predominantly of the *Fusobacterium* genus, were found to be associates with CRC progression and metastasis[149–152].

On the other hand, certain beneficial bacteria, acting as probiotics, are often depleted in CRC, suggesting an imbalance between pro-tumorigenic and anti-tumorigenic species[148].

Emerging evidence suggests a causal relationship between gut dysbiosis and CRC. Experimental models have shown that fecal microbiota from CRC patients can accelerate tumorigenesis in germ-free mice[153,154]. Moreover, specific bacterial strains have demonstrated tumorigenic potential in mono-colonization studies. The gut microbiota also interacts with environmental factors like diet and smoking to influence CRC risk. Various molecular mechanisms, including genotoxicity and inflammation, are implicated in this process.

Inflammation is a recognized risk factor for CRC, and gut dysbiosis plays a critical role in mediating inflammation in the gastrointestinal tract. Specific pathobionts, such as *F. nucleatum* and *Bacteroides fragilis*, are associated with colonic inflammation and CRC development. Mechanistically, these pathobionts activate inflammatory pathways like IL-17 and NF-κB, contributing to both CRC initiation and progression[148,155].

Metabolites produced by the gut microbiota, such as secondary bile acids and hydrogen sulfide, are also implicated in CRC risk. Elevated levels of these metabolites, often influenced by high-fat diets, promote tumorigenesis through mechanisms like oxidative DNA damage and NF-κB activation[156,157].

Integrated omics analyses have identified correlations between the gut microbiome and metabolome, offering further insights into CRC pathogenesis[158–161].

The current research frontier focuses on the precise and personalized modulation of the gut microbiota for clinical applications. The aim is to transition from broad-spectrum interventions like fecal microbiota transplantation to more targeted approaches, to leverage the gut microbiota for CRC prevention, treatment optimization, and early diagnosis.

1.3 Scope of the research

In this thesis project, we conducted a comprehensive investigation of the interplay between modifiable risk factors of CRC – specifically, vitD status/levels, circulating biomarkers of inflammation, obesity, and diet – and the gut microbiome. This multifaceted approach aimed to provide a comprehensive understanding of how the interrelationships between these factors can affect the prognosis and progression of CRC, accounting for possible sources of bias and confounding.

To accomplish this, we employed a multi-step approach.

First, we designed a case-control study comparing patients with CRC to healthy controls. In this study, we collected data on gut microbiome, a range of circulating biomarkers related to inflammation, adipocytes, and vitD, as well as information on diet and lifestyle.

Second, we conducted a systematic review of the literature on the relationship between vitD and microbiota both in humans, including both studies on healthy individuals and on individuals with dysbiosis conditions.

Lastly, we designed a RCT study involving survivors of CRC, who were randomly assigned to either placebo or daily vitD3 supplementation for one year. For these patients, we collected data on gut microbiome, circulating and vitD-related biomarkers, diet, and lifestyle at baseline and at the end of the treatment. Additionally, for a subgroup of these patients, we had gene expression profiles from tumor tissues for a set of immune-related genes from the Oncomine Immune Response Research Assay (OIRRA).

The statistical methodology employed for the analysis and integration of these diverse data sets was another major focus of this thesis project. Particularly, the microbiota data posed unique methodological challenges, as its high dimensionality, sparseness, and compositional nature precluded the use of conventional statistical methods. In light of these complexities, we employed advanced statistical techniques to ensure a more robust and insightful interpretation of the interrelationships among the variables under investigation.

Following is a more detailed summary of the research outline:

1. **Study Design: Case-Control**
    - **Primary Objective**: To identify the risk factors for CRC, including gut microbiome, diet, lifestyle, circulating biomarkers.
    - **Secondary Objectives**:
        - To investigate the relationship between diet/lifestyle and gut microbiome.
        - To assess the effect of high-risk diets and lifestyle on CRC.
        - To investigate the role of microbiome as a mediator of the effect of a high-risk diet on CRC risk.
        - To integrate the different risk factors to identify patterns in CRC patients.
        - To investigate the role of microbiome on CRC prognosis and progression.
2. **Study Design: Systematic Review on Human Studies**
    - **Primary Objective**: To synthesize existing literature on the relationship between the microbiota and vitD, focusing on both healthy subjects and those with dysbiosis conditions.
3. **Study Design: Randomized Controlled Trial (RCT)**
    - **Primary Objective**: To assess the impact of vitD supplementation on microbiome composition in CRC survivors.
    - **Secondary Objectives**:
        - To explore the role of gut microbiome as a mediator of vitD supplementation on 25(OH)D levels.
        - To investigate the role of circulating biomarkers, diet, obesity, and sex/gender and their interplay on the outcomes of vitD supplementation.
        - To estimate the associations between the immune-related transcriptomic profile and the investigated risk factors on the incidence of relapse and adenoma under vitD supplementation.

Methodological Challenges:

- To address the challenges related to the high dimensionality, sparseness, and compositional nature of omics data.
- To formulate effective strategies for the integration of multiomics data to obtain meaningful insights.
- To develop robust methodologies to estimate (causal) associations in the omics framework.

2. CASE-CONTROL STUDY: CRC PATIENTS AND HEALTHY CONTROLS

2.1 Rationale of the study

As we described, CRC arises from a multifaceted interaction of both non-modifiable and modifiable risk factors. These include obesity, dietary habits, lifestyle choices, inflammatory markers, and vitD levels. The gut microbiome has also emerged as a significant player in CRC, though determining if this relationship is causal remains intricate. These risk factors are intricately interconnected, each influencing and being influenced by the others.

Inflammation markers, which are associated with tumor initiation and progression[162–164], are modulated by adipose tissue through the release of adipokines, such as adiponectin. Adiponectin levels are usually reduced in obese individuals[165], with obesity being another important risk factor for CRC. Obesity usually leads to vitD deficiency[166], a condition linked to various chronic diseases, including cancer. Numerous meta-analyses of RCTs have shown that vitD supplementation is associated with decreased total and cancer mortality[118,167,168]. Additionally, some evidence suggests that vitD may modulate the gut microbiome[169–171], fostering an anti-inflammatory environment. The gut microbiome has been extensively studied in the context of CRC, leading to the identification of several predictive biomarkers for the disease[146,147]. The microbiome is also influenced by dietary habits, with diets high in red meat and low in fruits and vegetables being associated with a higher risk of CRC.

Despite the substantial body of evidence, the concurrent interplay among these risk factors and CRC are not yet fully understood. For this reason, the first step in our research was to design a case-control study, including both CRC patients and healthy controls, to delve deeper into the intricate network of risk factors involved in CRC etiology[172] (Figure 2.1).

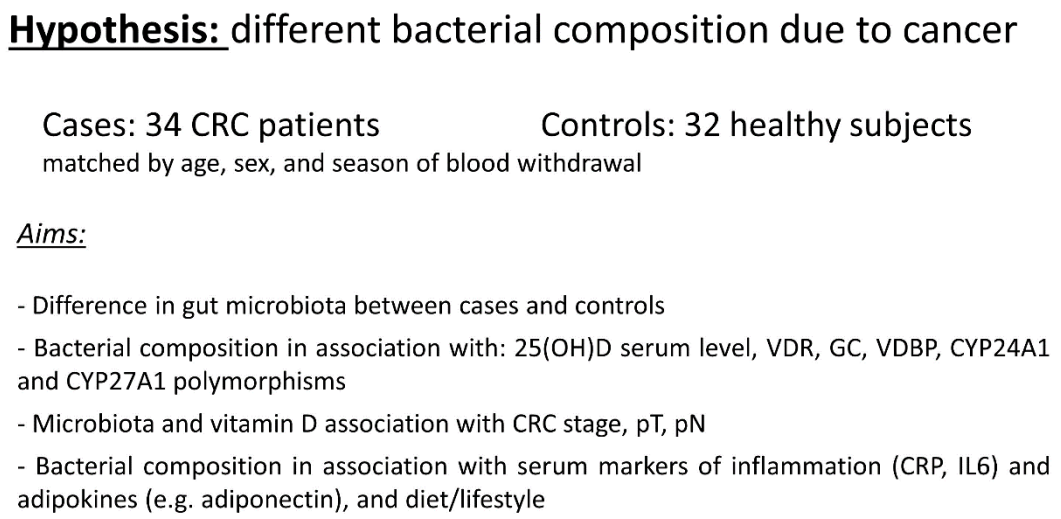


Figure 2.1. Graphical representation of the selected population for the case-control study. The diagram illustrates the division between cases (CRC patients) and controls (healthy individuals), as well as the primary and secondary objectives of the study.

2.2 Study design and participants

A total of 66 participants, including 34 CRC cases and 32 controls, were recruited and screened at the European Institute of Oncology (Milan, Italy). Cases with recent CRC diagnosis and aged between 35 and 70 years were enrolled before surgery or neoadjuvant treatment for resectable CRC.

Exclusion criteria for cases were:

- previous history of any cancer (5 years, other than cervical intraepithelial neoplasia or non-melanoma skin cancer);
- presence of mutations known to be associated to familial CRC (FAP, Lynch syndrome);
- current daily supplementation of vitD or calcitriol or high dose of calcium;
- history of malabsorption syndrome or any chronic IBD;
- use of antibiotics in the last 6 weeks, chronic alcoholism, and any medical condition that in the physician's opinion could potentially interfere with vitD metabolism.

Controls were subjects with a recent negative colonoscopy and no other relevant gastrointestinal disorder. Cases and controls were matched for age (±5 years) and season at blood collection (±2 months). However, 2 patients were lost after enrollment.

The study (IEO #118) was approved by the Institutional Review Board (European Institute of Oncology Ethical Committee), and all subjects gave their written informed consent according to ICH-Good Clinical Practice.

2.3 Materials and methods

2.3.1 Circulating Biomarkers

Morning fasting blood samples were collected at baseline. Serum was separated by 10 min of centrifugation at 1350× g and stored at −80 °C for subsequent biomarker quantification. Serum concentrations of 25(OH)D were measured by a chemiluminescence microparticle immunoassay (CMIA) designed for the automated instrument Architect (Abbott Diagnostics, Lake Forest, IL, USA). Due to high seasonal variability, different cut-off points were considered to define 25(OH)D deficiency in different seasons (<20 ng/mL in summer/autumn and <10 ng/mL in winter/spring). For the high-sensitivity C-reactive protein (hs-CRP) analysis, we employed a latex immunoturbidimetric high-sensitivity method on the same instrument. IGF-II was measured by sandwich ELISA from Mediagnost (Bensheim, Germany). IGFBP-3, IL-6, vitamin D binding protein (VDBP), leptin, and adiponectin were determined by ELISA (R&D Systems). Serum zonulin was determined using an ELISA kit from Elabscience (Wuhan, China). Subjects with non-detectable IL-6 levels were imputed with the lowest detectable value (3.13 pg/mL).

2.3.2 Single Nucleotide Polymorphism (SNPs) Analysis

Genomic DNA was extracted from whole blood specimens using a QIAamp DNA blood kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions on the automated platform "QIAcube" (Qiagen, Valencia, CA, USA), and quantified using NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA). DNA samples were genotyped for a comprehensive set of single nucleotide polymorphisms (SNPs). We analyzed Bsm1 (rs1544410), Taq1 (rs731236), Fok1 (rs228570), and Apa1 (rs7975232) in the VDR gene; 3 SNPs involved in vitD metabolism (CYP24A1-rs6013897, CYP27B1-rs10877012, CYP2R1-rs10741657); and rs2282679, rs7041, and rs4588 in the GC gene coding for the main transporter of vitD in the circulation. SNPs genotyping was performed by the TaqMan SNP Genotyping Assays using an ABI PRISM 7500 FAST Real-Time polymerase chain reaction (PCR) System (Thermo Fisher Scientific). Briefly, nearly 10 ng of DNA in 2 µL was added to a 10-µL reaction well together with 8 µL of reaction mix containing forward and reverse primers and 2 allele-specific fluorescent labelled probes (1 wild-type and 1 variant allele-specific). Control samples, representing a complete set of genotypes for all SNPs, were processed in each run. Hardy–Weinberg equilibrium (HW) for genotype frequencies was tested using a chi-squared test in controls.

2.3.3 Microbiota Analysis

Freshly voided stool samples were collected from controls and cases (before surgery, or any other treatment), and transported refrigerated to the laboratory within 6 hours from collection and immediately frozen at −80 ° C.

For metagenomic analysis, genomic bacterial DNA was isolated from feces of CRC patients and healthy controls with the G'NOME isolation kit (MP Biomedicals) following a published protocol[173]. The V5-V6 hypervariable regions of 16S rRNA gene were amplified and sequenced using the Illumina MiSeq platform, following library preparation and sequencing procedures previously described[174]. No evidence of batch effect related to the 2 sequencing runs was observed (intraclass correlation coefficient at phylum level: 0.98; 95% CI: 0.95–0.99).

Whole metagenome shotgun sequencing[175] was also applied on the same DNA samples. Metagenomic libraries were generated with a Nextera XT DNA Sample Prep Kit (Illumina, San Diego, CA, USA) and sequencing was carried out on the HiSeq2500 platform (Illumina) at a targeted depth of 5.0 Gb (100-bp paired end reads). Shotgun metagenomics sequencing samples were pre-processed as previously described[146].

2.3.4 Dietary Assessment

The dietary habits of participants at the time of enrollment were collected using a short questionnaire (Supplementary Table S1). Our questionnaire assessed the consumption of groups of food usually included in the Italian diet and that allowed to measure adherence to a Mediterranean diet. Participants were prompted to indicate their typical weekly intake for each food group by selecting an option from a five-tier scale, which ranged from 'never or seldom' to 'highly frequent.' To aid in precise reporting, a reference portion size was provided for each food category. The questionnaire was also adapted to estimate the dietary consumption of vitD, distinguishing between the consumption of fatty fish (rich in vitD) and the consumption of other fish.

To avoid the problem of sparse data, we created categories of similar food types by grouping the answers on the intake of the single types of foods belonging to the same category (such as sweets, any meat, diary products etc.). We then identified high-risk consumptions for each category following the WCRF recommendations for cancer prevention[176,177].

2.4 Statistical methods

2.4.1 Bioinformatic analysis for microbiome data (16S and shotgun)

The taxonomic annotation of raw Illumina MiSeq reads was performed using the pipeline BioMaS[178] . In particular: (i) the overlapping Paired End (PE) reads were merged into consensus sequences using pair-end read merger (PEAR)[179] and sequences shorter than 50nt were removed. Non-overlapping PE reads were further denoised by removing low-quality regions (quality-score threshold equal to 25) and discarding PE reads containing sequences shorter than 50nt by using Trim-Galore; (ii) both consensus and unmerged PE reads were mapped on the 11.5 release of the

RDP II database[180,181] using Bowtie2[182]; (iii) to obtain the taxonomic classification, mapping data were filtered according to two parameters: identity percentage (≥90%) and query coverage (≥ 70%). In particular, sequences matching on RDP II with an identity percentage of at least 97% were directed to species classification[183], while the others were classified at higher taxonomic levels. The NCBI taxonomy was used as reference taxonomy. Raw Illumina sequences in Operational Taxonomic Units (OTUs) were filtered by applying QIIME. In particular: (i) adaptor trimming: Illumina Nextera adaptor was removed by applying trim galore; (ii) PE reads were merged by applying PEAR; (iii) OTU definition was achieved by applying the QIIME open-picking procedure (reference database and taxonomy: greengenes 13.8); (iv) chimeric sequences were removed using Chimera-Slayer. The OTU table was re-generated by excluding chimeric OTUs and normalized by rarefaction. 12.5 million PE reads were produced in the 16S rRNA analysis. The mean number of PE reads per sample was about 187,000. About 97.5% of the sequences were taxonomically annotated using BioMaS. In particular, 89.5% and 68.2% of the sequences were taxonomically annotated at genus and species rank, respectively. In total, 744 operational taxonomic units (OTUs) were detected using 97% similarity threshold. The total number of sequences represents ranges from 115,437 to 204,779 sequences (Median 169,084, Average 168,265). Taxonomic data were summarized at phylum, class, order, family, genus and species level, normalized by applying DESeq2[184]. To better understand microbiota composition and its role in the CRC carcinogenesis, we also analyzed microbiota species and pathways from shotgun metagenomics. Shotgun data were missing for 2 CRC patients and 5 controls. Quantitative taxonomic profiling was performed using MetaPhlAn2[185], whereas HUMANn2[186] was used to profile pathway and gene family abundances. The generated profiles are available through the curated MetagenomicData R package.

2.4.2 Statistical analysis to investigate independent role of microbiome and interactive factors

Baseline demographic and epidemiological characteristics of cases and controls were summarized in terms of median and interquartile range for numerical variables and absolute frequency and percentages for categorical variables. Differences between groups were tested with Wilcoxon rank-sum test for numerical variables and with Chi-square test (Fisher exact test for sparse data) for categorical variables.

Circulating biomarkers were investigated as continuous variables and categorical variables. Because 25(OH)D levels significantly vary according to the season[187,188,188], we compared levels of 25(OH)D by using different cut-offs point for vitD deficiency depending on the season of blood collection: values below 10 ng/ml in winter, spring and initial summer (from November to June) and below 20 ng/ml in late summer and autumn[189]. The cut-off point for hs-CRP was chosen based on the median value of controls, the cut-off point for adiponectin was chosen based on first quartile among controls. The cut-off point chosen for IL-6 was based on the literature[190].

The WCRF score was built based on the cancer prevention lifestyle guidelines set by the WCRF/AICR[176,177]. Participants were deemed to adhere to the recommendations if their BMI was lower than 25, they reported engaging in regular physical activity, and they had a healthy diet, which was defined as a diet high in fruit and vegetable consumption or low in meat desserts, cakes, and pastries consumption. We also build two risk dietary scores: one was a categorical variable and the other was a continuous variable obtained as a linear combination of the food groups found to be significantly associated to CRC in the multivariable logistic model.

The first step in the investigation was to identify the taxa whose normalized abundances were significantly different between cases and controls. We did this by using Linear Discriminant Analysis (LDA) effect size (LEfSe)[191], which allows to identify the taxa statistically different among groups and to estimate their effect size. All p-values were set at 0.05, two-sided, adjusting for False Discovery Rate (FDR) using the Benjamini–Hochberg correction. Differences by cases and controls were also assessed using multivariable logistic regression models, which included CRC status as independent variable, the normalized abundance of each taxa as a covariate and that were adjusted for confounding factors.

After identifying the taxa significantly different between cases and controls, we employed different statistical methodologies – both supervised and unsupervised – to integrate the microbiota data with the other sources of data at our disposal.

We employed network analysis to visualize the interrelationships among the circulating biomarkers, the CRC-associated taxa, BMI and the continuous diet score obtained from multivariable logistic model. The network was based on the Spearman's rank-order correlation matrix, with the correlations computed on pairwise complete observations. The network visualization was generated using a force-directed ("spring") layout, which positions nodes based on their relationships, treating nodes as repelling objects and edges as connecting springs. To ensure the reliability of our network, we applied the BH procedure to control the FDR during multiple hypothesis testing. This ensured that only statistically significant relationships were retained in the final network visualization. The networks were generated using the “qgraph” package in R.

We further explored the potential relationship(s) between circulating biomarkers and gut microbiota by performing Canonical Correspondence Analysis (CCA)[192]. CCA is an unsupervised multivariate method usually employed in ecological studies to investigate the relationships between species abundances and sets of environmental variables (in our case, the circulating biomarkers). CCA, unlike ordinary correspondence analysis (CA), constrains the ordination of the biological variables (in our case, the species abundances) by the set of environmental variables so that the variation in species composition is explained in terms of environmental gradients. The method is useful to visualize the complex relationships in a reduced-dimensional space, where both the species and environmental variables are plotted. In our study, we used a triplot to display the

first two CCA components, incorporating both the taxa and circulating biomarkers for the case and control groups.

We integrated the microbiota data with the biomarkers, BMI and the diet score using the Data Integration Analysis for Biomarker Discovery (DIABLO)[193]. DIABLO is a framework of the mixOmics R package designed for supervised analysis of multiple datasets[194] [25]. We implemented a block sparse Partial Least Squares Discriminant Analysis (sPLS-DA) model to discriminate between the CRC patients and healthy controls, accounting for all the types of information at our disposal.

Partial Least Squares Discriminant Analysis (PLS-DA) is a statistical method used for classification and predictive modeling. It is an extension of Partial Least Squares (PLS) regression tailored for categorical response variables. In PLS-DA, the goal is to find the combination of variables that best separates different classes usually in a high-dimensional dataset. This is achieved by projecting the original variables onto a new set of variables, called latent variables or components, which are linear combinations of the original variables. These new components are constructed so that they maximize the covariance between the observed data and the class labels.

PLS-DA is particularly useful when dealing with "small n, large p" problems, where the number of observations (n) is small compared to the number of variables (p). It is also beneficial when the predictor variables are highly collinear.

In the context of block sPLS-DA, “blocks” refer to the different types of data included in the analysis. In our case, we have two blocks: the microbiota dataset and the clinical dataset (including the circulating biomarkers, BMI and the diet score). The “sparse” modality applied to the PLS-DA model aims to provide a selection of the variables that are most relevant for the discrimination between classes, so to reduce the dimensionality of the data. This is useful in high-dimensional settings like ours, where the number of variables is much higher than the number of the samples.

Each block is then analyzed simultaneously in a supervised manner through PLS-DA after the step of variable selection to identify the variables that mostly contribute to the discrimination of the groups. Results from each block is then integrated and visualized simultaneously (usually through heatmaps) allowing for a more comprehensive and holistic analysis.

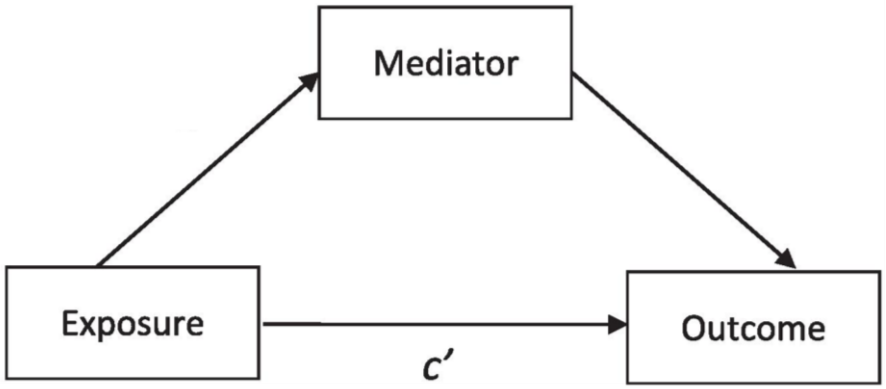
In our analysis, we estimated the first two components by tuning the optimal number of variables for each block through 10x100-fold cross-validation (CV). The final sPLS-DA model was then fit by using the tuning setting so defined.

A heatmap plot was then generated to graphically integrate the results on the first two components of each block.

2.4.3 Mediation analysis in the counterfactual framework

To estimate the role of microbiome as mediator of dietary factors on CRC development, we performed a mediation analysis based on a counterfactual framework approach[195,196].

In this context, a mediator is a variable that lies on the causal pathway between an exposure (in our case, diet) and an outcome (CRC). This implies that the mediator is influenced by the exposure and, in turn, has an effect on the outcome.



c'=vector of confounders.

Figure 2.2. DAG (Directed Acyclic Graph) showing the causal pathway in mediation analysis under the counterfactual framework.

The counterfactual framework allows for a formal, causal interpretation of this mediation effect by considering potential (or counterfactual) outcomes under different scenarios in which, for each level of the exposure, there is a potential outcome for the mediator, and for each level of the mediator, there is a potential outcome for the final outcome (Figure 2.2).

This happens through the decomposition of the Total Effect (TE) of the exposure on the outcome into the Natural Direct Effect (NDE) and the Natural Indirect Effect (NIE):

- **Natural Direct Effect (NDE)**: it quantifies how much of the TE goes directly from the exposure to the outcome without passing through the mediator. Specifically, it measures the effect of the exposure on the outcome while holding the mediator at the level it would have taken had the exposure not occurred.
- **Natural Indirect Effect (NIE)**: it quantifies how much of the TE occurs through the mediator. Specifically, it measures how the outcome would change if the mediator were modified to the level it would take under the exposure, while keeping the exposure constant.

By definition:

$$TE = NDE + NIE$$

Distinguishing between a *mediator* and a *confounder* is a crucial step in mediation analysis. A *mediator* is a variable that lies in the causal pathway between the exposure and the outcome, while a *confounder* is a variable that is associated with both the exposure and the outcome but is not part of the causal pathway.

Misclassification of both during the analysis can lead to different types of bias, affecting both the validity and interpretation of the results.

Directed Acyclic Graphs (DAGs) can be instrumental to ensure that all the hypothesis in a mediation analysis are correctly made. They provide a visual depiction of the causal structure among variables by representing variables as nodes and causal relationships as arrows (Figure 2.2), simplifying the task of identifying variables that lie in the causal pathway (mediators) and those that are outside of it (confounders).

In our analysis, we decomposed the TE of diet on CRC into a NDE and a NIE acting through the microbiota, considering alcohol consumption and physical activity as confounders of the pathway. In agreement with previous publications[197,198], the microbiota was evaluated considering the ratios of *Bifidobacteria* to *Escherichia* genera and Firmicutes to Bacteroides phyla, which have been shown to be modified with obesity and inflammation[199]. Both ratios were log-transformed, adding 1 unit, and modelled as mediators with linear regression models, whereas ORs with 95% CIs were obtained for NIE using unconditional logistic regression models adjusting for alcohol and physical activity. We also evaluated BMI as mediator of diet. The formulas used to calculate each effect are as follows.

Let Y be the binary outcome (CRC status), A the exposure (diet), M the mediator variable (microbiota) and C a set of multiple confounders (alcohol and physical activity).

The outcome Y was modelled using logistic regression as follows:

$$logit\{P(Y = 1|a, m, c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$

where *c* is the vector of confounders.

The mediator *M* was modelled using linear regression as follows:

$$M = \beta_0 + \beta_1 a + \beta_2' c$$

where *c* is the vector of confounders.

Provided that the assumption that the outcome *Y* is rare holds, we derived NDE and NIE on the Odds Ratio scale as following:

$$log\{OR^{NDE}\} = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' c + \theta_2 \sigma^2)\}(a - a^*) + 0.5\theta_3^2 \sigma^2 (a - a^2)$$

where σ^2 is the variance of the error term in the regression model on mediator *M*. For the binary exposure *A*, the two levels being compared are *a**=0 and *a*=1. Thus, the NDE provides an estimate of how much the outcome Y would change if the exposure A were set at level *a*=1 versus level *a**=0, having the mediator set to level it would naturally have in the absence of exposure.

$$log\{OR^{NIE}\} = \theta_2 \beta_1 (a - a^*)$$

Thus, the NIE estimates how much the outcome Y would change if the exposure A was set to level *a*=1, but the mediation *M* changes from the level it would have if *a**=0 to the level it would take if *a*=1.

2.4.4 Subgroup analyses to evaluate associations with prognostic factors and CRC recurrence

We evaluated differences in taxa abundances between CRC cases and healthy controls by pT, lymph-nodes involvement (pN) and early recurrence. Comparisons were made between healthy controls, pT1-2 and pT3-4 CRC patients, between healthy controls and CRC patients with or without involvement of lymph nodes and between healthy controls, CRC patients with no recurrence and CRC patients with recurrence. Non-parametric Kruskal-Wallis tests were used to assess differences by pT and pN. Time to relapse for cases was calculated from the day of diagnosis to the day of first cancer relapse (recurrence or adenoma) or last follow-up. We calculated Kaplan-Meier curves for disease free survival. Log-rank tests were used to investigate differences between survival curves. Cox proportional hazard models were employed to adjust for confounders. All statistical tests were two-sided.

All the statistical analyses were performed using the SAS statistical software (version 9.4) and R, version 3.4.

2.5 Results

2.5.1. Risk Factors and Circulating Biomarkers

Demographic, epidemiological, and clinical characteristics for cases and controls are summarized in Table 2.1.

Table 2.1. Descriptive characteristics of CRC patients (*n* = 34) and controls (*n* = 32).

		CRC (*N*, %)	Controls (*N*, %)	Total (*N*, %)	*p*-Value
Sex	Females	10 (29.4)	14 (43.7)	24 (36.4)	0.23
	Males	24 (70.6)	18 (56.3)	42 (63.6)	
Age	≤60 years	18 (52.9)	20 (62.5)	38 (57.6)	0.43
	>60 years	16 (47.1)	12 (36.5)	28 (42.4)	
BMI	≤25	11 (33.3)	20 (62.5)	31 (47.7)	0.02
	>25	22 (66.7)	12 (37.5)	34 (52.3)	
Regular physical activity	No	20 (58.8)	8 (25.0)	28 (42.4)	0.006
	Yes	14 (42.2)	24 (75.0)	38 (57.6)	
Regular alcohol consumption	No	5 (14.7)	15 (46.9)	20 (30.3)	0.005
	Yes	29 (85.3)	17 (53.1)	46 (69.7)	
Colon cancer family history	No	25 (73.5)	16 (50.0)	41 (62.1)	0.05
	Yes	9 (26.5)	16 (50.0)	25 (37.9)	
Smoking	Never	12 (35.3)	14 (75.0)	36 (54.5)	0.005
	Current	9 (26.5)	3 (9.4)	12 (18.2)	
	Former	13 (38.2)	5 (15.6)	18 (27.3)	

CRC, colorectal cancer; BMI, body mass index. *p*-values were obtained with chi-squared test.

Compared to healthy subjects, CRC cases were significantly more frequently obese (BMI>25; 66.7% vs 37.5% for cases and controls, respectively; p=0.02), regular alcohol consumers (85.3% vs. 53.1%; p=0.005) and smokers (64.8% vs. 25.0%; p=0.005), and engaged less frequently in regular physical activity (42.2% vs. 75.0%; p=0.006). However, no differences in terms of comorbidities (including diabetes and hypercholesterolemia) and current or recent use of drugs (including metformin, aspirin and statin) were observed.

Table 2.2. Descriptive statistics of circulating biomarkers in CRC patients and controls.

		CRC			Controls			
		Median	**Lower Quartile**	**Upper Quartile**	**Median**	**Lower Quartile**	**Upper Quartile**	***p*-Values**
25(OH)D (ng/mL) [1]		19.8	11.2	25.1	23.4	16.1	31.4	0.12
VDBP (µg/mL)		235	166	295	249	209	309.5	0.58
Zonulin (ng/mL)		119	74	178	109	54	315	0.94
IGFII (ng/mL)		671	578	769	695	614	806	0.41
IGFBP3 (µg/mL)		2.17	1.95	2.59	2.36	2.16	2.64	0.09
CRP (mg/dL)		0.23	0.12	0.39	0.10	0.05	0.20	0.01
Adiponectin (µg/mL)		4.87	3.41	9.48	7.77	6.23	12.39	0.03
Leptin (ng/mL)		6.56	4.25	14.15	6.71	5.19	15.43	0.67
			N. (%)			N. (%)		
Vitamin D (ng/mL) [1]	Sufficient		24 (70.6)			29 (90.6)		0.04
	Deficient		10 (29.4)			3 (9.4)		
hs-CRP (mg/dL) [2]	≤0.1		7 (20.6)			16 (50)		0.012
	>0.1		27 (79.4)			16 (50)		
Adiponectin (µg/mL) [3]	≤6		20 (58.8)			7 (21.9)		0.002
	>6		14 (41.2)			25 (78.1)		
IL-6 (pg/mL) [4]	≤4		25 (73.5)			30 (93.8)		0.03
	>4		9 (26.5)			2 (6.3)		

Differences between median values were assessed with Wilcoxon rank-sum tests and differences in frequencies with chi-squared tests.[1] Vitamin D deficiency is defined relative to the season: <20 ng/mL in summer/autumn and <10 ng/mL in winter/spring.[2] Cut-off point chosen on the basis of median value of controls. [3] Cut-off point chosen on the basis of first quartile among controls.[4] Cut-off point chosen on the basis of the literature.

Circulating biomarkers were also significantly different between cases and controls (Table 2.2). Cases had a significantly higher inflammation status, with higher hs-CRP (>0.1; 79.4% vs. 50.0% for cases and controls, respectively; p=0.012) and IL-6 (>4; 26.5% vs. 6.3%; p=0.03), and lower adiponectin (≤6; 58.8% vs. 21.90%; p=0.002). CRC patients were also significantly more often deficient in vitD than controls (29.4% vs. 9.4%; p=0.04), with vitD deficiency status defined considering different cut-offs based on the season of blood withdrawal.

These cut-offs were identified based on the existing literature (<20 ng/mL in summer/autumn and <10 ng/mL in winter/spring; see Statistical Methods) and consistent with what we observed in our sample (Table 1.3).

Table 2.3. Descriptive statistics of 25(OH)D levels (ng/mL) by season and CRC status

Season of blood withdrawal	Status	N	Median	Mean	Lower Quartile	Upper Quartile
Summer-Autumn	Cases	12	22.9	24.4	17.1	34.4
	Controls	14	30.8	31.4	25.5	35.3
Winter	Cases	8	23.5	20.9	15.9	27.4
	Controls	5	18.8	19.1	13.9	19.6
Spring	Cases	14	13.1	14.2	7.3	21.6
	Controls	13	17.9	16.6	12.7	22.6

In spring, levels of 25(OH)D were quite low (<20 ng/mL) in both cases and controls, especially in cases. However, 25(OH)D tended to be lower in cases compared to controls throughout the year, including in the summer-autumn period, when 25(OH)D levels were the highest and exceeded 20 ng/mL.

To better understand the relationship between CRC status and vitD, we also looked at VDR, GC (which encodes VDBP), and polymorphisms of vitD-metabolizing enzymes (CYP24A1, CYP24A1, and CYP24A1, as well as at the dietary intake.

ff FokI polymorphism and AA CYP24A1 polymorphism were significantly more frequent in CRC patients compared to controls (ff FokI : 20.6% vs. 3.1% for cases and controls, respectively, p=0.03; AA CYP24A1: 14.7% vs. 0%, p=0.02; Table 2.4), while no differences were observed for the other polymorphisms.

Table 2.4. Frequencies of CRC patients and controls by mutation status of polymorphisms.

VDR, GC, and CYP SNPs		CRC *n* = 34(%)	Controls *n* = 32 (%)	Total *n* = 66 (%)	*p*-Value
Fok1 rs2228570 (A > G) *FokI*	GG (*FF*) or GA (*Ff*)	27 (79.4)	31 (96.9)	58 (87.9)	0.03
(A = rare nucleotide)	AA (*ff*)	7 (20.6)	1 (3.1)	8 (12.1)	
Bsm1 rs1544410 (C > T) *BsmI*	CC (*bb*) or CT (*Bb*)	31 (91.2)	27 (84.4)	58 (87.9)	0.39
(T = rare nucleotide)	TT (*BB*)	3 (8.8)	5 (15.6)	8 (12.1)	
Taq1 rs731236 (A > G) *TaqI*	AA (*TT*) or AG (*Tt*)	32 (94)	27 (84)	58 (89)	0.20
(G = rare nucleotide)	GG (*tt*)	2 (6)	5 (16)	7 (11)	
Apa1 rs7975232 (C > A) *ApaI*	AA (*AA*) or AC (*Aa*)	27 (79.4)	25 (78.1)	52 (78.8)	0.9
(C = rare nucleotide)	CC (*aa*)	7 (20.6)	7 (21.9)	14 (21.2)	
GC rs2282679 (T > G)	TT or TG	31 (91.2)	31 (96.9)	62 (93.9)	0.33
(G = rare allele)	GG	3 (8.8)	1 (3.1)	4 (6.1)	
GC rs4588 (G > T)	GG or GT	31 (91.2)	31 (96.9)	62 (93.9)	0.33
(T = rare nucleotide)	TT	3 (8.8)	1 (3.1)	4 (6.1)	
GC rs7041 (A > C)	CC or CA	27 (79.4)	28 (87.5)	55 (83.3)	0.38
(A = rare nucleotide)	AA	7 (20.6)	4 (12.5)	11 (16.7)	
CYP24A1 rs6013897 (T > A)	TT or TA	29 (85.3)	32 (100)	61 (92.4)	0.02
(A = rare nucleotide)	AA	5 (14.7)	0 (0.0)	5 (7.6)	
CYP27B1 rs10877012 (G > T)	GG or GT	31 (91.2)	29 (90.6)	60 (90.9)	0.93
(T = rare nucleotide)	TT	3 (8.8)	3 (9.4)	6 (9.1)	
CYP2R1 rs10741657 (A > G)	GG or GA	31 (91.2)	32 (100)	63 (95.5)	0.09
(A = rare nucleotide)	AA	3 (8.8)	0 (0)	3 (4.5)	

VDR, vitamin D receptor; GC, Vitamin D Binding Protein gene; CYP, cytochrome P450; SNPs Single Nucleotide Polymorphism; *p*-values were obtained with chi-squared test and Fisher's exact test.

The dietary intake of vitD was quantified considering the level of consumption of fatty fish (intended as salmon, herring, mackerel), which is known to be rich in cholecalciferol. This data was collected through a short diet questionnaire that was administered to each patient at enrollment (see Statistical Methods; Supplementary Table S1).

Table 2.5. Descriptive statistics of 25(OH)D levels (ng/mL) by consumption of fatty fish (dose: 150 gr) and CRC status

Frequency of fatty fish consumption	Status	N	Median	Lower Quartile	Upper Quartile
Rarely (once/twice a month)	Cases	17	19.3	12.8	24.5
	Controls	10	18.5	13.0	23.4
Once a week	Cases	13	15.2	9.7	23.8
	Controls	11	24.5	14.2	30.0
Two/three times a week	Cases	4	31.1	17.7	34.4
	Controls	11	26.0	17.9	35.3

We found a dose–response trend of 25(OH)D with increasing consumption of fatty fish among controls (Table 2.5), whereas levels of vitD in cases remained low independently of the fatty fish consumption.

2.5.2. Microbiome Biomarkers and Functional Profiles

We performed shotgun metagenomic analysis and 16S sequencing to characterize the fecal microbiota in cases and controls (Figure 2.3). In line with the literature, we found significantly higher abundances of *Fusobacterium nucleatum*, *Escherichia coli*, *Parvimonas micra*, and *Bacteroides dorei* in CRC cases. Conversely, probiotic species, such as *Bifidobacterium longum* and *Bacteroides dorei*, were significantly more abundant in controls, even after adjusting for confounders such as age, smoking, and alcohol consumption (Figure 2.3b). In addition, significantly higher abundances of metabolic pathways associated with gluconeogenesis, putrefaction, and fermentation was detected in cases.

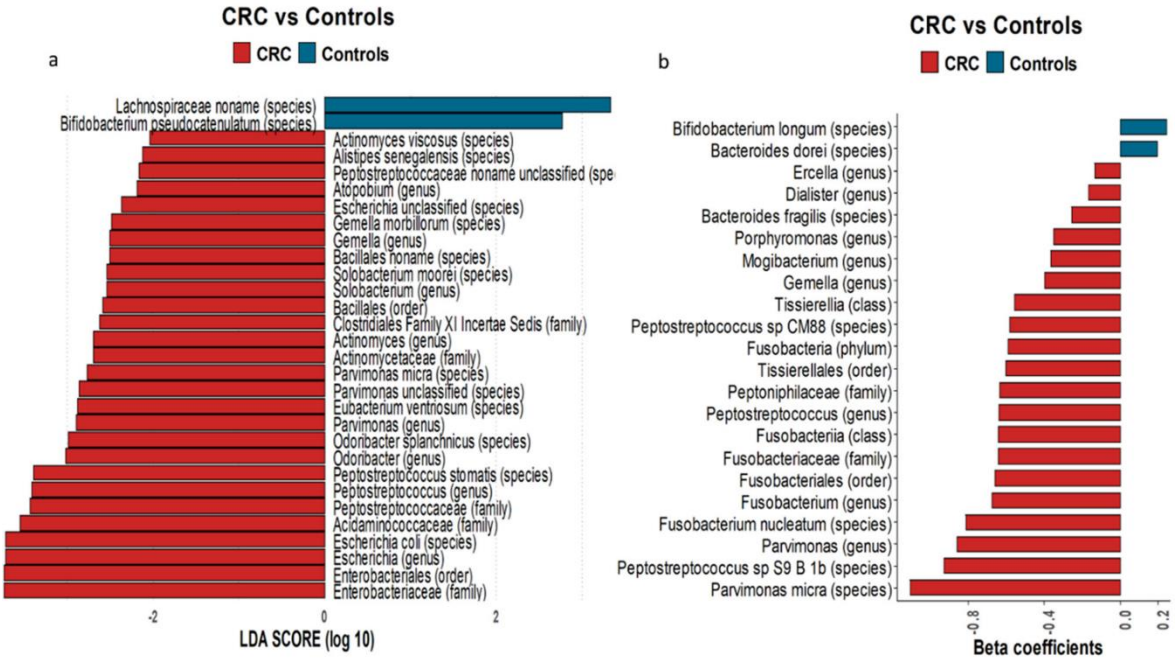


Figure 2.3. Microbiome composition in CRC patients and healthy controls. (**a**) Bar plot representing the result obtained by applying linear discriminant analysis effect size (LEfSe) on metabarcoding shotgun data. The bar length represents the linear discriminant analysis (LDA) score as a measure of the significant differences between the CRC (red) and control (blue) subjects (LDA score > 2). (**b**) Bar plot representing the taxa associated with CRC obtained through applying multivariable logistic model on metabarcoding 16S data, adjusted for age, smoking, and alcohol consumption. The bar length represents the significant beta-coefficient as a measure of the association with CRC (red) or healthy control (blue) subjects ($p < 0.05$).

2.5.3. Interplay between Vitamin D, Dietary Habits, and Microbiota in CRC

Overall, cases reported to have a diet with a significantly higher consumption of pasta, rice, and bread (food rich in carbohydrates) (67.6% vs. 28.1% for cases and controls, respectively, for once a

day; p=0.001) and with a significantly reduced consumption of fatty fish (11.8% vs. 34.4% for cases and controls, respectively, for 2–3 times a week; p=0.03; Table 2.6).

Table 2.6. Food intake frequencies for CRC patients and controls

Foods consumptions	Categories	CRC (N, %)	Controls (N, %)	Total (N, %)	P-value
Dairy products (milk/cheese/yogurt)	Not every day	5 (14.7)	8 (25.0)	13 (19.7)	0.29
	Once a day	29 (85.3)	24 (75.0)	53 (80.3)	
Pasta, rice and bread	Not every day	11 (32.4)	23 (71.9)	34 (51.5)	0.001
	Once a day	23 (67.6)	9 (28.1)	32 (48.5)	
Fruit and vegetables	Not every day	6 (17.6)	5 (15.6)	11 (16.7)	0.83
	Once a day	28 (82.4)	27 (84.4)	55 (83.3)	
Meat or processed meat	At most once a week	4 (11.8)	4 (12.5)	8 (12.1)	0.93
	At least twice a week	34 (88.2)	28 (87.5)	58 (87.9)	
Eggs	Rarely	9 (26.5)	9 (28.1)	18 (27.3)	0.49
	Once a week	15 (44.1)	17 (53.1)	32 (48.5)	
	2-3 times a week	10 (29.4)	6 (18.8)	16 (24.2)	
Fatty fish (salmon, herring, mackerel)	Rarely	17 (50.0)	10 (31.3)	27 (40.9)	0.03
	Once a week	13 (38.2)	11 (34.4)	24 (36.4)	
	2-3 times a week	4 (11.8)	11 (34.4)	15 (22.7)	
Fish (other)	Rarely	10 (29.4)	6 (18.8)	16 (24.2)	0.57
	Once a week	15 (44.1)	14 (43.8)	29 (43.9)	
	2-3 times a week	8 (23.5)	10 (31.3)	18 (27.3)	
Sweet/cakes/chocolate	≤1 a week	8 (23.5)	14 (43.8)	22 (33.3)	0.08
	≥2 times a week	26 (76.5)	18 (56.3)	44 (66.7)	

P-values were obtained with Chi-squared test. Meat consumption includes any type of meat (white and red), including processed meat and liver consumption. CRC=colorectal cancer.

Information on lifestyle (including smoking, alcohol consumption and physical activity) and diet were included as covariates in a multivariable logistic regression model to identify the risk factors significantly associated with CRC status. To avoid the problem of sparse data in the diet assessment, we grouped the answers on the consumption of every type of food enquired in the questionnaire into high-risk groups of foods. Based on the information collected with the questionnaire (Supplementary Table S1), we defined the following groups:

- High consumption of sweets and cakes, defined as: high ice cream (**Q22**≥3) or high chocolate (**Q23**≥3) or high sweets (**Q24**≥3).
- High consumption of cereals/carbohydrates, defined as: [high bread (**Q5**≥4) and pasta/rice (**Q3**≥4)] or [high crackers/breadsticks (**Q6**≥4) and pizza (**Q7**≥4)].

- Low consumption of fatty fish, defined as: low fish including salmon, herring, and mackerel (**Q12**<1).
- High consumption of meat, defined as: high meat (**Q9**≥3) and high processed meat (**Q11**≥3).
- Low consumption of fruit and vegetables, defined as: high soups (**Q4**≥4) or high vegetables (**Q18**≥4) or high fruit (**Q20**≥4).

High consumption in alcohol was significantly associated with CRC (Odds Ratio (OR)=6.20 [95% Confidence Interval (CI): 1.27-30.20]; p=0.024), whereas an inverse significant association was observed for regular physical activity (OR=4.31 [95%CI: 0.08-0.99]; p=0.049). Regarding diet, high consumptions of sweets and cakes and a diet low in fatty fish and high in cereals and carbohydrates were significantly associated with CRC status (high sweets and cakes: OR=4.31 [95%CI: 1.02-18.28]; p=0.048; low fatty fish and high cereals/carbohydrates: OR=5.88 [95%CI: 1.49-25.0; p=0.048; Table 2.7).

Based on these findings and using the data available to us, we built a dichotomous variable to indicate whether or not the patient was adhering to the WCRF/AICR guidelines for cancer prevention summarized in Figure 2.4.

Figure 2.4. Cancer prevention recommendations provided by WCRF/AICR. Source: WCRF. Adapted from: https://www.wcrf.org/diet-activity-and-cancer/cancer-prevention-recommendations/after-a-cancer-diagnosis-follow-our-recommendations-if-you-can/

A patient was considered to adhere to the recommendations if his/her BMI was lower than 25 or if he/she was physically active, and if his/her diet was high in fruit and vegetables or low in meat consumption or low in sweets.

This score was found to be significantly and strongly associated with CRC status, with higher odds of CRC for those not following the WCRF guidelines [OR=0.23; 95%CI: 0.08-0.67; p=0.007; Table 2.7).

Table 2.7. Multivariable logistic models: diet and risk factors associated with CRC.

	Lifestyle Risk Score	OR	Lower 95% CI	Upper 95% CI	*p*-Values
Risk factors	Regular physical activity	0.28	0.08	0.99	0.049
	Ever smoking	3.21	0.85	12.14	0.086
	High alcohol	6.20	1.27	30.20	0.024
Diet	High sweets and cakes	4.31	1.02	18.28	0.048
	Low fatty fish and highcereals/carbohydrates [2]	5.88	1.49	25.0	0.011
	WCRF score [1]	0.23	0.08	0.67	0.007

p-values were obtained from multivariable logistic models. [1] WCRF score: adherent if BMI < 25, high physical activity and a healthy diet (high consumption of fruit and vegetables, or low consumption of meat or low consumption of sweets, cakes, and pastries). [2] Low fatty fish and high cereals/carbohydrates: Low fatty fish (salmon, herring, mackerel) less than twice a week and high cereals (pasta, rice, and bread) at least once a day.

Next, we investigated whether specific species were enriched in subjects adhering to WCFR recommendations or to other dietary habits. Because both diet, lifestyle behaviors and microbiota are strongly dependent on socio-demographic characteristics and because we suppose a relationship between them and CRC, we investigated the association between microbiota and WRCR/diet through multivariable logistic regression, adjusting for CRC status, age and sex/gender. We found that a diet rich in fatty fish and with reduced cereals and carbohydrate consumption was significantly associated with higher abundances of *Lactobacillus* species (Figure 2.5a). On the other hand, an opposite diet – rich in carbohydrate and low in fatty fish – was significantly associated with higher abundances of *Clostridium ramosum* (belonging to the Firmicutes phylum).

Patients following WCRF guidelines showed an enrichment of *Bacteroides salyersiae*, which is a normal part of the gut flora and was found to be significantly more abundant in vegans[200], and in *Phascolarctobacterium succinatutens*, which is known to convert succinate into propionate[201]. Conversely, subjects not adhering to the recommendations showed higher abundances of species belonging to the oral microbiome, such as *Streptococcus sanguinis* and *Eubacterium infirmum* (Figure 2.5b).

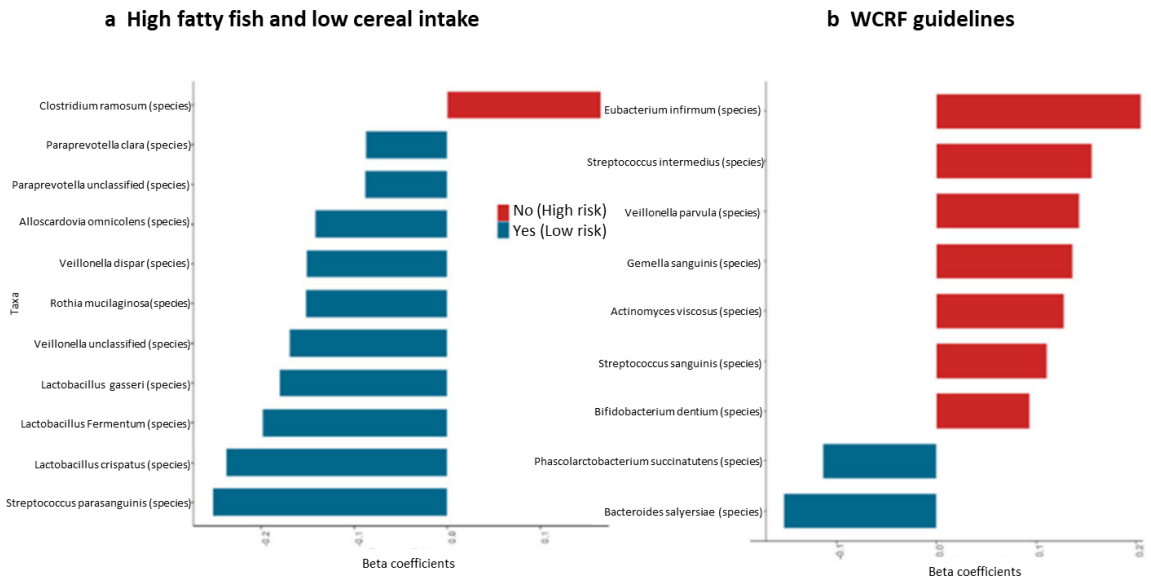


Figure 2.5. Species associated with diet and World Cancer Research Fund International (WCRF) guidelines. Results from logistic models of shotgun data. Bar plot representing the result obtained by applying multivariable logistic models, adjusted CRC status, age, and sex. The bar length represents the significant beta-coefficients of the models ($p < 0.05$). High-risk diet or not following the WCRF (red) and low-risk diet or following WCRF (blue). "Yes" indicates low-risk diet—"high fatty fish and low carbohydrates/cereals"; "No" indicates high-risk diet. "Yes" indicates those who follow WCRF guidelines; "No" indicates those who do not follow WCRF guidelines. (**a**) for high fatty fish and low cereals intake. (**b**) for adherence to WCRF guideline.

2.5.4. Microbiome-Mediated Diet Effect on CRC Risk

To understand if the relationship between a high-risk diet and CRC was – at least partially – mediated by the modulation of the gut microbiota, we conducted a mediation analysis under the counterfactual framework.

Through a DAG, we visually represented the causal pathway(s) we hypothesized to link the exposures (high-risk diets) to the outcome (CRC), both directly and indirectly through the microbiota, and identified the potential confounders affecting this pathway.

Diet information was summarized with a binary variable assessing a "low fatty fish and high carbohydrates/cereals" consumption, whereas the microbiota was summarized through the *Bifidobacteria/Escherichia* genera ratio, an indicator of "healthy" intestinal flora. We found that in subjects consuming a "low fatty fish and high carbohydrates/cereals" diet (which we found to be associated with CRC and is poor in vitD), the odds of CRC decreased at increasing levels of the log-transformed ratio of *Bifidobacteria* over *Escherichia* genera (Indirect Effect through microbiome: OR=0.31 (95% CI: 0.10–0.94), p=0.03), confirming a mediating effect of the microbiota. The direct effect of diet on CRC, independent of microbiota, was also statistically significant (p= 0.001), as well as the total effect (p=0.03) (Figure 2.6).

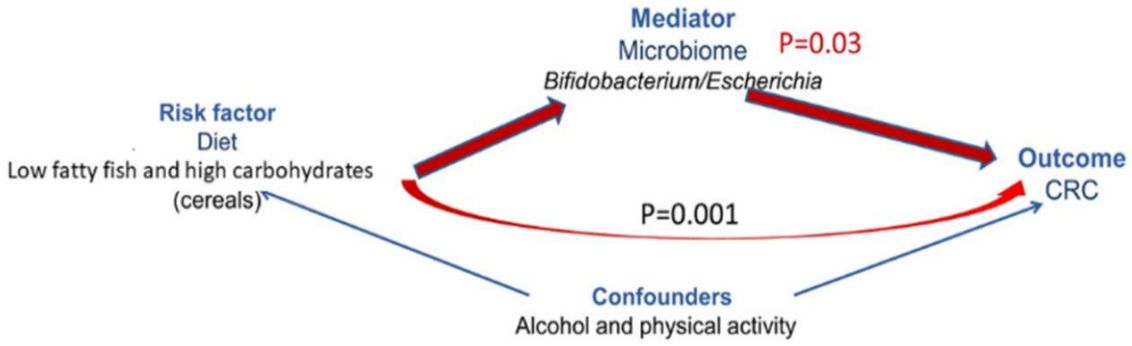


Figure 2.6. Direct acyclic graph of mediation model analyses. Gut microbiota as mediator of the effect of "low fatty fish and high carbohydrates/cereals" diet (exposure) on CRC risk (outcome). In red, natural indirect effect (NIE) and natural direct effect (NDE); in blue, the effect of confounders on exposure–outcome relationship. p-values were obtained from mediation analysis. The gut microbiota was summarized through *Bifidobacterium/Escherichia* ratio. Alcohol and physical activity were confounders. No exposure-mediator interaction was assumed, as it was not statistically significant in multivariable analysis.

We also considered body-mass index (BMI) as mediator of the effect of diet on CRC. However, the NIE was not statistically significant (p = 0.73), suggesting an independent role of diet and obesity as risk factors for CRC.

Firmicutes over Bacteroides ratio, another indicator of normal intestinal homeostasis associated with obesity and inflammation[202], was also evaluated as a mediator of the effect of diet, but no significant indirect effect was observed (NIE: OR = 0.96 (95% CI: 0.13–6.80; p = 0.97).

2.5.5. Integrative Data Analysis

The interrelationships between circulating biomarkers, vitD, BMI and diet were further investigated through a network analysis based on the Spearman correlation matrix. Only significant correlations were plotted after adjusting for the false discovery rate (FDR). The diet score was created as a linear combination of the regression coefficients and dietary groups associated with CRC in the multivariable analysis shown in Table 2.7.

BMI was directly and positively correlated with the diet score (Spearman correlation coefficient, R=0.41; p< 0.001), and the diet score was positively correlated with both hs-CRP (R=0.37; p=0.002) and IL-6 (R=0.27; p=0.027). IL-6 was inversely correlated with 25(OH)D (R=-0.27; p=0.0027) and positively correlated with the cluster of CRC-associated taxa, with a direct correlation with *F. nucleatum* (R=0.31; p=0.01) and *P.micra*. Adiponectin showed a significant inverse correlation with BMI (R=−051; p<0.001), while BMI was positively correlated with zonulin (R=0.36; p=0.004), a protein modulating the intestinal barrier function (Figure 2.7).

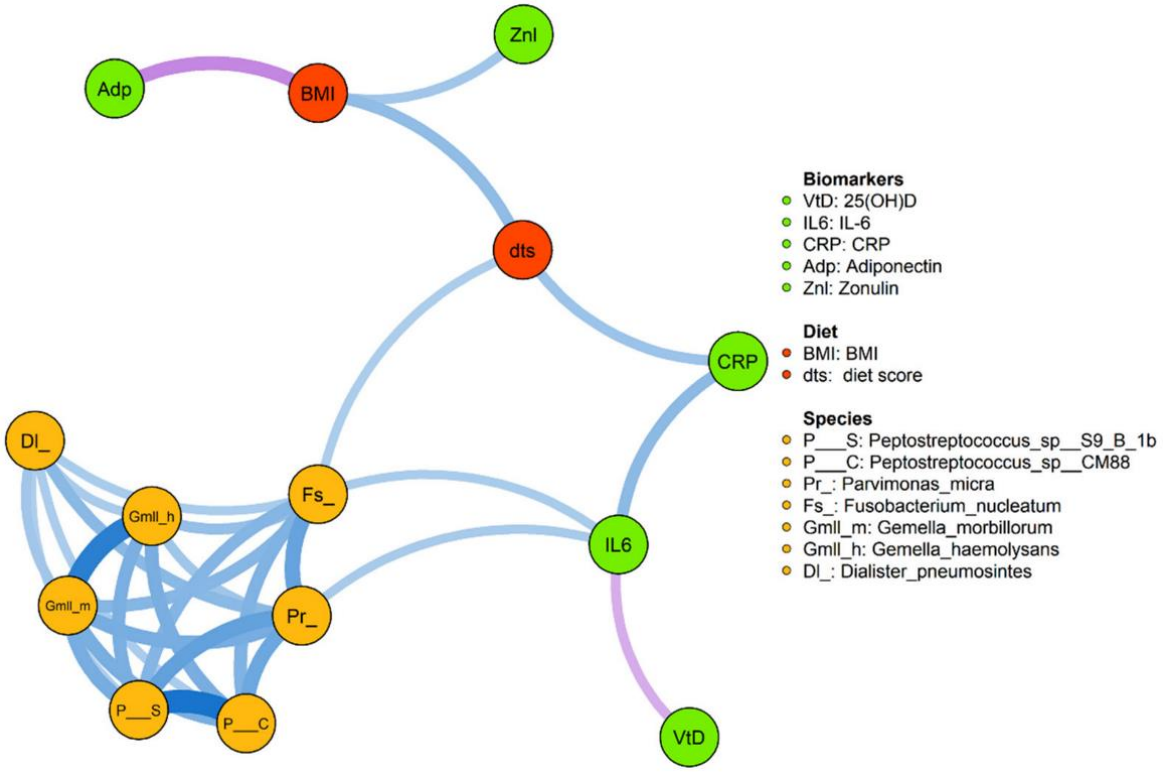


Figure 2.7. Correlation network analysis among circulating markers, BMI, dietary score, and CRC-associated species. The width of each edge corresponds to the absolute values of Spearman correlation coefficients and the transparency of edge represents the p-value after Benjamini-Hochberg (BH) correction. The line color indicates the direction of the correlation (blue for positive and violet for negative). Correlations with p-values less than 0.05 after BH correction are displayed. In the network, the CRC-associate taxa identified in the multivariable analyses in Figure 2.3b were included.

To better understand the correlation between circulating biomarker and the microbiota community in CRC cases and controls, we performed an unsupervised multivariate analysis based on canonical correspondence analysis (CCA). In the triplot in Figure 2.8 the first two components are shown. Each factor's weight is proportional to its arrow length.

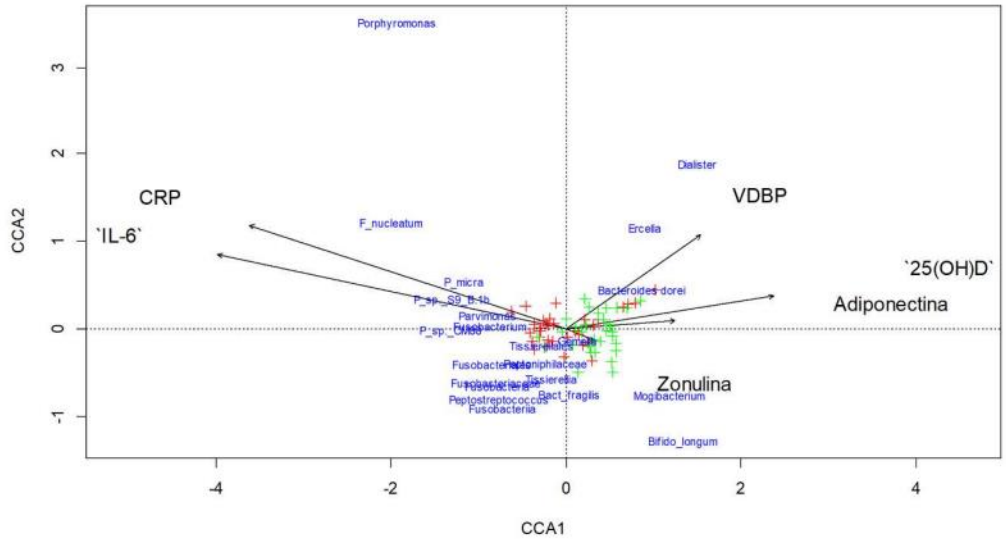


Figure 2.8. Triplot of Canonical Correspondence Analysis. Arrows indicate the direction and magnitude of the circulating biomarkers, systemic inflammatory markers correlation with bacterial community structures. Healthy controls are in green and colorectal cancer cases in red. All the CRC-associate taxa identified in the multivariable analyses in Figure 2.3b were included.

The first component of the CCA was the only one that was significantly associated with CRC status (p=0.001). This component correlated negatively with IL-6 and hs-CRP, with the negative side mostly characterizing CRC cases, and positively with 25(OH)D, VDBP, and adiponectin, with the positive size mostly characterizing healthy subjects (Figure 2.8). *F. nucleatum*, *Parvimonas micra*, and *Porphyromonas* positively correlated with hs-CRP and IL-6, whereas *Bacteroides dorei* and *Bifidobacterium longum* positively correlated with 25(OH)D and adiponectin. The second component was mostly characterized by the separation between Zonulin, on the negative axis, and the other biomarkes. Zonulin was positively – although weakly – correlated with the first component of CCA, characterized by healthy controls, and positively correlated with *Bifidobacterium longum* and *Mogibacterium*.

A supervised integrative analysis was eventually performed using block sPLS-DA based on the Data Integration Analysis for Biomarker Discovery (DIABLO) framework (see *Statistical Methods*).
This approach allowed us to distinguish between CRC patients and healthy controls by combining the "discriminative power" of each block of information (the 'clinical' block and the 'microbiota' block), after a step of feature selection. As a result, we were able to identify the taxa most discriminative of CRC status after accounting for the clinical condition and diet of each patient.
We compared the results of this model with the results of the model including only the microbiota data. As shown in Figure 2.9, the inclusion of clinical factors and diet led to a better discrimination between CRC patients and healthy controls compared to using microbiota data alone. Specifically,

we observed that CRC cases exhibited higher BMI, elevated levels of inflammation, increased diet scores, and greater abundances of several taxa like Fusobacteria, Tissierellales, and *Parvimonas micra*, especially in PT 3-4 patients. Conversely, healthy controls showed higher levels of vitD, adiponectin, zonulin, and Bacteroides, including *Bacteroides dorei*. These results regarding microbiota are consistent with those obtained from the multivariable analysis in Figure 2.3b, which was adjusted for confounders.

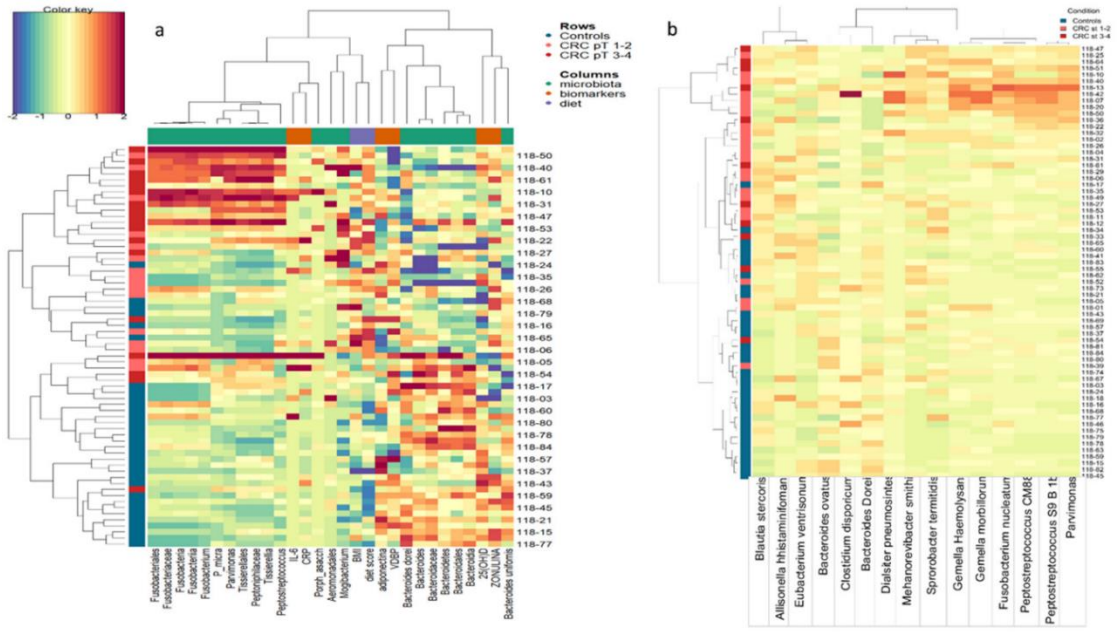


Figure 2.9. Block sPLS-DA for data integration. (**a**) Heatmap for data integration including the scaled variables selected from the first two components of each block. Plot generated by performing a block sparse partial least square-differential analysis (sPLS-DA) (10-fold cross-validation and 100 repeats) and selecting the most discriminative species, circulating biomarkers, BMI, and diet score. (**b**) Heatmap plot generated by performing a sparse partial least squares differential analysis (sPLS-DA) (10-fold cross-validation and 100 repeats) and selecting the most discriminative species, including the scaled abundances of taxa selected from the first two components.

2.5.6. Association of gut microbiota with CRC Prognostic Factors and Relapse

We also conducted an exploratory analysis to investigate associations between gut microbiota and tumor size (pathological T, pT), lymph node involvement (pathological N, pN) and early recurrence. *Parvimonas* and *Dialister* genera were very low among controls and the abundance increased among cases with worse prognosis (*pT3-4* and pN+; p<0.0001 and p=0.03, respectively; Figure 2.10).

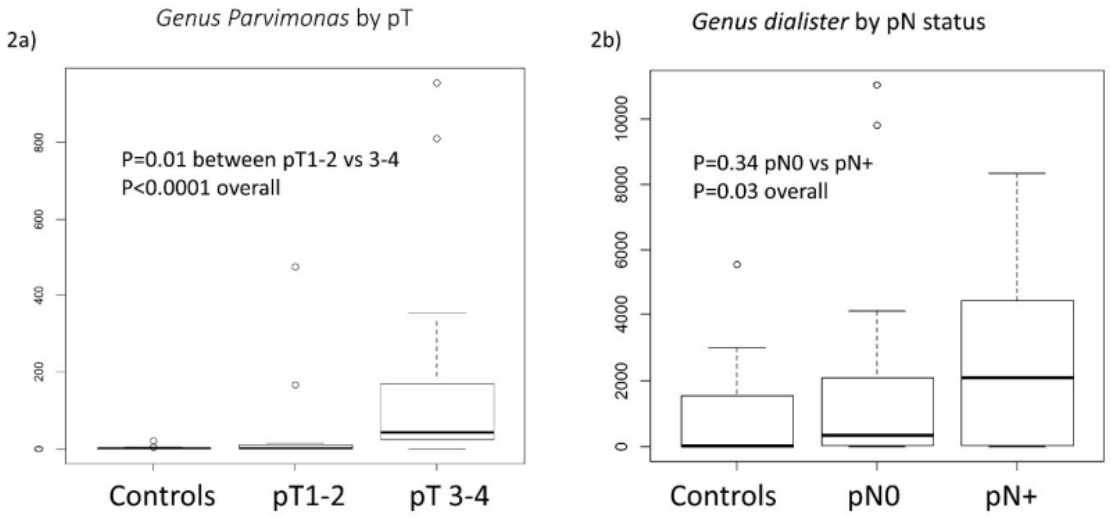


Figure 2.10. Association of microbiota with colorectal cancer (CRC) prognostic factors. On the left, boxplots of genus *Parvimonas* abundances by pT (controls, pT1-2 and pT3-4). On the right, boxplots of genus *Dialister* abundances by lymph-nodes involvement (controls, CRC patients without lymph-node (pN0) and CRC patients with lymph-node involvement (pN+)).

At 29-month median follow-up, we had four patients with cancer recurrence and five patients with adenomas.

Abundances of *F. nucleatum*, genus *Parvimonas*, and Tissierella class were significantly lower in healthy controls, higher in cases with no recurrence, and very high in cases with cancer recurrence (Kruskall–Wallis test: p=0.0002, p=0.0003, p=0.0006, respectively; Figure 2.11).

Upon categorizing *F. nucleatum* into high and low abundance based on the upper quartile of the distribution among cases, we found that patients with high *F. nucleatum* had a significant higher risk of recurrence (log-rank test: p=0.03; Figure 2.11d). This association remained statistically significant also in Cox proportional hazard model, after adjusting for lymph node involvement (p=0.02). Altogether, these data suggest that the microbiota composition plays a significant role throughout the tumorigenic process, including progression, and may influence prognosis.

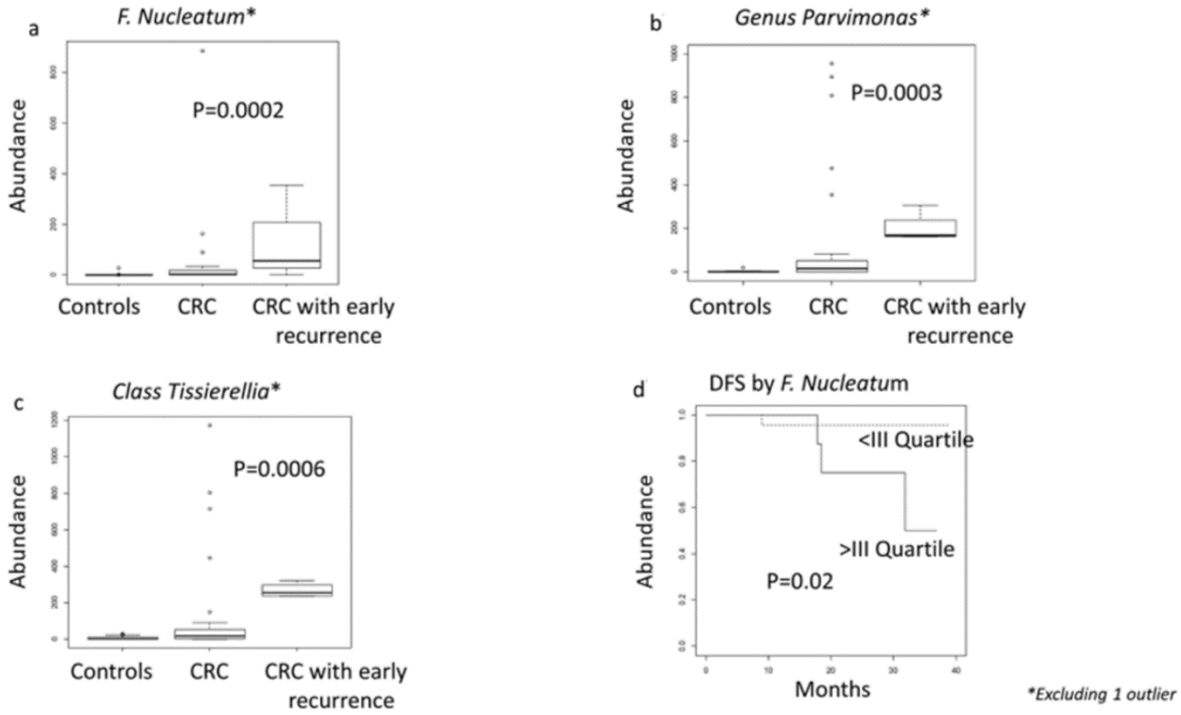


Figure 2.11. Box plots of taxa significantly associated with CRC status (**a**–**c**). Panel (**d**) Kaplan–Meier curves for disease-free survival p-value indicated in panel 8d was obtained from multivariate Cox regression models adjusting for lymph node status. * Excluding 1 outlier.

3. SYSTEMATIC REVIEW ON MICROBIOTA AND VITAMIN D IN HUMANS

3.1 Rationale of the review

VitD is a fat-soluble vitamin essential for maintaining bone health by facilitating the absorption of calcium and phosphate in the intestines. It exists in two primary forms, VitD2 and VitD3, which differ by a double bond between C22 and C23 and a methyl group at C24; this effects the bioavailability of the two forms, with vitD3 being more readily absorbed at the intestinal level.

The richest source of dietary intake of vitD is fatty fish, followed by egg yolk, liver, meat, and fortified dairy products[203,204]. However, the greatest amount of vitD is synthesized at skin level after UVB exposure.

Associations between vitD and immune modulation, cardiometabolic disorders, cancer risk and overall mortality have been reported in several epidemiological and clinical studies[205–208]. However, these encouraging findings have not been consistently replicated in interventional studies. To establish a causal link and advocate for vitD utilization, it is imperative to delve deeper into its non-skeletal roles, especially its interplay with the immune system.

A mechanism through which vitD exerts its effects could be by modulating the gut microbiota, whose alterations have been implicated in diseases such as cardiovascular disease, diabetes, and cancer.

VitD receptor (VDRs), which mediate the actions of active vitD3, are abundantly expressed in the gut and are instrumental in immune regulation and maintaining intestinal equilibrium[209,210]. While the direct impact of vitD on bacterial populations is not fully elucidated, few studies have shown its potential antimicrobial properties, both in vitro and in humans.

On the other hand, the microbiota is involved in maintaining the integrity of the intestinal mucosal barrier, protecting against pathogens, providing vitamins and metabolites, and shaping and regulating the immune response. This last function seems to be the key point linking the condition of dysbiosis with various diseases such as cancer, diabetes and cardiovascular or autoimmune diseases[211]. Consequently, a possible role of vitD in modulating the microbial composition of the gut could prove to be crucial in maintaining the function of the immune system and, consequently, human health[210].

In order to provide a deeper understanding of the association between vitD and gut microbiota alterations and/or composition, we conducted a systematic review of the existing literature of human studies[212] providing estimates of the association/correlation between microbiota and either vitD supplementation, vitD serum concentration (quantified with 25(OH)D) or dietary intake (Figure 3.1).

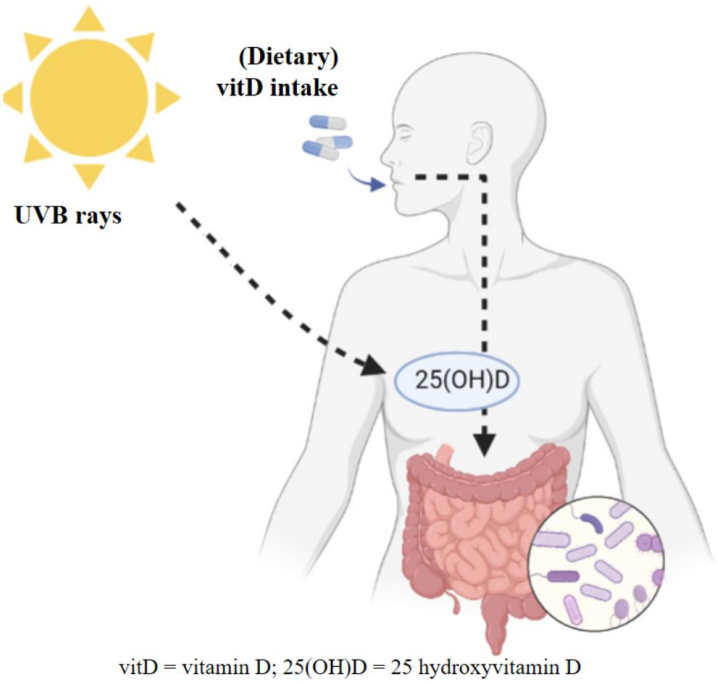


Figure 3.1. Graphical representation of the possible pathway going from vitD (either dietary, supplemented or synthesized following sun exposure) to the gut microbiota in humans.

3.2. Search Strategy

The literature search was carried out in compliance with PRISMA guidelines and extended up to January 2021. The databases consulted included PubMed, EMBASE, CINAHL, and Cochrane, and the search was limited to peer-reviewed articles published in English. The keywords "vitamin D," "vitamin D3," "cholecalciferol," and "25 Hydroxyvitamin D" were used in conjunction with "microbiota," "gut microbiota," "microbiome," or "dysbiosis."

The eligibility criteria were based on the PICOS framework[213]. We included human studies involving participants of all ages and health statuses (healthy and non-healthy). Both interventional and observational studies focusing on vitD supplementation, dietary vitD intake, and serum 25(OH)D levels were considered. All the selected studies had to provide at least one estimate of the association between the microbiota and vitD. Data on alpha diversity, beta diversity, species richness and the prevalence of bacterial taxa were collected. In interventional studies, outcomes regarding the microbiota were compared either to baseline (in case of single-arm trials) or to a control group (in case of two-arms studies).

We organized and reviewed the collected studies into separate categories based on:

- Scope of the analysis :
    - Effect of vitD supplementation on microbiota.

    - Association/correlation between vitD serum levels or vitD and microbiota.
- Health status of the study population:
    - Healthy individuals.
    - Individual with dysbiosis, pregnancies, obesity or diabetes.
- Microbiota sample type:
    - Stool samples.
    - Biopsy samples.

We provided the phylogenetic classification for all the taxa that were found to be significantly correlated/associated with vitD in the collected studies.

For the studies involving healthy participants and with microbiota evaluation from fecal samples, we calculated and displayed the percentage of taxa either increasing or decreasing in relation to vitD within each phylum over the total number of identified taxa. These were presented in frequency or mirror plots, differentiated by whether the study was focused on supplementation or on assessing serum vitD levels or dietary intake. Where applicable, plots at the family level were also included.

3.3 Results

The PRISMA flowchart in Figure 3.2 illustrates the study selection criteria for this systematic review.

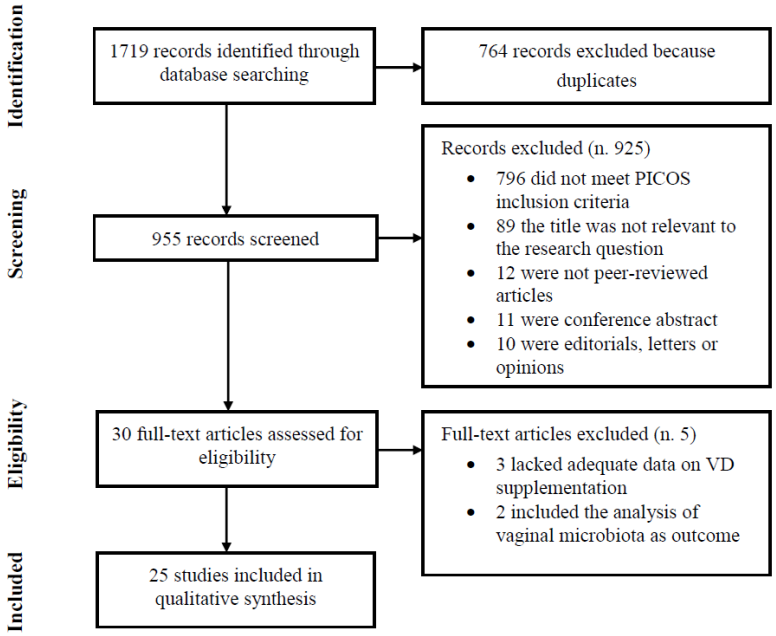


Figure 3.2: Flowchart of the section of the studies included in the systematic review.

Out of 955 publications, 25 were considered eligible for the analysis. As detailed in Table 3.2, the study designs included 14 interventional studies, comprising 7 RCTs[214–220], and 11 observational studies, of which 4 cohort studies[221–224], 6 cross-sectional studies[225–230], and 1 case-control study[231]. The majority of the studies were conducted in the United States and Europe, four in East Asia and the Middle East, two in Canada, and the remaining two in Africa and Brazil.

Most of the studies included healthy populations. However, specific cohorts were also investigated: five studies were conducted on pregnant women[218,221,232–234], four studies recruited individuals with ulcerative colitis and Crohn's disease[230,231,235,236], and individual studies focused on participants with multiple sclerosis[237], cystic fibrosis[217], HIV[216], and prediabetes[220]. Only one study included a population of overweight/obese individuals[216].

In all interventional studies, the duration of vitD supplementation was 14–15 weeks on average. Dosages varied, ranging from a minimum of 400 international units (IU) to a maximum of 10,000 IU per day. Eight [215,217,219,220,235–238] enrolled participants with serum levels of 25(OH)D less than 30 ng/mL, usually considered a cut-off of vitD deficiency status.

The analysis of microbiota was performed on stool samples, although four studies[216,230,231,239] also examined biopsies of gastrointestinal tissue. Methodological heterogeneity was observed in DNA extraction, amplification, and 16S rRNA sequencing procedures, largely attributable to the analysis

of different hypervariable regions. Only two studies employed shotgun metagenomic sequencing[225,231].

Table 3.1. Characteristics of selected studies.

Author, Year	Participants (n°)	Country, Cohort Name	Health status, Inclusion Criteria	Vitamin D Supplementation, Dietary Vitamin D Intakes or 25(OH)D Measure	Microbiota Analysis	Hypervariable Region of 16 sRNA Gene
Double-blind, randomized controlled trials						
Ciubotaru, 2015[220]	115	US	Prediabetes, AAM veteran, aged 35–85 years, BMI 28–39, serum 25(OH) D < 29 ng/mL	ARM1: 400 IU/week + placebo; ARM2: 400 IU/week + 50,000 UI/week for 12 weeks	Ion Torrent Personal Genome Machine	V4
Charoenngam,2020[219]	20	US	Healthy adults, serum 25(OH)D levels < 30 ng/mL	Three different arms: 600, 4000 or 10,000 UI/day for 8 weeks	uBiome Inc.	NR
Hjelmsø, 2020[218]	580	DK, COPSAC2010 cohort	Pregnant women, gestational age 24 weeks	2800 UI/day from 12 to 16 weeks	Illumina MiSeq	V4
Kanhere, 2018[217]	38	US	Patients with CF, age ≥ 18 year, no contraindication to oral high-dose vitamin D. Serum 25(OH)D level at baseline 37 ± 6 ng/mL	50,000 UI/week for 12 weeks	Illumina MiSeq	V4
Missailidis, 2019[216]	23	ET	ART-naïve HIV-positive individuals > 18 years, CD4+ T cells counts > 350 cells/mL, and plasma viral loads > 1000 copies/mL	5000 UI/day (plus phenylbutyrate suppl) for 16 weeks	Illumina MiSeq	V4
Naderpoor,2018[215]	26	AU	Healthy adults, serum 25(OH)D levels < 20 ng/mL, BMI > 25, stable weight	100,000 UI at baseline followed by 4000 UI/day for 16 weeks	Illumina MiSeq platform	V6-V8
Sordillo, 2016[214]	261	US	Pregnant women, aged 18–40 years, gestational age 10–18 weeks	Maternal vitDS with 400 or 4000 UI/day for 22–30 weeks	Pyrosequencing 16S RNA gene	V3–V5
Non-randomized interventional studies						
Bashir, 2016[239]	16	AT	Healthy adults, BMI 20–30, non-smokers	980 UI/Kg (week 1–4), 490 UI/Kg (week 5–8)	GS FLX	V1–V2
Bosman, 2019[240]	21	CA	Healthy adults, aged 19-40 years, Fitzpatrick skin types I-III	Average 1389 UI/day	Illumina MiSeq	V6–V8
Cantarel,2015[237]	15	US	Multiple Sclerosis/Healthy women, 25 (OH)D < 30 ng/mL, BMI 18-30	5000 UI/day for 90 days	PhyloChip Array	NR
Garg, 2018[235]	25	GB	25(OH)D < 50 ng/mL; For UC patients: partial Mayo index of ≤ 4, and stable therapy	40,000 UI/week for 8 weeks	Illumina MiSeq	V3- V4
Schäffler, 2018[236]	17	DE	CD/Healthy adults, serum 25(OH)D levels < 30 ng/ml	20,000 UI/day 1-3 + 20,000 UI every other day for 4 weeks	Illumina MiSeq	V3- V4
Singh, 2020[238]	80	QA	Healthy students, serum 25(OH)D levels < 30 ng/ml	50,000 UI/week for 12 weeks	Illumina MiSeq. Metagenomic analysis PICRUST	V3–V4
Tabatabaeizadeh, 2020[241]	50	IR	Healthy young girls, no history of diabetes, hypertension, or chronic disease	50,000 UI/week for 9 weeks	TaqMan assays	NR
Observational studies—Cohort						
Drall, 2020[224]	1157	CA, CHILD cohort	Pregnant women, gestational age 28 weeks	Maternal and infant vitDS of 400 UI/day	Illumina MiSeq platform	V4
Kassem,2020[223]	499	US, WHEALS cohort	Pregnant women, aged 21–49 years, gestational ages from 25 to 44 weeks	Maternal serum 25(OH)D and cord blood 25(OH)D levels	Illumina MiSeq	V4

Mandal, 2016[222]	60	NO, NoMIC cohort	Pregnant women	Dietary vitD intakes during 22 weeks of pregnancy: 3.13 µg/day (median)	Illumina MiSeq platform	V4
Talsness, 2017[221]	913	NL, KOALA cohort	Pregnant women, gestational age 14–18 weeks	Maternal vitDS: < or > 400 UI/day for 22–30 weeks. Infant vitDS: classified as yes or no	5'- nuclease technique	NR
Observational studies—Cross-sectional						
Jackson, 2018[229]	1724	GB, TwinsUK	Healthy adults	Use of vitDS	Illumina MiSeq technology	V4
Luthold, 2017[228]	150	BR, NutriHS Study	Healthy students, aged 18–40 years, undergraduate or graduate from nutrition colleges	Dietary vitD intakes (I: 1.66–4.95/ II: 4.97–7.18/ III: 7.56–39.87 µg/day)	Illumina MiSeq technology	V4
Seura, 2017[227]	28	JP	Healthy young women, aged 20–22 years, normal weight	Dietary vitD intakes (3.5 ± 2.5 µg/day)	T-RFLP method	NR
Soltys,2020[230]	87	SK	UC and CD	Serum 25(OH)D levels	Illumina MiSeq	V4
Thomas, 2020[226]	567	US	Healthy men (community-dwelling), aged 65 years or older	vitDS presents in 424 participants, not quantified. Measure of 25(OH)D; 1,25(OH)2D; 24,25(OH)2D	Illumina bcl2fastq	V4
Wu,2011[225]	98	US	Healthy volunteers, aged 2 to 50 years	Dietary vitD intakes	454/Roche pyrosequencing. Additional metagenomic analysis with shotgun method	V1-V2
Observational studies—Case-control						
Weng, 2019[231]	113	CN	Age >18 years and confirmed diagnosis of IBD (CD); BMI within the normal range and have not taken any antibiotics, probiotics, prebiotics or yogurt within the previous 4 weeks	Dietary vitD intakes	Illumina MiSeq System. Additional metagenomic analysis with shotgun method	V4

25(OH)D—25 hydroxyvitamin D; AAM—African-American men; ART- antiretroviral therapy; BMI—body mass index; CD—Crohn's disease; FDR—false discovery rate; HIV—human immunodeficiency virus; IBD—inflammatory bowel disease; NR—not reported; PICRUST- phylogenetic investigation of communities by reconstruction of unobserved states; rRNA—ribosomal RNA; SNPs—single nucleotide polymorphism; T-RFLP—terminal restriction fragment length polymorphism; UC—ulcerative colitis; UI—international units; vitD—vitamin D; VDR—vitamin D receptor; vitDS—vitamin D supplementation.

3.3.1 Alpha and beta diversity in relation to vitamin D

For the assessment of microbial diversity, various metrics were employed across the studies; however, the Shannon index and weighted Unifrac distance were most frequently used for evaluating alpha and beta diversity, respectively.

Table 3.2 summarizes, for each study, the main findings on changes in 25(OH)D levels and in alpha and beta diversity following vitD supplementation; Table 3.3 summarizes the findings on the correlation between dietary vitD intakes or serum 25(OH)D levels and alpha and beta diversity in the selected studies.

Overall, vitD supplementation increased serum levels of 25(OH)D. Regarding alpha diversity, only 7 studies have found an association with vitD, albeit with inconsistent findings (Table 3.2 and Table 3.3). Specifically, two interventional studies reported a decline in community richness following vitD supplementation[215,236]. In contrast, Bosman et al. found a significantly lower diversity and richness in the non-supplemented group compared to the supplemented. Singh et al. found a significant increase in alpha diversity following vitD supplementation, but only in observed OTUs and Chao1 indices, and not in the Shannon Index, both species richness and evenness. In cohorts of pregnant women, both maternal serum 25(OH)D or dietary vitD intake were significantly and inversely correlated with infant richness and diversity (Table 3).

With respect to beta diversity, significant alterations were found after vitD supplementation in biopsies of the upper gastrointestinal tract, but not in fecal samples or in lower gastrointestinal biopsies (Table 3.2). Some evidence showed significant shifts in bacterial community composition, related to both vitD supplementation and serum 25(OH)D levels (Table 3.2 and Table 3.3).

Table 3.2. Results of selected studies on alpha and beta diversity with vitamin D supplementation.

Author, Year	Comparison	Serum 25(OH) Levels	Sample	Alpha and Beta Diversity
Double-blind, randomized controlled trials				
Ciubotaru, 2015[220]	Serum 25(OH)D: quintiles	**Baseline**: 14 ± 6 ng/mL **Post**: 36 ± 24 ng/mL	Stool	**Alpha diversity**: NS **Beta diversity**: significant different bacterial composition found in Q1 vs. Q4 of 25(OH)D at genus and family levels
Charoenngam,2020[219]	Different doses of vitDS	**Baseline**: 16.9 ± 6.0 ng/mL; 20.3 ± 6.3 ng/mL; 18.5 ± 3.5 ng/mL **Post**: 20.0 ± 3.4 ng/mL; 39.0 ± 8.7 ng/mL; 67.3 ± 3.1 ng/ml	Stool	**Alpha diversity**: NS **Beta diversity**: NR
Hjelmsø, 2020[218]	Different doses of prenatal vitDS	Not reported	Infant stool	**Alpha diversity**: NS **Beta diversity**: NS
Kanhere, 2018[217]	Supplemented group vs placebo group in vit D insufficient at baseline	**Baseline**: vitD suff: 37 ± 6 ng/mL; vitD insuff, Pl.: 22 ± 6; vitD insuff, suppl.: 25 ± 5 ng/mL **Post**: vitD insuff, Pl.:25 ng/mL; vitD insuff, suppl.: 45 ng/ml	Stool	**Alpha diversity**: NS **Beta diversity**: significantly different composition at follow-up in the supplemented group compared to the placebo
Missailidis, 2019[216]	Supplemented group versus placebo group	**Baseline**: NR **Post**: NR	Mucosal gut biopsy	**Alpha diversity**: NS **Beta diversity**: NS
Naderpoor,2019[215]	Supplemented group versus placebo group	**Baseline**: vitD group 31.54 ± 4.4 vs. Pl 31.07 ± 4.1 nmol/L **Post**: vitD group 91.14 ± 25.8 vs. Pl 31.58 ± 14.11 nmol/L	Stool	**Alpha diversity**: significant reduction in richness at follow-up in the supplemented group **Beta diversity**: significant difference in composition between groups at follow-up at the genus level
Sordillo,2016[214]	Maternal vitDS Umbilical cordon 25(OH)D levels	**Baseline**: 22.7 ± 11.9 ng/ml	Stool	**Alpha diversity**: NS **Beta diversity**: NR
Non-randomized interventional trials				
Bashir, 2016[239]	Post- versus pre-supplementation	**Baseline**: 22.3 ± 13.1 ng/mL **Post**: 55.2 ± 13.3 ng/ml	Biopsy and stool	**Alpha diversity**: significant increased richness in GA **Beta diversity**: significant change in composition only in upper GI tract
Bosman, 2019[240]	Prior vitD supplemented group (vitDS+) vs prior non-vitD supplemented group (S-) before UVB exposure	**Baseline**: NR **Post**: NR	Stool	**Alpha diversity:** vitDS- showed significantly lower diversity and richness before UVB exposure than vitDS+ **Beta diversity:** NR
Cantarel,2015[237]	Post- versus pre-supplementation in healthy controls and in patients with multiple sclerosis	**Baseline**: 23.2 ± 5.7 ng/mL in the HCs; 25.9 ± 4.4 ng/mL in MS **Post**: 59.8 ± 11.7 ng/mL in the HCs; 55.6 ± 17.0 ng/mL in MSs	Stool	**Alpha diversity:**NS **Beta diversity:** NS
Garg, 2018[235]	Post versus pre-supplementation	**Baseline**: 34 (range 12–49) nmol/L **Post**: 111 (range 71–158) nmol/l	Stool	**Alpha diversity**: NS **Beta diversity**: NS
Schäffler, 2018[236]	Post versus pre-supplementation in healthy controls and in patients with CD	**Baseline**: in CD 39.7 ± 23 nmol/L, in HC 29.6 ± 6.3 nmol/L **Post**: in CD 121.4 ± 43.2 nmol/L, in HC 143.0 ± 25.2 nmol/L	Stool	**Alpha diversity**: In HC, NS; in CD taxa significantly decreased after vitDS. **Beta diversity**: NS
Singh, 2020[238]	Post- versus pre-supplementation	**Baseline**: 11.03 ± 0.51 ng/mL **Post**: 34.37 ± 1.47 ng/mL	Stool	**Alpha diversity**: Significant increase in observed OTUs

				and Chao1 indices, no difference in Shannon Index. **Beta diversity**: significant difference in composition between vs pro supplementation
Tabatabaeizadeh, 2020[241]	Post- versus pre-supplementation	**Baseline**: 11 ± 9 ng/mL **Post**: 40 ± 17 ng/mL	Stool	**Alpha diversity**: NR **Beta diversity:** NR
Observational studies—Cohort				
Drall, 2020[224]	Pre vs post maternal vitDS and Infant vitDS	**Baseline**: NR **Post**: NR	Stool	**Alpha diversity**: NR **Beta diversity:** NR
Talsness, 2017[221]	Comparisons of 3 levels of maternal vitDS; Infant vitDS vs non-infant vitDS	**Baseline**: 44.3 ± 18.3 nmol/L **Post**: NR	Stool	NR

25(OH)D—25 hydroxyvitamin D; CD—Crohn's disease; FDR—false discovery rate; HC—healthy controls; ND-UVB—narrow-band ultraviolet-B; NR—not reported; NS –not significant; SNPs—single nucleotide polymor-phisms; Pl—placebo; UC—ulcerative colitis; vitD—vitamin D; VDR—vitamin D receptor; vitDS—vitamin D supple-mentation; OTUs: operational taxonomic units.

Table 3.3. Results on alpha and beta diversity with vitamin D Intake or Serum Concentrations

Author, year	Comparison	Serum 25(OH) levels	Sample	Alfa and beta diversity
Observational studies—Cohort				
Kassem,2020[223]	Maternal serum 25(OH)D levels Umbilical cord blood 25(OH)D levels	**Baseline maternal serum 25(OH)D**: 25.04 ± 11.62 ng/mL **Baseline umbilical cord blood 25(OH)D levels:** 10.88 ± 6.77 ng/mL	Stool	**Alpha diversity**: Prenatal 25(OH)D level significantly associated with decreased infant richness and diversity at 1 month; cord 25(OH)D level was positively associated with infant gut evenness in White women and negatively associated with infant evenness at 6 months. **Beta diversity**: both prenatal and cord 25(OH)D were significantly associated with 1-month composition
Mandal, 2016[222]	Dietary maternal vitD intakes	**Baseline**: Not reported **Post**: Not reported		**Alpha diversity**: vitD intake was significantly and inversely associated to whole tree phylogenetic and Shannon diversity **Beta diversity**: NS
Observational studies—Cross-sectional				
Jackson, 2018[229]	Intake of vitD supplements: yes versus no	NR	Stool	**Alpha diversity**: NS **Beta diversity**: NS
Luthold, 2017[228]	Dietary vitD intakes: high versus low-tertile Serum 25(OH)D leves: high- versus low tertile	**Baseline**: 23.9 ± 9.7 ng/mL	Stool	**Alpha diversity**: NR **Beta diversity**: NR
Seura, 2017[227]	Dietary vitD intakes	**Baseline**: NR	Stool	**Alpha diversity**: NR **Beta diversity**: NR
Soltys,2020[230]	Serum vitD levels in patients with UC and CD	**Baseline**: in winter/spring 25.05 ng/mL and summer/autumn period 37.26 ng/ml	Biopsy and stool	**Alpha diversity**: NR **Beta diversity**: NS
Thomas, 2020[226]	vitD metabolites	25(OH)D (34.2 ng/mL), 1,25(OH)2D (56 pg/mL), and 24,25(OH)2D (3.2 ng/mL)	Stool	**Alpha diversity**: 1,25(OH)2D, active ratio and catabolism ratio are positively and significantly associated with diversity. **Beta diversity**: $1,25(OH)_2D$, $24,25(OH)_2D$, activation ratio, catabolism ratio significantly define clusters of microbial composition
Wu,2011[225]	Dietary vitD intakes	**Baseline**: NR	Stool	**Alpha diversity**: NR **Beta diversity**: NR
Observational studies—Case-control				
Weng, 2019[231]	Dietary vitD intakes in healthy controls and patients with UC and CD	**Baseline**: NR	Biopsy and stool	**Alpha diversity**: NR in UC **Beta diversity**: NR

25(OH)D—25 hydroxyvitamin D; CD—Crohn's disease; FDR—false discovery rate; HC—healthy controls; ND-UVB—narrow-band ultraviolet-B; NR—not reported; NS—not significant; SNPs—single nucleotide polymorphisms; Pl—placebo; UC—ulcerative colitis; vitD—vitamin D; VDR—vitamin D receptor.

3.3.2 Distribution of Taxa at Phylum Level

Firmicutes and Bacteroidetes were the phyla that increased the most frequently following vitD supplementation, followed by Proteobacteria and Actinobacteria. (Figure 3.3). Similar trends were observed in relation to vitD dietary intake or serum concentrations levels. Because we found contradictory results for Firmicutes, which either increased or decreased in relation to vitD, we carried out a more granular inspection at the family level (Figure 3.4). Taxa from Veillonellaceae and Oscillospiraceae families decreased more frequently at increasing levels of 25(OH)D or following vitD supplementation. For Lachnospiraceae family we did not identify a clear trend in relation to vitD, therefore we further investigated the differences at the genus level. We found that the genus *Blautia* mostly decreased after vitD supplementation and at increasing 25(OH)D. However, Thomas et al. found a significant positive correlation between the species *Blautia obeum*, a member of the genus *Blautia*, and the active form of vitD, 1,25(OH)D. The genus *Roseburia* was also inversely correlated with vitD levels, although Singh et al. found a positive correlation in the group of non-responders to vitD supplementation (25(OH)D: < 20 ng/mL at follow-up). Conversely, *Funicanibacter*, *Lachnospira* and *Lachnobacterium* were significantly more abundant in vitD - supplemented individuals.

Conflicting results were found for the genus *Coprococcus*, which was significantly more abundant in the supplemented cohort of untreated Multiple Sclerosis women patients[237] and in the subjects with a high response to the supplementation (25(OH)D: > 75 ng/mL vs. < 50 ng/mL)[216]. It correlated positively with 1,25(OH)D and the activation ratio in Thomas et al. [42], but was inversely correlated with 25(OH)D levels in the healthy cohort by Luthold et al.

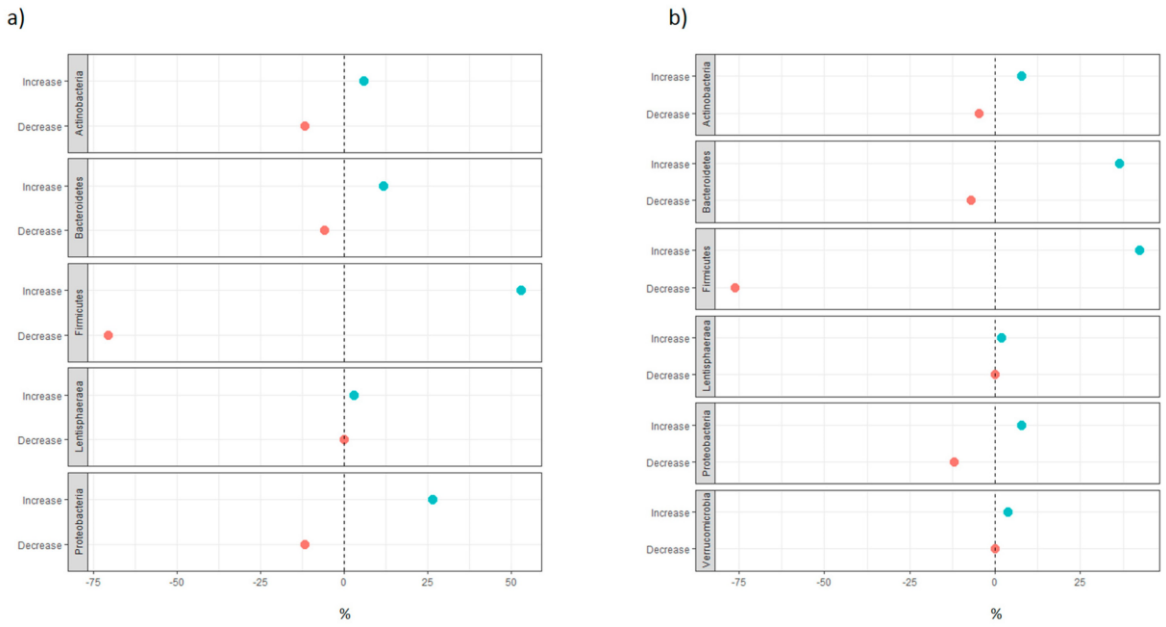


Figure 3.3. (**a**) For each phylum, the blue dot indicates the number of significant taxa in the phylum that increased with increasing levels of vitD serum levels or dietary intake, over the total number of significant taxa that increased with increasing levels of vitD serum levels or dietary intake in the non-supplementation group of studies (expressed in percentages); the red dot indicates the number of significant taxa in the phylum that decreased with increasing levels of vitD serum levels or dietary intake, over the total number of significant taxa that decreased with increasing levels of vitD serum levels or dietary intake in the non-supplementation group of studies (expressed in percentages). (**b**) For each phylum, the blue dot indicates the number of significant taxa in the phylum taxa that increased after vitD supplementation, over the total number of significant taxa that increased after vitD supplementation in the supplementation group of studies (expressed in percentages); the red dot indicates the number of significant taxa in the phylum that decreased after vitD supplementation, over the total number of significant taxa that decreased after vitD supplementation in the supplementation group of studies (expressed in percentages).

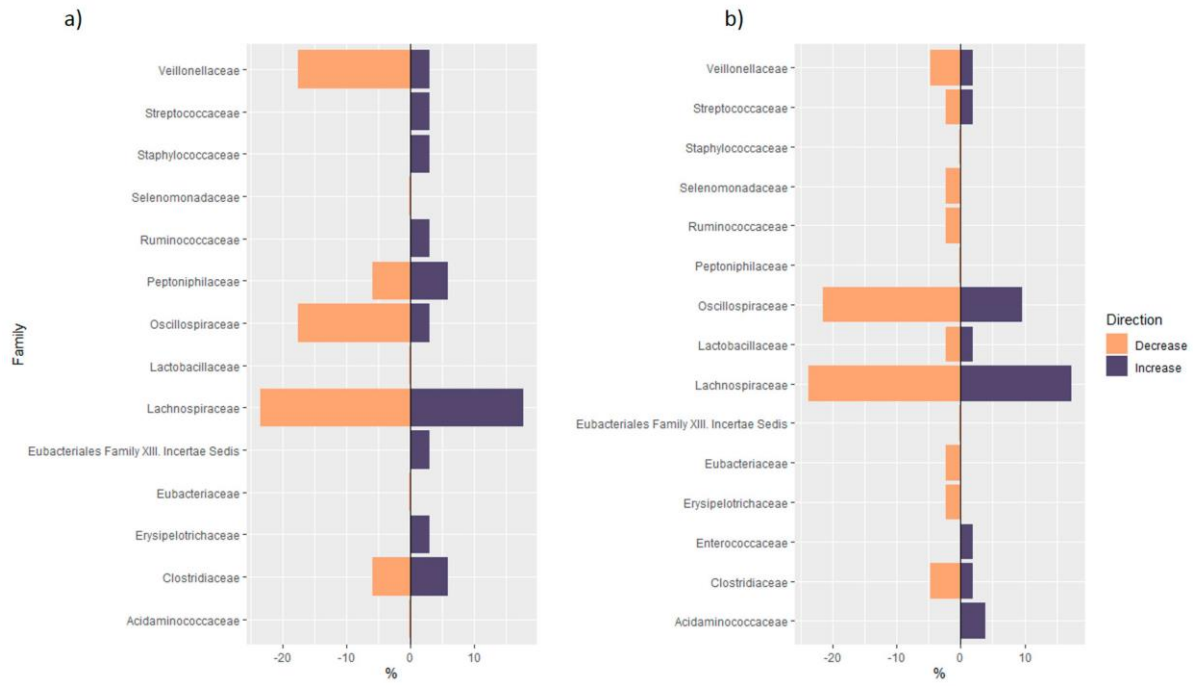


Figure 3.4. (**a**) mirror bar chart of the number of significant taxa in each family of Firmicutes phylum over the total number of significant taxa in each family of Firmicutes phylum found in the non-supplementation group of studies: for each family, the violet bar indicates the number of significant taxa in the family that increased with increasing levels of vitD serum levels or dietary intake, over the total number of significant taxa in Firmicutes that increased with increasing levels of vitD serum levels or dietary intake (expressed in percentages); the orange bar indicates the number of significant taxa in the family that decreased with increasing levels of vitD serum levels or dietary intake, over the total number of significant taxa in Firmicutes that decreased with increasing levels of vitD serum levels or dietary intake (expressed in percentages); (**b**) Mirror bar chart of the number of significant taxa in each family of Firmicutes over the total number of significant taxa of Firmicutes phylum found in the supplementation group of studies: for each family, the violet bar indicates the number of significant taxa in the family that increased after vitD supplementation, over the total number of significant taxa in Firmicutes that increased after vitD supplementation (expressed in percentages); the orange bar indicates the number of significant taxa in the family that decreased after vitD supplementation, over the total number of significant taxa in Firmicutes that decreased after vitD supplementation (expressed in percentages).

3.3.3. Analysis of Phylogenetic Trees of Studies

In studies involving vitD supplementation and including healthy subjects, Firmicutes, Actinobacteria and Bacteroidetes were the most recurrent phyla that either increased or decreased following supplementation (Figures S1 and S2). In Firmicutes phylum, several core genera from the Lachnospiraceae family, like *Lachnospira*, *Fusicatenibacter* and *Lachonacterium*, increased following vitD supplementation. Conversely, two studies reported a decrease in the *Faecalibacterium* genus from the Oscillospiraceae family[219,238]. Moreover, several genera from the Lactobacillales order, such as *Lactococcus* and *Lactobacillus*, were found to decrease after vitD supplementation, except for *Enterococcus*, which increased in the female cohort of adolescents by Tabatabaeizadeh et al.

Increasing abundances were also found in Actinobacteria phylum, in particular in the *Bifidobacterium* genus, and in other genera from Bacteroidetes, such as *Bacteroides*, *Parabacteroides* and *Alistipes*.

In the group of studies not involving vitD supplementation and including healthy subjects (Figures S3 and S4), the associations between microbial taxa and vitD serum levels or dietary intake were investigated. In *Veillonella* (Firmicutes phylum) and *Haemophilus* (Proteobacteria phylum) were found to be significantly more abundant in the lowest compared to the highest tertile of both vitD intake and serum 25(OH)D levels[228]. *Coprococcus* (Firmicutes phylum) and *Bifidobacterium* (Actinobacteria) genera were also found to be inversely correlated with 25(OH)D levels, even after adjustment for confounders. On the other hand, the *Megasphera* genus from the Negativicutes order (Firmicutes phylum) was significantly more abundant in the highest tertile of 25(OH)D levels compared to the lowest[228]. In the community-dwelling older-men cohort by Thomas et al., they observed a significant positive correlation between *Coprococcus catus* and *Blautia Obeum* species (Firmicutes phylum; Clostridia class) and the active form of vitD, 1,25(OH)2D. Moroever, they found a positive correlation between the Eubacteriales order, Ruminococcaceae, Lachnospiraceae, Victivallaceae families, *Coprococcus* and *Mogibacterium* genera, and the ratio of active vitD. Conversely, Blautia and Oscillospira, belonging to the Firmicutes phylum and Clostridia class, were significantly and negatively associated with both 1,25OH2D levels and the vitD active ratio (1,25OH2D/25(OH)D).

In Tables S2–S7, we provided the phylogenetic reconstruction of the taxa that significantly decreased or increased after vitD supplementation, whereas in Tables S8-S11 we provided the phylogenetic reconstruction of the taxa that were significantly and positively or negatively correlated with either vitD serum levels or dietary intake. The tables were stratified according to health status of the enrolled populations and type of microbiome samples.

One of the five RCTs included in the review showed a dose-response effect of vitD supplementation on microbiota composition, with increased abundances of *Bacteroides* and *Parabacteroides* in the

supplemented group[219]. In overweight or obese patients, a first loading dose of cholecalciferol (100,000 UI) followed by 4000 UI/day was significantly associated with a higher abundance of the genus *Lachnospira* and with a lower abundance of the genus *Blautia*[215].

In Singh et al, the changes in the microbiota following vitD supplementation occurred only in the superior gastro-intestinal tract and were detected in biopsies but not in fecal samples. On the other hand, Bosman et al. showed significant results only after the exposure to narrow-band ultraviolet B light, especially for the vitD -deficient group.

Prenatal higher doses of vitD were also associated with significant changes in infant microbial composition, resulting in decreased abundances of *Bilophila* and *Lacnospiraceae*. Conversely, infant vitD supplementation did not show a significant effect on gut microbiota, except for the lower abundance of the genus *Megamonas*, as reported by Drall et al. Talsness et al. found that *Bifidobacterium* abundance was inversely related to higher levels of maternal 25(OH)D.

Significant changes in microbiota in relation to vitD frequently occurred in IBD patients. One interventional non-randomized trial[236] showed a significant increase of Firmicutes following vitD supplementation in patients with CD, whereas no significant change was observed in the healthy controls. A positive correlation between Firmicutes and dietary vitD intakes in CD was also found in a case-control study[231]. In UC patients, *Enterobacteriaceae* were found to significantly increase following vitD supplementation, while the *Desulfovibrio* genus increased at increasing levels of vitD intakes. Both taxa belong to the Proteobacteria phylum.

Among the included studies, three did not find any association between vitD and gut microbiota. One study was a RCT[216] including HIV patients and involving vitD supplementation for 16 weeks with a dose of 5000 UI/day. The second was a cross-sectional study[227] of young Japanese women, where the association between vitD dietary intake and gut microbiota was investigated. The last one was a large UK twins cohort study[229], where the authors did not find any significant association between self-reported use of vitD supplements and change in gut microbiota.

4. RANDOMIZED PHASE II TRIAL ON VITAMIN SUPPLEMENTATION

4.1 Rationale of the study

The final step in our research was to design a phase II randomized placebo-controlled trial to evaluate whether VitD supplementation and changes in serum levels of 25(OH)D could determine a microbial composition modification in CRC survivors. We enrolled patients with resected stage I and III CRC after surgery, and completion of the adjuvant therapy if occurred. Participants were randomly assigned to 2000 IU/day versus placebo and treated for 1 year. For each patient, we collected data on gut microbiome, diet and lifestyle, circulating markers at both baseline and at the end of the treatment. For a subgroup of patient, we also had the gene expression (GE) profiling of a set of 395 immuno-related genes evaluated in the tumor tissue.

The primary endpoint of the study was to assess if vitD supplementation modulated the microbial composition of patients, whereas secondary analysed focused on the analysis of the interplay between the investigated factors to understand if and how they affected both gut microbiome and CRC progression.

All the steps undertaken in the trial have been comprehensively summarized in Figure 4.1.

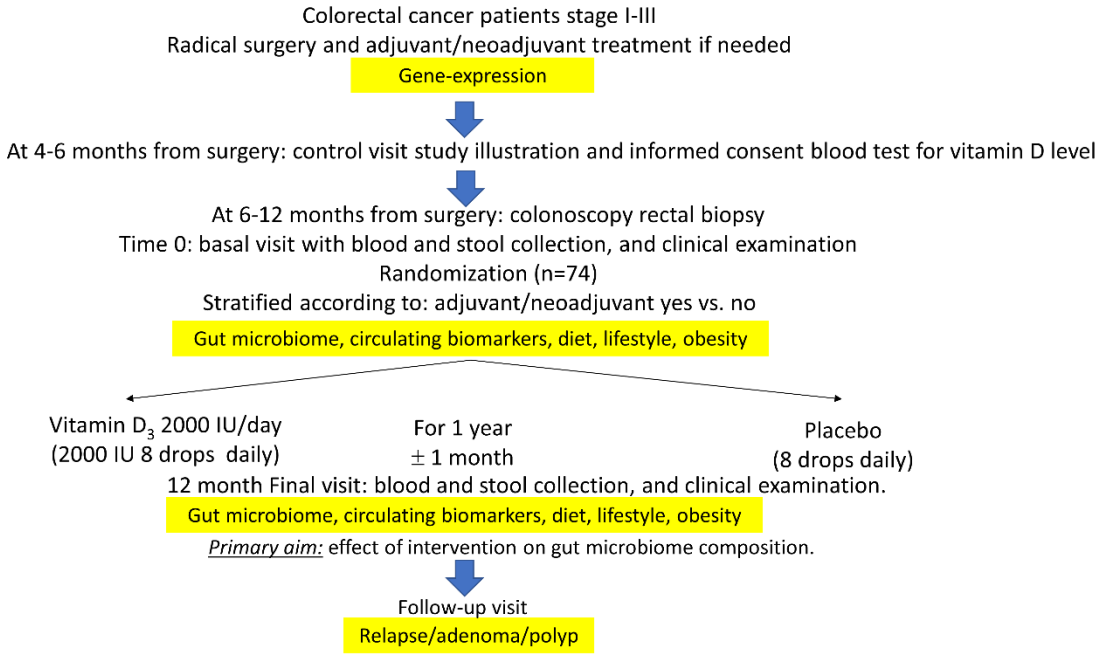


Figure 4.1. Graphical representation of the steps undertaken in the trial, detailing the samples and data collections at the various time points throughout the research process.

4.2 Study design and participants

We designed a randomized double-blind placebo controlled phase II trial to evaluate whether vitD supplementation (and changes in serum 25(OH)D levels) affect the composition of gut microbiome. We enrolled 74 patients with resected stage I-III CRC after surgery, undergoing or not chemotherapy and/or radiotherapy, randomly assigned to either vitamin D3 2000 IU a day or placebo for 1 year with allocation 1:1. Stratification was made for chemotherapy (adjuvant or neoadjuvant) or not.

The Istituto Europeo di Oncologia institutional review board approved the study with the number IEO 223 and Eudract number 2015-000467-14.

Inclusion criteria were:

- Patients with resected stage I-III CRC in the last 24 months.
- Aged 35-75 years old.
- Signed informed Consent according to the International Committee on Harmonization of Good Clinical Practice (ICH-GCP) guidelines.
- Willingness to provide stool, blood samples and rectal mucosa biopsies.
- Performance Status of 0-1 (ECOG).
- Hematopoietic hepatic, renal functionality and serum calcium lower then grade 2 base on the common terminology criteria.

Exclusion criteria were:

- History of cancer in the prior five years (other than cervical intraepithelial neoplasia and non-melanoma skin cancer).
- Carrier of a pathogenic mutation for the main syndrome for CRC (FAP, Lynch, other).
- Clinical/radiological evidence or laboratory/pathology report of residual neoplasia or recurrence.
- Vitamin D level ≥ 30 ng/ml (external exams).
- Current daily supplementation of vitamin D (e.g. calcium citrate with vitamin D).
- History of recurrent renal calculi.
- History of malabsorption syndrome (e.g., pancreatic insufficiency, celiac disease, Crohn disease, any chronic IBD).
- Chronic liver disease and/or renal disease with altered biochemical functions, or renal dialysis.
- Pregnancy or breast feeding or planning on becoming pregnant during the study.
- Known chronic alcoholism.
- Known hypersensitivity to vitamin D.

- Any medical condition that in the physician's opinion would potentially interfere with the subjects' health.

To avoid any effect of tumour treatment on the composition of the microbiome, a wash-out period of 4-6 months after surgery was planned for each patient enrolled before baseline visit.

During the baseline visit, full medical history was collected, along with anthropometric measurements, blood and stool samples and information on smoking habits and any concurrent medications. A self-administered food frequency questionnaire was also administered (Supplementary Table 1).

The therapy boxes of vitD and matching placebo were prepared by the pharmacists at the European Institute of Oncology. VitD3 was in an oily solution and the placebo was made to be visually identical to the active formulation. A six-month supply was provided to the participant at baseline visit.

After 3 months, a phone call was made to check safety and compliance. At 6-month visit, safety, clinical examination and concomitant medications were assessed, and the new 6-month drug supply was provided. At 12-month final visit, safety, clinical examination, concomitant medications and all biological samples were collected.

Compliance was assessed though a self-reported diary collected at each visit and graded according to the level of adherence (1=83-100%, 2= 66-82%, 3=25-65%, 4=25%, 5=none). A patient was considered compliant if their level of adherence was of grade 1 or 2 at all times. Since we did not measure the returned leftover agent, 25(OH)D measurements were considered as an additional compliance control.

Targeted gene expression profiling was also conducted on the tumour tissue of participants using the RNA-based Next Generation Sequencing (NGS) panel Oncomine Immune Response Research Assay (OIRRA) (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. This assay allows for the simultaneous evaluation of expression of 395 immune-related genes. However, data from this analysis was only available for 46 patients, resulting in a total of 48 profiles. Notably, two patients had the analysis performed on two tumour samples because of different characteristics of the tumour in the two (i.e. different infiltration level of the neoplasia).

4.3 Materials and methods

4.3.1 Sampling of biological specimens

Morning fasting samples of whole etheylenediaminetetraacetic acid (EDTA)-treated blood and serum were collected at baseline and after 12 months following storage at -80°C until biomarker measurement.

Freshly voided stool samples were collected at both timepoints. The stool sample was collected in a tube, stored at -20°C and then transported to the laboratory in a plastic bag containing an ice pack. Upon arrival to the laboratory, each sample was immediately frozen at -80°C.

Serum 25(OH)D concentrations were determined by a commercially available chemiluminescent immune assay (Immunodiagnostic Systems, Pantec S.r.l., Turin, Italy). This method recognizes both metabolites of vitD (D2-D3). Concentrations of adiponectin, leptin, IL-10, IL-6 TNF-α were determined using an ELISA kit, whereas concentrations of CCL2/MCP1, CD27, CD40 Ligand, CXCL6/GCP-2, Galectin-3, IL-8/CXCL8, CD40, CXCL2/GROβ, Galectin-1, Galectin-9, IL-7 and B7-H1/PD-L1 were obtained using the Immuno-Oncology Checkpoint LXSAHM-31 kit by R&D Systems.

4.3.2 Microbiome Analyses

For metagenomic analysis, genomic bacterial DNA was isolated from feces of patients using G'NOME isolation kit (MP Biomedicals) following a published protocol[242]. Whole metagenome shotgun sequencing[175] was applied on the DNA samples. Metagenomic libraries were generated with a Nextera XT DNA Sample Prep Kit (Illumina, San Diego, CA, USA) and sequencing was carried out on the HiSeq2500 platform (Illumina) at a targeted depth of 5.0 Gb (100-bp paired end reads).

DNA sequences were aligned to a curated database containing all representative genomes in RefSeq[243] for bacteria with additional manually curated strains. Alignments were made at 97% identity against all reference genomes. Every input sequence was compared to every reference sequence in the CoreBiome Venti database using fully gapped alignment with BURST[244] . Ties were broken by minimizing the overall number of unique Operational Taxonomic Units (OTUs). For taxonomy assignment, each input sequence was assigned the lowest common ancestor that was consistent across at least 80% of all reference sequences tied for best hit. The number of counts for each OTU was normalized to the OTU's genome length. OTUs accounting for less than one millionth of all species-level markers and those with less than 0.01% of their unique genome regions covered (and < 1% of the whole genome) were discarded. Samples with fewer than 10,000 sequences were also discarded. Count data was converted to relative abundance for each sample. The normalized and filtered table was used for all downstream analyses.

4.3.4 Pathway Analyses

To analyze the gut microbiome, we applied bioBakery tools[245] on whole shotgun metagenomic data of stool samples. To quantify the relative abundance of microbial species, we carried out MetaPhlAn 3 pipeline[246] on raw reads. MetaPhlAn profiles the microbial community with 1.1 million microbial protein-coding gene markers (circa 50-400 marker genes for each bacterial species). The relative abundances of microbial pathways and functional potentials were computed utilizing the HUMAnN 3[246]. HUMAnN provides the contribution of each species to the gene families and pathways.

4.3.5 Targeted next-generation sequencing gene expression analysis

Gene expression analysis was performed using the RNA-based NGS panel Oncomine Immune Response Research Assay (OIRRA) (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions. This assay enables the simultaneous evaluation of the expression of 395 immune-related genes, including subgroups of leukocytes, antigen presentation, checkpoint pathways and tumor progression including low-expression genes involved in inflammatory signaling. Briefly, 25ng of RNA was used for the library preparation and the subsequent chip loading, both automatically performed on the Ion Chef System (ThermoFisher Scientific, Waltham, MA, USA). The sequencing run was done on Ion S5 System (ThermoFisher Scientific, Waltham, MA, USA) and genes expression data were obtained using the TorrentSuite ImmuneResponseRNA plugin software (ThermoFisher Scientific, Waltham, MA, USA).

Tissue specimens were available for 62 patients, resulting in a total of 64 tumor samples. Out of these, 48 samples met the quality criteria set for the sequencing run (mapped reads> 1000000; valid reads> 80000). Ultimately, gene expression data was available for 46 patients. For two of these patients, the analysis was performed on two distinct tumour samples due to differences in tumour characteristics in the two samples, such as varying levels of neoplastic infiltration.

4.4 Statistical methods

Demographic and clinical characteristics of the enrolled population were summarized by treatment arm using median and interquartile range for numerical variables and absolute frequencies and percentages for categorical variables. Wilcoxon rank-sum test and Chi-square test (or Fisher exact test, when appropriate) were used to test differences between treatment arms. Changes in 25(OH)D concentrations between the two timepoints were tested within each arm using Wilcoxon signed-rank test.

4.4.1 The challenge of compositional data analysis (CoDA)

High-throughput sequencing technologies, such as 16S rRNA gene sequencing or metagenomic sequencing, have revolutionized the field of microbiome research, providing unprecedented insights into microbial diversity, function, and their roles in health and disease.

However, the multitude of data that these technologies produce comes with its own set of challenges, especially in the realm of data analysis.

These technologies generate a number of sequence reads for each sample, which are then mapped to known microbial taxa. The number of reads corresponding to each taxon is counted, but these counts are constrained by the total number of reads generated, which can vary between samples and sequencing runs. Given this constraint to the arbitrary total sum of reads, the raw read counts are often normalized to their relative abundances within each sample[247].

This means that the count of each taxon is divided by the total counts for all taxa in that sample, converting them into proportions that sum to 1 or any other constant or percentages that sum to 100%. This normalization process allows for the comparison of microbial communities across different samples and conditions, even when the total number of sequence reads varies across samples. While relative abundances facilitate comparisons, it is important to recognize that they are **compositional data**.

In a compositional dataset, the value of each component (in this case, each microbial taxon) is inherently dependent on the values of all other components. For example, an increase in the relative abundance of one taxon will necessarily result in a decrease in the relative abundance of one or more other taxa, even if their absolute abundances remain constant (Figure 4.2).

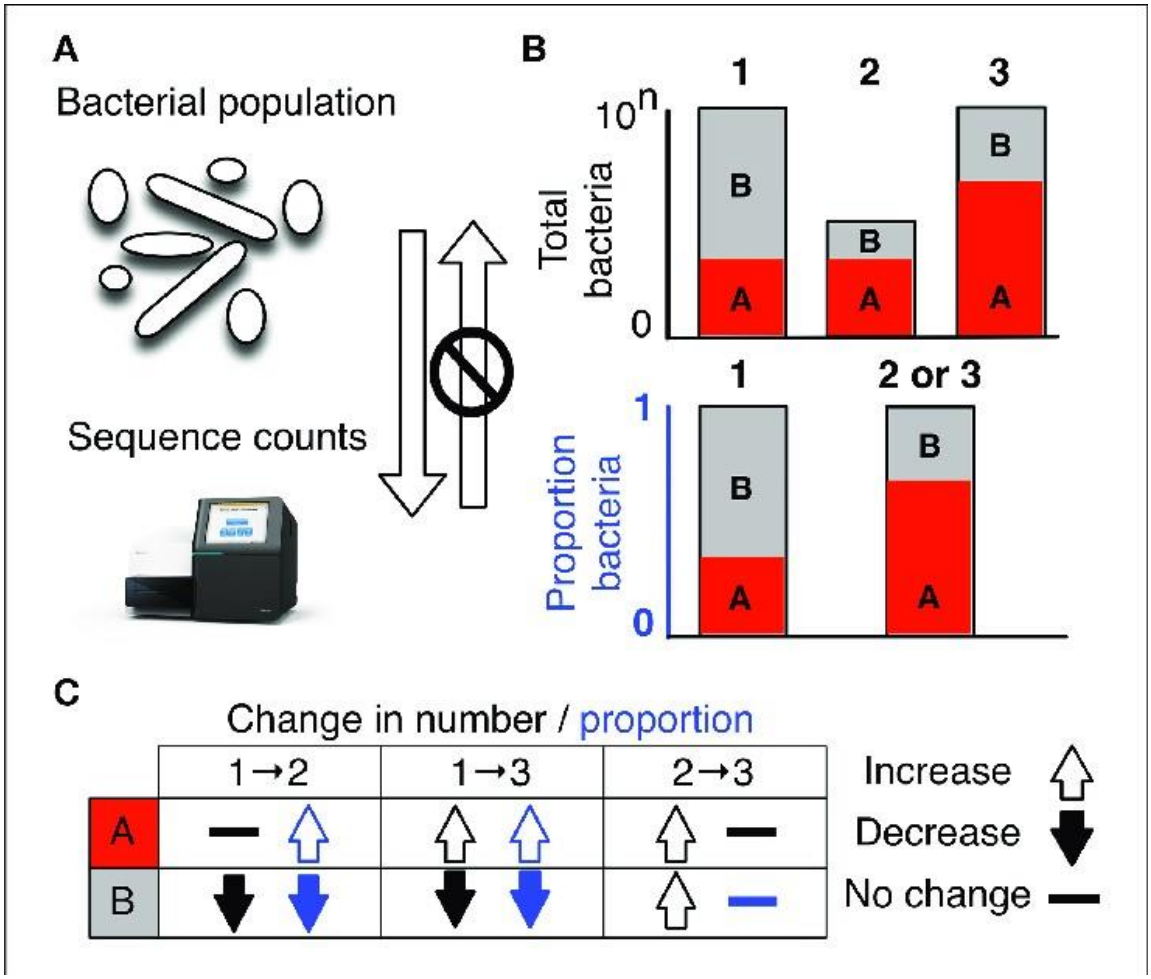


Figure 4.2. Adapted with permission from Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front Microbiol.* 2017 Nov 15;8:2224. doi: 10.3389/fmicb.2017.02224[248].
High-throughput sequencing data are compositional. (A) illustrates that the data observed after sequencing a set of nucleic acids from a bacterial population cannot inform on the absolute abundance of molecules. The number of counts in a high throughput sequencing (HTS) dataset reflect the proportion of counts per feature (OTU, gene, etc.) per sample, multiplied by the sequencing depth. Therefore, only the relative abundances are available. The bar plots in (**B**) show the difference between the count of molecules and the proportion of molecules for two features, **A** (red) and **B** (gray) in three samples. The top bar graphs show the total counts for three samples, and the height of the color illustrates the total count of the feature. When the three samples are sequenced we lose the absolute count information and only have relative abundances, proportions, or "normalized counts" as shown in the bottom bar graph. Note that features **A** and **B** in samples 2 and 3 appear with the same relative abundances, even though the counts in the environment are different. The table below in (**C**) shows real and perceived changes for each sample if we transition from one sample to another.

This interdependence among parts – defined *compositional dependence* - violates the assumptions of *independence* that underlie most of the traditional statistical methods, necessitating the use of specialized techniques for compositional data analysis[249,250]. Additionally, the constrained nature of these data to a constant sum challenges other key assumptions of classical statistical

methodologies, specifically *normality* and *homoscedasticity*, as the data is bounded and not free to vary across the entire real number line[251].

The geometric implications of these violations are profound and can lead to misleading results and incorrect interpretations.

4.4.1.1 CoDA: the simplex space and properties of compositional data

By definition, *compositional data* are non-negative multivariate data, which carry only relative information. They usually have a constant-sum constraint, which implies a composition. The components of a composition are called *parts*, which are always positive and add up to the total[252]. Unlike most data types that are naturally located in an unconstrained *real* space, compositional parts are projected into a constrained geometric space known as the *simplex* space[253] $\mathbb{S}$:

$$\mathbb{S}^D = \left\{\mathbf{x} = [x_1, x_2, \ldots, x_D] \in \mathbb{R}_+^D : \sum_{i=1}^{D} x_i = k\right\}, \quad k > 0$$

where $\mathbf{x}$ is the composition, $x_i,\ i = 1, \ldots, D$ are the parts of the composition, D is the number of parts and k is the positive closure constant. (Figure 4.3)

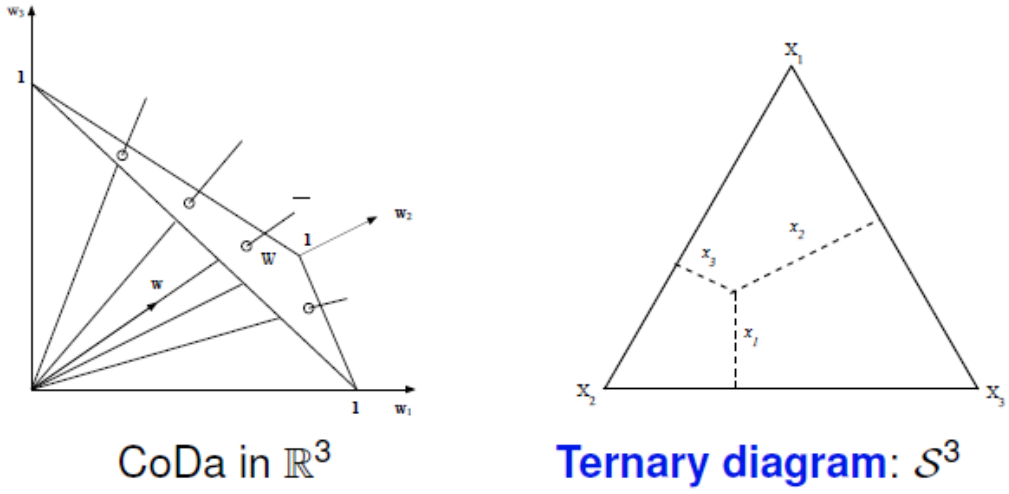


Figure 4.3. Graphical representation of compositional data in: **A**. a three-dimensional real space; **B**. three-dimensional simplex space (ternary diagram). The figure was adapted from the notebook in *Compositional Analysis of Data with CoDaPack* provided by the CoDa-Research Group during the *10th Course and Open Seminar on Compositional Data Analysis* held in Girona (Spain) in July 3-7, 2023.

The three fundamental properties of compositional data are[254]:

- *Scale invariance*: the results of the analysis should not depend on the scale or unit of measurement, as proportional positive parts carry the same information in terms of compositions (*compositional equivalence*).

- *Permutation invariance*: the order of parts is not relevant for the analysis.
- *Subcompositional coherence*: the analysis of subcompositions after closure is not in contradiction with that obtained from the full composition.

4.4.1.2 The log-ratio approach and data transformation

The properties just described imply that the classical statistical methodologies are inadequate for the analysis of compositional data, especially the methods based on the covariance or correlation matrix of the observations.

Pearson was the first to address this problem in 1897, when he pointed out the risk of obtaining spurious correlations in case of ratios whose numerators and denominators have common parts[255]. However, it was John Aitchison who, in the eighties, introduced the definition of "compositional data" for this kind of data and proposed a *log-ratio approach* to address these challenges[252].

He underlined that the key information in constrained data does not reside in the absolute values of the parts but rather in their ratios. This is because the ratio between any two parts remains constant, regardless of the other parts included in the composition and the constraint of closure (for the property of *subcompositional coherence*)[253].

However, ratios operate on a multiplicative scale, which can be difficult to handle. The log-ratio approach addresses this by applying a logarithmic transformation to the ratios, thereby converting the data to an additive scale.

The simplest form of log-ratio function, which guarantees scale invariance, is the logarithmic transformation of a ratio between two parts of a composition:

$$\ln\left(\frac{x_i}{x_j}\right) \qquad \boldsymbol{x} \in \mathbb{S}^D$$

A generalization of the log-ratio transformation is the log-contrast function

$$\sum_{i=1}^{D} a_i \ln\, x_i = \ln\left(\prod_{i=1}^{D} x_i^{a_i}\right) \quad \text{with} \quad \sum_{i=1}^{D} a_i = 0$$

which is the linear combination of the logarithm of the parts with the coefficients a_i. The restriction $\sum_{i=1}^{D} a_i = 0$ guarantees scale invariance.

Several transformations have been proposed to shift compositional data from the simplex space to the real space, allowing the application of many of the conventional statistical methodologies.

Given $\boldsymbol{x} = [x_1 + x_2 + \cdots + x_D]$ a composition of $\mathbb{S}^D$, the three main transformations based on the log-ratio approach are[256]:

- **The additive log-ratio (alr) transformation**

$$alr\boldsymbol{x} = \left[\ln\frac{x_1}{x_D}, \ln\frac{x_2}{x_D}, \ldots, \ln\frac{x_{D-1}}{x_D}\right] \in \mathbb{R}^{D-1}$$

In this transformation, one part of the composition x_D is chosen as a reference, and the ratios of all other components to this reference part are calculated. The natural logarithm of these ratios is then calculated to convert the data into an additive scale. This transformation results in a $(D-1)$-dimensional real vector, reducing the dimensionality of the original compositional data by one.
Although the *alr* transformation is dependent on the choice of the reference part x_D, it has been shown that this choice does not affect the resulting inference[257]. However, this transformation is not an isotropy, meaning that the results obtained from statistical methods based on distances between *alr* vectors are not the same that would be obtained on the compositional distances in the simplex. For this reason, this transformation is not recommended for statistical analysis.

- **The centered log-ratio (clr) transformation**

$$clr\boldsymbol{x} = \left[\ln\frac{x_1}{g(\boldsymbol{x})}, \ln\frac{x_2}{g(\boldsymbol{x})}, \dots, \ln\frac{x_D}{g(\boldsymbol{x})}\right] \in \mathbb{R}^D$$

where $g(\boldsymbol{x}) = \left(\prod_{i=1}^{D} x_i\right)^{1/D}$ is the geometric mean of **x** and $clr\boldsymbol{x}$ is a constrained vector of $\mathbb{R}^D$ so that $\sum_{i=1}^{D} clr_i\boldsymbol{x} = 0$.
This transformation returns a D-dimensional real vector, with the inherent constraint of the components summing to zero. Consequently, the covariance matrix of the *clr*-transformed data is singular and the Pearson correlation coefficients $corr(x_i, x_j)$, $i, j = 1, \dots, D$ are non-informative.
Despite these drawbacks, the *clr* transformation is an isometry that preserves the compositional distances in the simplex space (see *The geometry of the simplex space*[258]). This property allows for the use of various standard statistical methodologies - especially those focused on clustering - in the transformed space, without altering the relationships between data points in the simplex. Additionally, it facilitates straightforward interpretation of results. For these reasons, we employed the *clr* transformation for our analysis of microbiome and gene expression data.

- **The isometric log-ratio (ilr) transformation**[259]

Similarly to the *clr* transformation, the *ilr* transformation is an isometry which preserves the geometry of the original compositional data into an unconstrained Euclidean space.
It is defined by creating $D-1$ orthonormal (olr) coordinates – called *balances* - that capture the relationships between parts of the composition. One of the most employed methods to calculate *olr* coordinates is Sequential Binary Partition (SBP). SBP is a hierarchical binary partitioning of the components of a composition which, at each level of the hierarchy, splits the components into two

groups, and the geometric means of these groups are used to compute a set of coordinates. This hierarchical splitting continues until no more splitting is possible.

Mathematically, the ILR transformation can be expressed as follows:

$$ilr(\boldsymbol{x}) = [z_1, z_2, \dots, z_{D-1}] \in \mathbb{R}^{D-1}$$

Each z_j is calculated based on the SBP and is defined as:

$$z_j = \sqrt{\frac{n_j . d_j}{n_j + d_j} \ln \frac{\left(x_{k_1} \dots x_{k_{n_j}}\right)^{1/n_j}}{\left(x_{l_1} \dots x_{l_{d_j}}\right)^{1/d_j}}}$$

where n_j is the number of parts in the numerator coded as +1 in the sign matrix, d_j is the number of parts in the denominator coded as -1 in the sign matrix.

$k_1, \dots, k_{n_j}$ are the labels of the parts in the numerator and $l_1, \dots, l_{d_j}$ are the labels of the parts in the denominator.

Although the *ilr* transformation is mathematically convenient and facilitates robust statistical analysis, the interpretation of results post-transformation can be very challenging. The creation of orthonormal coordinates based on SBP of the components usually leads to new transformed data that is difficult to relate back to the original components, especially in the case of high-dimensional datasets, like in microbiome studies. For this reason, we decided not to use this transformation in our study.

4.4.1.3 The geometry of the simplex space

The *simplex space* is a constrained geometric space where each point represents a composition. The geometry of the simplex is inherently different from Euclidean geometry due to the constant sum constraint. The concept of "distance" in the simplex space is not straightforward and requires a specialized metric, one of which is the ***Aitchison distance***[258].

The *Aitchison distance* is a metric specifically designed for measuring the dissimilarity between two compositions in the simplex space and is usually denoted as d_a. Named after John Aitchison, this distance metric is based on the log-ratios of the components of the compositions being compared.

Mathematically, the Aitchison distance d_a between two compositions $\mathbf{x} = [x_1, x_2, \dots, x_D]$ and $\mathbf{y} = [y_1, y_2, \dots, y_D]$ is defined as:

$$d_a(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{D} \sum_{i=1}^{D-1} \sum_{j=i+1}^{D} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2} = \sqrt{\frac{1}{D} \sum_{i=1}^{D} \left(\ln \frac{x_i}{g(\mathbf{x})} - \ln \frac{y_i}{g(\mathbf{y})} \right)^2}$$

where D is the number of parts in $\mathbf{x}$ and $\boldsymbol{y}$, $g(\mathbf{x})$ and $g(\mathbf{y})$ are the geometric means of $\mathbf{x}$ and $\boldsymbol{y}$, respectively.

The equation below shows that the Aitchison distance between two compositions is equal to the Euclidean distance between the *clr*-transformed compositions:

$$d_a(\mathbf{x}, \mathbf{y}) = d(clr\mathbf{x}, clr\mathbf{y}) = \sqrt{\sum_{i=1}^{D} (clr_i\mathbf{x} - clr_i\,\mathbf{y})^2}$$

where d is the Euclidean distance. For this reason, the clr transformation is an *isometry*.

Aitchison distance has *scale invariance*, *permutation invariance* and *subcompositional dominance*, making it a robust and reliable metric for statistical analysis[260].

4.4.1.4 Perturbation and power operations in the simplex

Because of the sum-constraint of compositional data, the simplex space has its own set of operations, which are different from those in the Euclidian space.

Let $\mathbf{x}$ and $\mathbf{y}$ be compositions in $\mathbb{S}^D$, $a \in \mathbb{R}$ and C the closuring function:

- **Perturbation** of $\mathbf{x}$ by $\mathbf{y}$

$$\mathbf{x} \oplus \mathbf{y} = \mathrm{C}[\mathrm{x_1 y_1}, \mathrm{x_2 y_2}, \ldots, \mathrm{x_D y_D}]$$

Perturbation in the simplex space is equivalent of vector addition in Euclidean space. Given two compositions $\mathbf{x}$ and $\mathbf{y}$, their perturbation is obtained through element-wise multiplication of the components, followed by closure.

- **Perturbation difference** between $\mathbf{x}$ and $\mathbf{y}$

$$\mathbf{x} \ominus \mathbf{y} = C[x_1/y_1, x_2/y_2, \ldots, x_D/y_D]$$

Perturbation difference is the inverse of the perturbation operation and corresponds to vector subtraction in Euclidian space. The perturbation difference between $\mathbf{x}$ and $\boldsymbol{y}$ quantifies the difference between the two compositions in the simplex and is obtained through element-wise division of each component of $\mathbf{x}$ by each component of $\mathbf{y}$, followed by closure.

- **Powering** of $\mathbf{x}$ by a

$$\mathrm{a} \ominus \mathbf{x} = C[{x_1}^a, {x_2}^a, \ldots, {x_D}^a]$$

Powering in the simplex corresponds to scalar multiplication in Euclidian space. Given a composition $\mathbf{x}$ and a real number a, the power transformation is obtained by raising each part of $\mathbf{x}$ to the power a, followed by closure.

In our study, we employed perturbation difference to evaluate changes in the microbiome following the 1-year treatment. Specifically, we calculated:

$$\boldsymbol{x_B} \ominus \boldsymbol{x_A} = C\left[x_{B_1}/x_{A_1}, x_{B_2}/x_{A_2}, \ldots, x_{B_D}/x_{A_D}\right]$$

where C is the closuring function, D is the number of taxa, $\boldsymbol{x_A}$ is the composition of the microbiome at baseline and $\boldsymbol{x_B}$ is the composition of the microbiome at follow-up.
After closuring, we applied the *clr* transformation for statistical analysis.

4.4.1.5 The handling of zeros in CoDA

One of the main hurdles in CoDA is managing zeros. This is because log-ratio approaches require logarithmic transformations, which are not defined for zeros.
Zeros in a composition can exist for various reasons, and they can be categorized as follows[261,262]:

- *essential* or *structural* zeros, which are real zeros indicating the absence of a specific part. Replacing these zeros with small values is not appropriate.
- *count* zeros, which can be present in discrete parts and are common in studies that involve counting, like those using data from high-throughput sequencing technologies. In such cases, Bayesian-multiplicative (BM) replacement methods are appropriate for imputation of zeros. These methods treat compositions as probability vectors in a multinomial model and replace zeros with small values[263].
- *rounded* zeros, which occur in continuous components and are typically small numbers that have been rounded to zero. This rounding usually happens because they fell below the maximum round-off error or below the detection limit (DL) of the instrument. In this case, zeros could be replaced using parametric/non-parametric and univariate/multivariate algorithms[263].

In our study, because we only had data on relative abundances for the gut microbiome, we assumed that the zeros were rounded. For zeros imputation, we employed the non-parametric multiplicative simple imputation method of left-censored data, using the smallest relative abundance observed across all samples as DL. Conversely, for the Gene Expression (GE) data, we had information on absolute read counts. As a result, we employed the BM replacement method for zero imputation. Both replacement methods were employed using the 'zCompositions' package in R[264].

4.4.1.6 Taxa selection with coda-lasso

We selected the taxa whose abundances were significantly associated with vitD supplementation using the *coda-lasso* (Compositional Data Analysis with Least Absolute Shrinkage and Selection Operator) model.

The *coda-lasso* is a method based on penalized regression designed for variable selection that acknowledges the compositional but also multivariate structure of the microbiome data[265]. The LASSO regularization employed in this model accounts for the sparsity and the high-dimensionality of the data by imposing a penalty term on the regression coefficients, shrinking them to zero.

As shown by the authors, the *coda-lasso* uses penalized regression with log-constraints to overcome the limitations of the centered log-ratio transformation (*clr*) in variable selection, which are due to its lack of subcompositional consistency.

In mathematical terms, the log-constrained model with penalization is defined as following:

$$y_i = \beta_0 + \beta_1 \log(x_{1i}) + \cdots + \beta_k \log(x_{ki}) + \varepsilon_i$$

where $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{ki})$ is the composition of k taxa in sample i and x_{ki} is the relative abundance of taxa k

with constraint $\sum_{j\geq 1} \beta_j = 0$, where the regression coefficients $\beta = (\beta_0, \ldots, \beta_k)$ are estimated to minimize

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 \log(x_{1i}) - \cdots - \beta_k \log(x_{ki}))^2 + \lambda \sum_{j\geq 1} |\beta_j| \quad \text{subject to } \sum_{j\geq 1} \beta_j = 0$$

This linear regression model thus establishes a relationship between log-transformed covariates and the outcome variable through a log-contrast function, which implies that the regression coefficients, with the exception of the intercept, are subject to a zero-sum constraint. This constraint serves to grant the principle of scale invariance. This model is then extended to incorporate a penalized term in the loss function for taxa selection and can be easily adapted to generalized linear models, including the logistic model that was implemented in our analysis.

In our study, we used the *coda-lasso* function available in R at https://github.com/UVic-omics/CoDA-Penalized-Regression, employing logistic regression and including vitD supplementation as outcome.

We selected the taxa that were significantly different at follow-up between vitD supplemented and non-supplemented patients, and repeated the analysis including baseline relative abundances to exclude from the set of selected taxa those that were already differently abundant at baseline.

In each *coda-lasso* model, we chose as optimal penalization term λ the one that provided the largest proportion of explained deviance without returning any warnings related to model convergence and overfitting.

4.4.1.7 Principal Component Analysis in Aitchison geometry

To account for the compositionality of the microbiome data, we used the *clr* transformation on taxa relative abundances because of its isometric properties (see *The log-ratio approach and data transformation*). Zero values were imputed using the non-parametric multiplicative imputation method of left-censored data, using the smallest relative abundance observed across all samples as DL[264]. We conducted a Principal Component Analysis (PCA) based on the covariance matrix, including the *clr*-transformed abundances of selected taxa during the follow-up period (*Aitchison distance*). Graphical representations, including scaled scores of the first two components and their corresponding biplots, were generated to identify patient clusters and examine the influence of each taxon on the component definition. If a component was deemed relevant to our study, the weight of each taxon was quantified using its loading value. A higher absolute value of the loading indicated a stronger correlation between the taxon and the component. Specifically, a positive loading value indicates a positive correlation, while a negative loading indicates an inverse correlation between the taxon and the component.

Significant associations between the selected *clr*-transformed taxa abundances and vitD supplementation/sufficiency were estimated through multivariable logistic regression models, adjusting for significant confounders. Interactions between the microbiome and sex/gender on 25(OH)D levels were also investigated in multivariable regression models.

For the functional analysis, we analyzed the abundances of pathways at the community level. Absolute counts were normalized with the counts per million (CPM) method. Only the pathways that where present in at least 10% of the patients at the end of the treatment were considered for the analysis.

Imputation of zeros, pathways selection and analysis, were carried out using the same methodologies described for taxonomic data.

4.4.2 Methods for alpha and beta diversity calculation

Alpha and beta diversity are two fundamental concepts in the context of microbiome studies.

Alpha diversity is a measure that quantifies the microbial diversity within individual samples. It provides insights into the *richness* and the *evenness* of the species within a sample. Specifically:

- The species ***richness*** is the number of species (or OTUs) present in a given sample, without accounting for their abundances.
- The ***evenness*** measures how similar the abundances of different species (or OTUs) are within a community. If the species in a sample are equally abundant, then evenness is maximum. Conversely, if one or a few species dominate on the others, evenness is low.

In the literature, there are several indices to quantify alpha diversity, which focus on different aspects of diversity. In our study, we used the *Shannon Index*, which combines richness and evenness into a single value, considering both the number of species and their abundance. Higher values of Shannon Index indicate greater diversity. We calculated the Shannon Index at baseline and post-treatment. Correlations and associations with the indices were calculated using non-parametric tests, such as Wilcox rank-sum test, or univariable/multivariable regressions models.

While alpha diversity focuses on the diversity within a single sample, *beta diversity* evaluates the differences between multiple groups of samples. Beta diversity is therefore very important for understanding how factors such as a disease or environmental conditions affect the composition of microbial communities.

There are various metrics used to quantify beta diversity, such as the *Jaccard Index*, which measures the proportion of shared species between two samples, *Bray-Curtis dissimilarity*, which considers both the presence/absence and the abundance of species, and the *UniFrac Distance*, which incorporates phylogenetic relationships between species.

In our study, we compared beta diversity by treatment arm using the Bray-Curtis dissimilarity. However, to account for the compositional nature of the data, we also used the Aitchison distance, i.e the Euclidian distance between the *clr*-transformed relative abundances of all taxa, as suggested by few papers in the literature[266,267].

We tested differences in beta diversity by treatment arm and by conditions using permutational multivariate analysis of variance (PERMANOVA).

4.4.3 Mediation analysis

We conducted a mediation analysis based on the counterfactual framework approach[268] to investigate if the taxa that were associated with the treatment also mediated the effect of vitD supplementation on the post-treatment 25(OH)D levels (Figure 4.4).

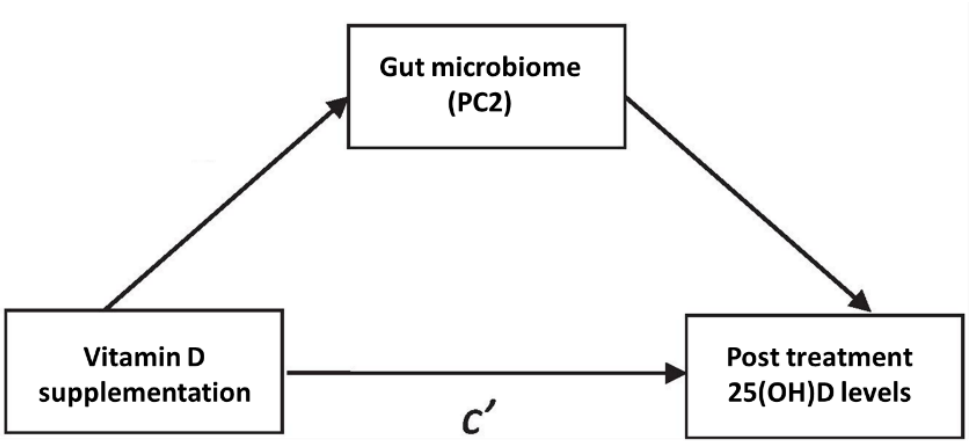


Figure 4.4. DAG showing the causal pathway behind the mediation analysis employed in the trial. Exposure: vitamin D supplementation. Mediator: microbiome summarized through the second principal component (PC2) on the selected taxa (see Results). Outcome: post-treatment 25(OH)D levels. c'=vector of confounders. 25(OH)D = 25-hydroxyvitamin D

We decomposed the total effect (TE) of vitD supplementation (exposure) on 25(OH)D levels (outcome) at follow-up into a natural direct effect (NDE) and a natural indirect effect (NIE) acting on serum vitD levels through the selected taxa (mediator), allowing for the interaction between vitD supplementation and the taxa (see *Mediation analysis*). Sex/gender and 25(OH)D levels at baseline were included as possible confounders.

We summarized the abundances at follow-up of the selected taxa through the latent variable PC2 obtained through PCA on the selected taxa (see *Results*) and included it in the model as mediator.

The NDE was estimated by comparing the effect of vitD supplementation versus placebo on post-treatment 25(OH)D levels, having PC2 set to the value it would naturally have under the placebo group, which is the condition of non-exposure. The NIE was estimated by comparing the effect of PC2 under vitD supplementation versus the effect of PC2 under placebo on post-treatment 25(OH)D levels.

By definition, the TE of the exposure on the outcome is equal to the sum of NDE and NIE:

$$TE = NDE + NIE$$

We thus calculated the three effects as follows.

Let Y be the outcome (25(OH)D levels at follow-up), A the exposure (vitD supplementation), M the mediator (the selected taxa summarized by PC2) and C the set of confounders (sex/gender and 25(OH)D levels at baseline).

The outcome Y was modelled using linear regression as follows:

$$E\{Y|A = a, M = m, C = c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4' c$$

where *c* is the vector of confounders.

The mediator M was modelled using linear regression as follows:

$$E[M|A = a, C = c] = \beta_0 + \beta_1 a + \beta_2{}' c$$

where *c* is the vector of confounders.

We derived NDE and NIE as following:

$$NDE(a, a^*; a^*) = \{\theta_1 + \theta_3(\beta_0 + \beta_1 a^* + \beta_2' E[C])\}(a - a^*)$$

$$NIE(a, a^*; a) = \{\theta_2\beta_1 + \theta_3\beta_1 a\}(a - a^*)$$

where the two levels of the binary exposure being compared are *a**=0 (placebo group) and *a*=1 (vitD supplementation group).

For each effect, the p-value for statistical significance was obtained by calculating the standard errors for each expression using the delta method (see VanderWeele and Vansteelandt, 2009[269], for more details). Because the numerical values of PC2 are not directly interpretable, we did not calculate the Controlled Direct Effect (CDE).

4.4.4 Scoring of diet and lifestyle based on WCRF/AICR recommendations

We built a diet/lifestyle score for each time point based on WCRF/AICR recommendations for cancer prevention published in 2018[176,177].

The score is comprised of 6 components of the total 10 recommendations: 1) be a healthy weight; 2) be physically active; 3) eat a diet rich in wholegrains, vegetables, fruit and beans; 4) limit consumption of "fast foods" and other processed foods high in fat, starches or sugars; 5) limit consumption of red and processed meat; 6) limit alcohol consumption. We did not include the recommendation on the sweet drinks consumption (this information was not available), on the breastfeeding component and the use of supplements for cancer prevention. The total value of the score can range from 0 to 6. Higher levels of the score indicate a higher adherence to the recommendations.

Where possible, we used a three-level scoring system, given that partially meeting a recommendation may confer some benefit. Within each recommendation, if more than one sub-recommendation is operationalized, the scoring weight is divided equally between them to retain a total of one point.

Each component concerning dietary habits was calculated on the basis of the answers provided to the questionnaire in Supplementary Table S1. When an item was missing at a specific time point, it was imputed with the answer to the same item at the other time point.

The scoring was defined as shown in Table 4.1.

Table 4.1 Scoring system of the diet/lifestyle score

2018 WCRF/AICR RECOMMENDATIONS	Reference questionnaire or data collected	ITEM DESCRIPTION	POINTS
1. Be a healthy weight	*Measured BMI*	***BMI (kg/m^2):***	
		18.5–24.9	*1*
		25.0–29.9	*0.5*
		<18.5 or ≥30	*0*
2. Be physically active	*Do you engage in physical activity regularly?*	***Regular physical activity***	
		Yes	*1*
		No	*0*
3. Eat a diet rich in wholegrains, vegetables, fruit and beans	*Questionnaire in Supplementary Table S1:*	***Frequency of wholegrains cereals consumption***	
	Question Q3	*5*	*0.25*
		4	*0.125*
		≤3	*0*
	Question Q20	***Frequency of fruits consumption***	
		5	*0.25*
		4	*0.125*
		≤3	*0*
	Question Q18	***Frequency of vegetables consumption***	
		5	*0.25*
		4	*0.125*
		≤3	*0*
	Question Q4 *used as a proxy for consumption of beans*	***Frequency of soups consumption***	
		5	*0.25*
		4	*0.125*
		≤3	*0*
4. Limit consumption of "fast foods" and other processed foods high in fat, starches or sugars	*Questionnaire in Supplementary Table S1:*	***Frequency of sweets consumption***	
	Question Q24	*1*	*0.5*
		2-3	*0.25*
		4-5	*0*
	Question Q7 *used as a proxy for consumption of fast foods*	***Frequency of pizza consumption***	
		1	*0.5*
		2-3	*0.25*
		4-5	*0*
5. Limit consumption of red and processed meat	*Questionnaire in Supplementary Table S1:*	***Frequency of processed meat consumption***	
	Question Q11 *used also as a proxy for consumption of red meat*	*1*	*1*
		2-3	*0.5*
		4-5	*0*
6. Limit alcohol consumption	*Do you consume alcohol during the week?*	***At least one portion of alcohol during the week***	
		No	*1*
		Yes	*0*
TOTAL RANGE SCORE			***0-6***

BMI=Body Mass Index.

4.4.5 Transcriptomic signature based on Consensus Clustering

We analyzed the expression profiles of 395 immuno-related genes from the OIRRA panel evaluated in the tumour tissue with the aim to build a "transcriptomic signature". To achieve this, we identified clusters of patients based on their gene expression (GE) profile using the consensus clustering approach implemented in the *ConsensusClusterPlus* package[270] in R.

Consensus clustering is an extension of the traditional clustering approach. Instead of producing a single clustering of units, consensus clustering aims to find a stable and robust clustering solution by aggregating multiple clustering results, often derived from different initial conditions, parameter settings, or algorithms. This method is particularly useful in scenarios where assessing the stability and robustness of the identified clusters is challenging, such as when the "large p, small n" problem occurs, as in the omics context.

The final averaged clustering result is typically represented as a ***consensus matrix***, which indicates the frequency with which pairs of data points co-cluster across multiple runs.

Specifically, *ConsensusClusterPlus* employs this methodology by randomly selecting subsets of items (referring to patients) and features (referring to genes). These subsets are then divided into up to *k* clusters using the chosen clustering algorithm. This entire procedure is conducted multiple times. Subsequently, pairwise consensus values are derived, which represent the frequency with which two items are clustered together across the different iterations[270].

To determine the optimal number of clusters, the package provides various graphical representations. These include the hierarchical clustering of the consensus matrix and a plot displaying the cumulative distribution function of the consensus distributions for each potential *k* value.

In our study, we applied consensus clustering to the GE data, post the filtering and normalization steps. For filtering, only genes expressed in at least 10% of the patients were included in the analysis (resulting in 371 genes selected), whereas normalization was conducted using the Reads Per Million (RPM) approach. This method accounts for the sequencing depth by dividing the count of mapped reads by a one million scaling factor of the total mapped reads for each patient. As a result, the normalized GE data were compositional. To address this, we applied the *clr*-transformation to the normalized data, imputing the zeros using the Bayesian-multiplicative replacement method (see *The handling of zeros in CoDA*) based on total read counts.

To ensure the reliability and robustness of the clustering partition, sensitivity analyses were performed. These involved conducting the consensus clustering analysis on the normalized data without *clr*-transformation, and including only one of GE profiles for the two patients whose GE was obtained from two distinct tumour samples.

Finally, the optimal number of clusters was determined through a graphical evaluation of the consensus matrix for each potential *k* number of clusters. The genes whose expression was significantly different across the optimal clusters were identified using the Kruskal-Wallis test.
To account for the FDR, we adjusted the p-values using the Benjamini-Hochberg[271] method. Only corrected p-values less than 0.05 were deemed to indicate statistical significance.

4.4.6 Integrative Analysis of Data

4.4.6.1 Network analysis

To better understand the interplay between the investigated factors, we performed network analysis based on graphical LASSO (GLASSO). GLASSO[272] is a statistical method which estimates the sparse inverse covariance by applying a penalization to the absolute values of the coefficients in the matrix. This results in a sparse network where only the most important edges (indicating partial correlations) are retained. The selection of the models was based on the Bayesian Information Criterion (BIC). All the networks were generated using the "qgraph" package in R, version 4.1.2.

4.4.6.2 Block sparse Partial Least Square-Discriminant Analysis

We employed the block sPLS-DA based on the DIABLO framework[193], to integrate and analyze multiple datasets available in our study, namely the microbiome dataset, the circulating biomarkers dataset, the GE dataset and the dataset on diet and lifestyle.

Mathematically, this method decomposes each data block into a product of two matrices: a matrix of loadings and a matrix of latent variables (scores). The primary objective is to identify latent variables, which are linear combinations of the original variables, which maximize the covariance between the data blocks and the categorical response variable (which, in our study, is the treatment arm and the occurrence of a clinical event).

Unlike traditional PLS-DA, block sPLS-DA introduces a sparsity constraint on the loadings, which allows the selection of a subset of the most discriminative variables from each block.
The scores from each latent variable are then included into a classifier – in our case, linear discriminant analysis - to classify samples into distinct groups.
The model is based on a design matrix, which describes whether the datasets in each block should be correlated, or not. In our study, we estimated the correlation between each pair of blocks empirically, by executing PLS on each block pair. The models were performed including the *clr*-transformed data in the microbiome block and in the GE block.
To ensure model robustness and determine the optimal number of variables in each component, we utilized cross-validation.

The analysis was conducted using the *mixOmics* package[194] in R, with parameters set based on cross-validation to optimize the number of components and the sparsity level.

4.4.7 Event-Free Survival analysis

For the Event-Free Survival (EFS) analysis, we included follow-up data collected until June 2022. Time-to-event was calculated as the time difference between the date of randomization and the date of first clinical event. For the patients who did not experience any clinical events, the time was calculated as the time between the day of randomization and the day of last visit. Because of the relative short follow-up period (median=3.7 years), we considered as clinical events not only tumour progression and death, but also colorectal adenomas and polyps. However, we distinguished between colorectal events, which included tumour relapses, adenomas, polyps and deaths, and any clinical events, which included also other tumours. The median follow-up was calculated as the median time of the patients who did not experience events. Comparisons in EFS by groups were carried out using Kaplan-Meier estimator and tested with log-rank test. Multivariable Cox proportional-hazards models were employed to estimate the risk of event in terms of hazard ratios (HRs); 95% confidence intervals were also provided.

All statistical analyses were performed using R version 4.1.2 and SAS 9.4.

4.5 Results

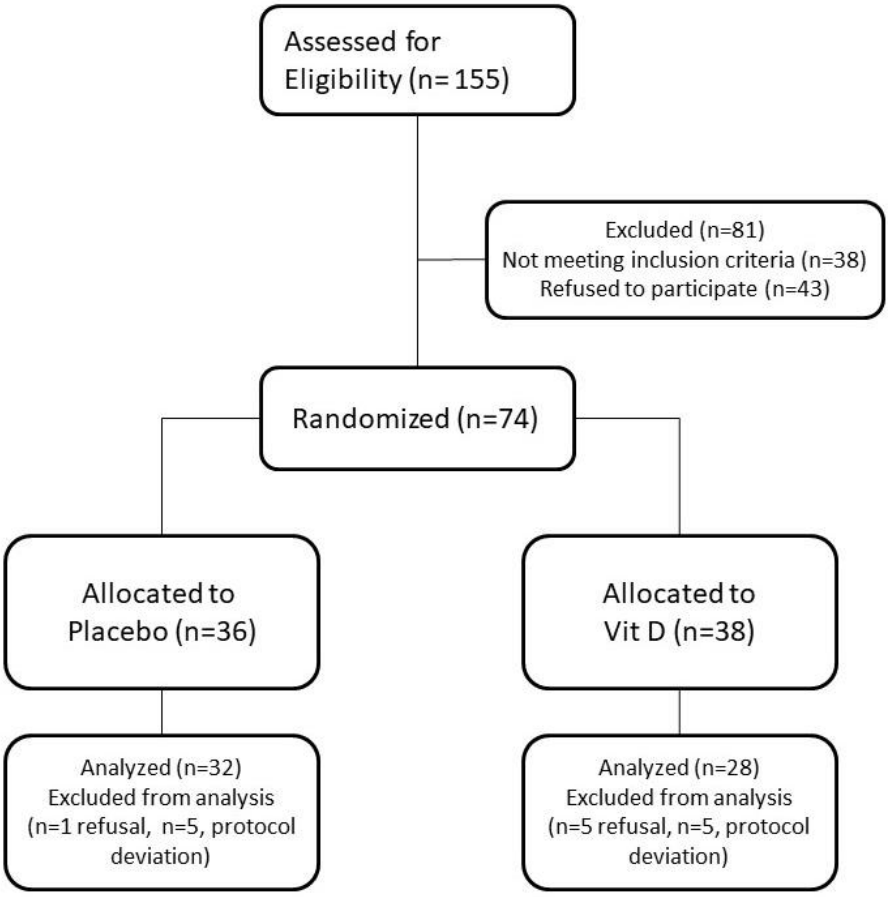


Figure 4.5. Overall study design flowchart and selection of patients included in the analysis. Vit D=Vitamin D.

We enrolled 74 patients in the trial, with 36 allocated to the placebo group and 38 to the vitD supplementation group (Figure 4.5). Notably, 85% of the participants were compliant to the treatment, with 77% consuming over 83% of the prescribed vitD/placebo dosage.

A summary of the clinical and demographic characteristics by treatment arm is provided in Table 4.2.

Table 4.2. Baseline characteristics of the study population by treatment arm

	Placebo (N=36)	Vitamin D (N=38)	P-value*
Chemotherapy[+], n (%)			
No	18 (50.0%)	18 (47.4%)	1
Yes	18 (50.0%)	20 (52.6%)	
Age, median [Q1, Q3]	61.7 [54.4, 67.1]	62.4 [53.9, 67.2]	0.86
Sex, n (%)			
Female	19 (52.8%)	15 (39.5%)	0.36
Male	17 (47.2%)	23 (60.5%)	
Body Mass Index, median [Q1, Q3]	26.2 [23.5, 27.9]	26.4 [23.9, 29.4]	0.45
Stage, n (%)			
0	2 (5.6%)	2 (5.3%)	1
I	11 (30.6%)	11 (28.9%)	
II	11 (30.6%)	12 (31.6%)	
III	10 (27.8%)	11 (28.9%)	
Missing	2 (5.6%)	2 (5.3%)	
pT, n (%)			
T0	2 (5.6%)	2 (5.3%)	0.59
T1	5 (13.9%)	2 (5.3%)	
T2	10 (27.8%)	12 (31.6%)	
T3	16 (44.4%)	20 (52.6%)	
T4	3 (8.3%)	1 (2.6%)	
Missing	0 (0%)	1 (2.6%)	
pN, n (%)			
N0	26 (72.2%)	26 (68.4%)	1
N1	7 (19.4%)	8 (21.1%)	
N2	3 (8.3%)	3 (7.9%)	
Missing	0 (0%)	1 (2.6%)	
Site, n (%)			
Left colon	17 (47.2%)	14 (36.8%)	0.23
Right colon	9 (25.0%)	17 (44.7%)	
Rectum	7 (19.4%)	5 (13.2%)	
Missing	3 (8.3%)	2 (5.3%)	
Histotype, n (%)			
Adenocarcinoma	26 (72.2%)	28 (73.7%)	0.4
Multiple histotypes	10 (27.8%)	6 (15.8%)	
Other	0 (0%)	1 (2.6%)	
Missing	0 (0%)	3 (7.9%)	
Grade, n (%)			
G1	2 (5.6%)	1 (2.6%)	0.5
G2	20 (55.6%)	19 (50.0%)	
G3	5 (13.9%)	11 (28.9%)	
Unknown	6 (16.7%)	5 (13.2%)	
Missing	3 (8.3%)	2 (5.3%)	
Radiotherapy°, n (%)			
No	32 (88.9%)	30 (78.9%)	0.55
Yes	4 (11.1%)	7 (18.4%)	
Missing	0 (0%)	1 (2.6%)	

Q1=first quartile. Q3=third quartile.

*p-values derived from Wilcoxon rank-sum test for numerical variables and from Chi-square test (or Fisher exact test, where appropriate) for categorical variables.

+ Chemotherapy was administered in both adjuvant or neoadjuvant settings.

° Radiotherapy was administered only in a neoadjuvant setting.

4.5.1 Vitamin D supplementation and circulating biomarkers

Levels of 25(OH)D significantly increased in the supplemented group, reaching a median concentration of 39.0 ng/ml (Interquartile Range (IQR): 34.4-43.2 ng/ml) at follow-up ($p < 0.001$). All the patients in the vitD supplementation group positively responded to the treatment, achieving vitD sufficiency (intended as 25(OH)D ng/ml>30) by the end of the study.

No significant change was observed in the placebo group ($p = 0.497$), although 25% (9/36) of the patients reached vitD sufficiency – defined as 25(OH)D>30 ng/ml – by the end of the treatment period (Third quartile (Q3): 29.8 ng/ml) (Table 4.3).

Table 4.3. Distribution of 25(OH)D levels by timepoint and treatment arm

	Baseline 25(OH)D	Post 25(OH)D	Change in 25(OH)D (Post-Baseline)	P-value*
All patients (N=74)				
Placebo, n=36				
Median [IQR]	23.8 [15.9, 25.9]	20.7 [14.6, 29.8]	0.70 [-2.91, 4.50]	0.497
Missing	0	1 (2.8%)	1 (2.8%)	
Vitamin D supplementation, n=38				
Median [IQR]	21.2 [14.5, 26.0]	39.0 [34.4, 43.2]	18.5 [10.8, 25.8]	<0.001
Missing	0	3 (7.9%)	3 (7.9%)	

**p-values* derived from Wilcoxon signed-rank test for paired data that compared baseline 25(OH)D with post-treatment 25(OH)D within each group; 25(OH)D = 25-hydroxyvitamin D; IQR = Interquartile range. Interquartile range is reported as [First quartile (Q1) – Third quartile (Q3)]. 25(OH)D values were obtained from the post-enrollment evaluation on serum samples. For the enrolment, self-reported data were considered.

VDBP also significantly increased following vitD supplementation ($p=0.033$), while no other circulated biomarker either increased or decreased post-supplementation (Table 4.4).

Table 4.4 Distribution of the change (post-treatment – baseline) in circulating biomarkers by treatment arm

	Placebo (N=36)	Vitamin D (N=38)	P-value
VDBP change (µg/mL), median [Q1, Q3]	-9.00 [-34.5, 17.5]	9.00 [-11.0, 45.5]	**0.033**
Missing	1 (2.8%)	3 (7.9%)	
25(OH)D change (ng/mL), median [Q1, Q3]	0.700 [-2.91, 4.50]	18.5 [10.8, 25.8]	**<0.001**
Missing	1 (2.8%)	3 (7.9%)	
Adiponectin change (µg/mL), median [Q1, Q3]	0.261 [-1.06, 1.86]	0.349 [-0.586, 1.32]	1
Missing	1 (2.8%)	3 (7.9%)	
Leptin change (ng/mL), median [Q1, Q3]	0.953 [-1.03, 8.30]	-0.195 [-2.15, 7.93]	0.632
Missing	1 (2.8%)	3 (7.9%)	
IL-10 change (pg/mL), median [Q1, Q3]	0.0900 [-0.220, 0.680]	0.110 [-0.155, 0.525]	0.747
Missing	1 (2.8%)	3 (7.9%)	
IL-6 change (pg/mL), median [Q1, Q3]	0.0600 [-0.525, 0.735]	0.0100 [-0.409, 1.34]	0.991
Missing	1 (2.8%)	3 (7.9%)	
TNFα change (pg/mL), median [Q1, Q3]	0.500 [-0.550, 1.31]	0.100 [-0.650, 0.750]	0.213
Missing	1 (2.8%)	3 (7.9%)	
CCL2/MCP1 change (pg/mL), median [Q1, Q3]	4.00 [-28.0, 60.5]	4.00 [-17.5, 38.0]	0.916
Missing	1 (2.8%)	3 (7.9%)	
CD27 change (pg/mL), median [Q1, Q3]	-126 [-505, 210]	138 [-238, 627]	0.0821
Missing	1 (2.8%)	3 (7.9%)	
CD40 Ligand change (pg/mL), median [Q1, Q3]	-1760 [-3120, -347]	-2400 [-3540, 740]	0.963
Missing	1 (2.8%)	3 (7.9%)	
CXCL6/GCP-2 change (pg/mL), median [Q1, Q3]	-3440 [-5720, -1870]	-3830 [-5090, -2310]	0.709
Missing	1 (2.8%)	3 (7.9%)	
Galectin-3 change (pg/mL), median [Q1, Q3]	1820 [1490, 2060]	1820 [1460, 2260]	0.492
Missing	1 (2.8%)	3 (7.9%)	
IL-8/CXCL8 change (pg/mL), median [Q1, Q3]	-2000 [-2240, -1590]	-2080 [-2470, -1610]	0.338
Missing	1 (2.8%)	3 (7.9%)	
CD40 change (pg/mL), median [Q1, Q3]	-4.00 [-38.5, 14.5]	10.0 [-10.0, 24.5]	0.188
Missing	1 (2.8%)	3 (7.9%)	
CXCL2/GROβ change (pg/mL), median [Q1, Q3]	-11.0 [-118, 43.5]	1.00 [-63.5, 85.0]	0.198
Missing	1 (2.8%)	3 (7.9%)	
Galectin-1 change (pg/mL), median [Q1, Q3]	-1130 [-2560, 706]	47.0 [-1330, 1960]	0.146
Missing	1 (2.8%)	3 (7.9%)	
Galectin-9 change (pg/mL), median [Q1, Q3]	169 [-330, 825]	171 [-554, 773]	0.452
Missing	1 (2.8%)	3 (7.9%)	
IL-7 change (pg/mL), median [Q1, Q3]	-1.13 [-2.77, 1.06]	0 [-2.08, 3.25]	0.247
Missing	1 (2.8%)	3 (7.9%)	

p-values derived by Wilcoxon rank-sum test. Q1= first quartile. Q3=third quartile.
Change is defined as post-treatment – baseline values of the circulating biomarkers.

However, increasing levels of 25(OH)D were significantly and positively correlated with increasing levels of Galectin-3 (p=0.03) and Galectin-9 (p=0.04), after adjustment for treatment arm, sex/gender and age.

We also observed differences by weight status, with adiponectin significantly increasing at increasing levels of 25(OH)D only in normal-weight individuals (p=0.047. Table 4.5; Figure 4.6). Change in IL-8/CXCL8 was also significantly and inversely correlated with increasing levels of 25(OH)D in normal weight only (p=0.03), however a borderline significant association was also observed in the whole sample (p=0.06; Table 4.5).

Table 4.5 Relationship between change in 25(OH)D and change in circulating biomarkers overall and by weight status.

	25(OH)D change (ng/mL)			25(OH)D change (ng/mL)			25(OH)D change (ng/mL)		
Biomarker	**Beta***	**95% CI**	**p-value****	**Beta***	**95% CI**	**p-value****	**Beta***	**95% CI**	**p-value****
		All (n=74)			*Normal weight (n=28⁺)*			*Overweight (n=44)*	
VDBP change (µg/mL)	0.31	[-0.92; 1.54]	0.62	0.26	[-2.22; 1.67]	0.77	0.47	[-1.15, 2.09]	0.56
Adiponectin change (µg/mL)	0.07	[-0.01; 0.15]	0.08	0.12	[0.04; 0.20]	**0.01**	0.04	[-0.07, 0.15]	0.46
Leptin change (ng/mL)	0.06	[-0.28; 0.41]	0.72	-0.05	[-0.69; 0.60]	0.88	0.09	[-0.35, 0.54]	0.67
IL-10 change (pg/mL)	0.01	[-0.02; 0.05]	0.46	0.03	[-0.03; 0.09]	0.28	0.01	[-0.03, 0.06]	0.62
IL-6 change (pg/mL)	0.05	[-0.21; 0.30]	0.71	0.21	[-0.41, 0.83]	0.49	-0.02	[-0.15, 0.12]	0.78
TNFα change (pg/mL)	0.04	[-0.01; 0.09]	0.13	0.02	[-0.08; 0.12]	0.67	0.05	[-0.003, 0.11]	0.06
CCL2/MCP1 change (pg/mL)	0.31	[-2.76; 3.38]	0.84	-1.57	[-9.06; 5.92]	0.67	1.47	-[0.748, 3.69]	0.19
CD27 change (pg/mL)	16.6	[-3.74; 37.0]	0.11	14.3	[-17.2; 45.8]	0.36	9.19	[-19.1, 37.5]	0.52
CD40 Ligand change (pg/mL)	-6.6	[-89.8; 76.6]	0.87	12.6	[-128.5; 153.8]	0.85	-54.5	[-171, 62.4]	0.35
CXCL6/GCP-2 change (pg/mL)	-4.73	[-69.8; 60.3]	0.88	-17.4	[-99.1; 64.2]	0.66	1.27	[-102.1, 104.6]	0.98
Galectin-3 change (pg/mL)	13.0	[1.0; 24.9]	**0.03**	23.8	[2.88; 44.6]	**0.03**	4.74	[-11.7, 21.2]	0.56
IL-8/CXCL8 change (pg/mL)	-11.2	[-23.1; 0.67]	0.06	-21.5	[-40.3; -2.74]	**0.03**	-3.95	[-21.7, 13.8]	0.65
CD40 change (pg/mL)	-0.09	[-1.19; 1.00]	0.87	-0.09	[-2.24; 2.07]	0.94	-0.29	[-1.65, 1.06]	0.66
CXCL2/GROβ change (pg/mL)	1.02	[-3.09; 5.14]	0.62	1.26	[-3.87; 6.38]	0.62	2.13	[-4.41, 8.68]	0.51
Galectin1 change (pg/mL)	-42.1	[-129.0; 44.9]	0.34	-85.9	[-178, 5.77]	0.07	-14.7	[-142.4, 112.9]	0.82
Galectin9 change (pg/mL)	0.7	[1.99; 59.5]	**0.04**	15.2	[-22.8; 53.2]	0.42	33.5	[-12.6, 79.6]	0.15
IL-7 change (pg/mL)	0.04	[-0.06; 0.13]	0.42	0.08	[-0.08; 0.25]	0.31	0.01	[-0.13, 0.14]	0.94

*Beta regression coefficient and **p-value of the change in 25(OH)D levels (post-treatment – baseline) included as a covariate in multivariable linear regression models, adjusted for sex/gender, treatment arm and age and having the change in levels (post-treatment – baseline) of each circulating biomarke as outcome.

⁺ Two underweight patients (BMI<18.5) were not included.

[1]CI = Confidence Interval.

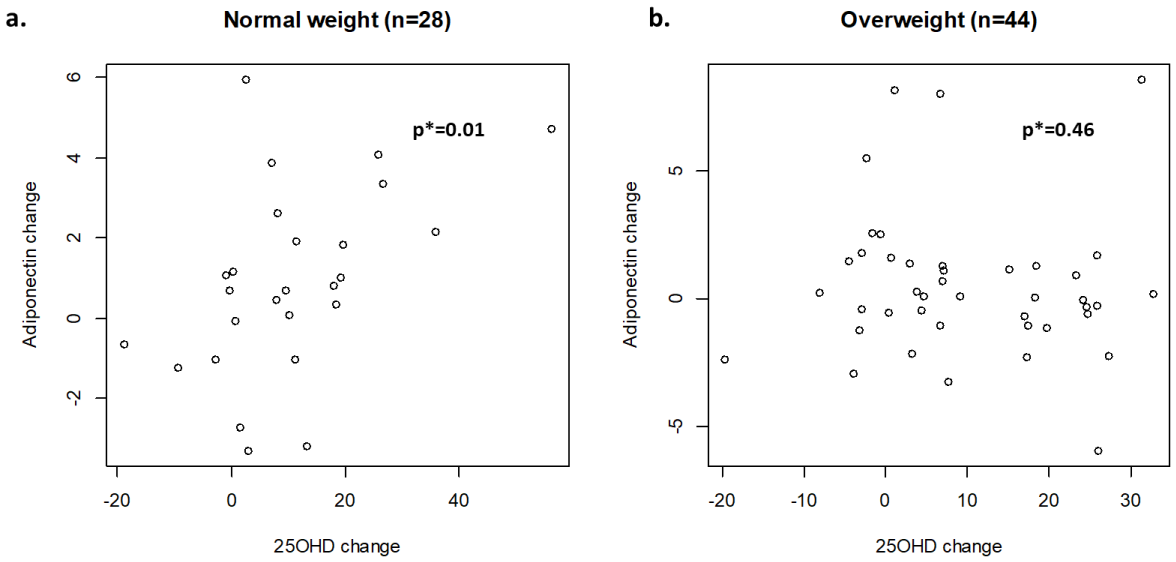


Figure 4.6. Scatterplots of change in 25(OH)D levels (post-treatment – baseline) and change in adiponectin levels (post-treatment – baseline) for (**a**) normal-weight (BMI.18.5-25) and (**b**) overweight (BMI>25) individuals.
*p-values were obtained from the multivariable models in Table 4.5.

Recent literature has indeed increasingly pointed out a relationship between obesity and vitD status, with emerging evidence suggesting a potential modifier effect of BMI on the response to vitD supplementation.

In this context, the VITAL trial, a randomized, double-blind, placebo-controlled 2 × 2 factorial study involving supplementation of vitD3 at 2000 IU/d and marine ω-3 fatty acids, 1 g/d and enrolling more than 16,000 healthy participants, has provided pivotal data.

In the trial, they found that vitD supplementation did not reduce the incidence of cardiovascular and cancer events in the whole sample. However, they observed a significant lower incidence of invasive cancer events in the vitD supplemented individuals compared to placebo, but only in the subgroup of normal-weight individuals (BMI<25)[124].

Moreover, the authors recently published results from a cohort study nested within the trial, including a subset of 2742 participants with available blood sample at 2-year follow-up. Within this cohort, vitD supplementation was significantly correlated with increasing levels of total 25(OH)D, 25(OH)D3, free vitD, and bioavailable vitD compared to placebo. However, a significant interaction between the treatment and BMI was observed, with these increases being significantly lower in participants with higher BMI categories[131].

Interestingly, we observed similar results in our study. We found a significant interaction between baseline BMI and vitD supplementation on 25(OH)D (p=0.0498), with post-treatment 25(OH)D levels decreasing with increasing BMI in the supplemented group. The model was adjusted for season of blood draw and baseline 25(OH)D. Sex/gender and age were not significantly associated

with post-treatment 25(OH)D and were thus excluded from the model. A borderline significant interaction between the supplementation and BMI was also observed on the 1-year change in 25(OH)D levels (p=0.085), which became significant after adjusting for baseline 25(OH)D levels (p=0.0498) (**Figure 4.7**).

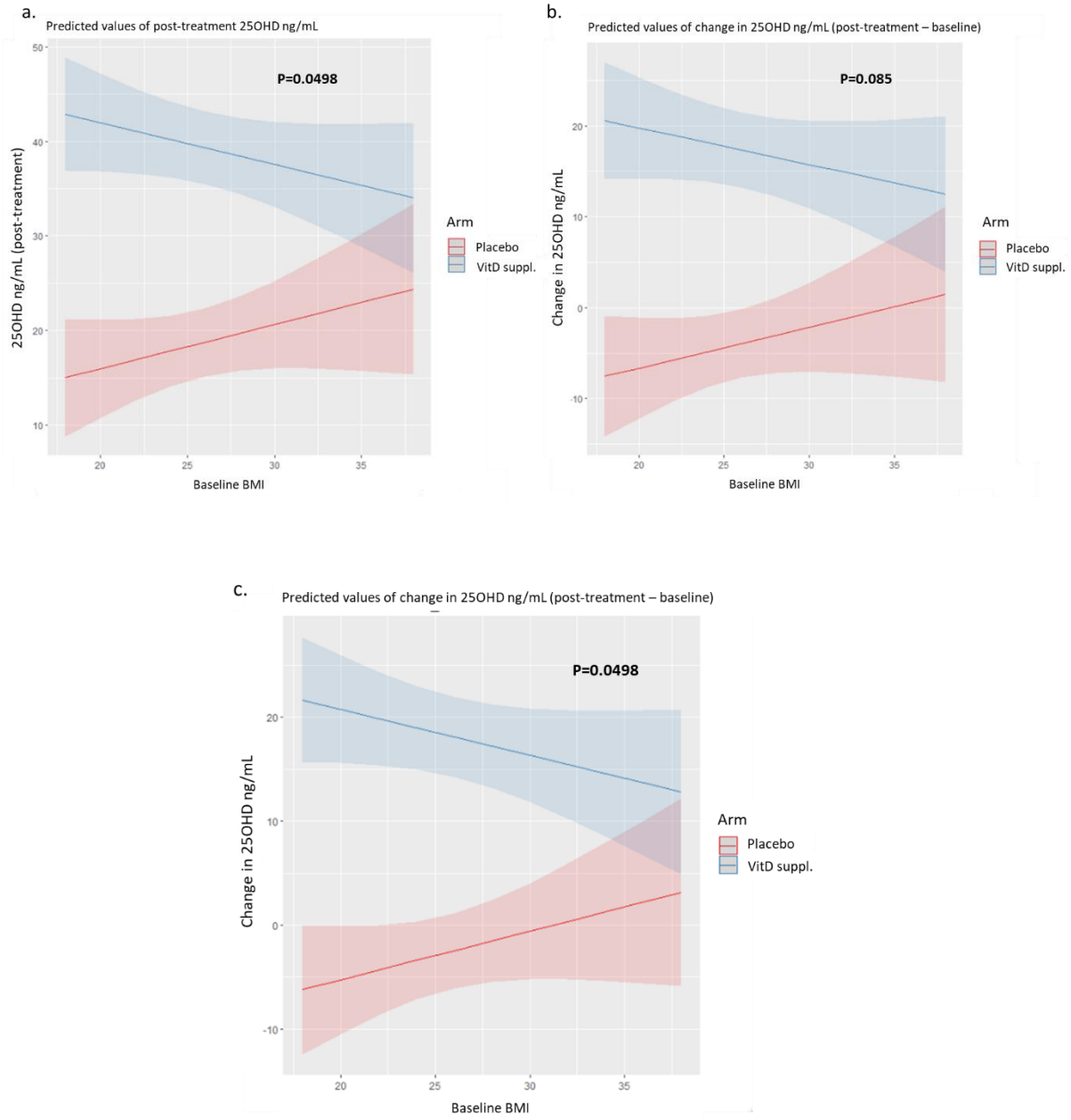


Figure 4.7. Predicted values of post-treatment 25(OH)D values and change in 25(OH)D values (post-treatment – baseline) from multivariable regression models including the interaction between the treatment arm and baseline BMI. Specifically:

a. Model: 25(OH)D levels (post) ~ BMI*Arm + season of blood draw + 25(OH)D levels (baseline).

b. Model: Change in 25(OH)D levels ~ BMI*Arm + season of blood draw.

c. Model: Change in 25(OH)D levels ~ BMI*Arm + season of blood draw + 25(OH)D levels (baseline).

4.5.2 Analysis of diet and lifestyle

For both timepoints, we built a score on lifestyle and diet habits for each patient. The score was built following the recommendations on cancer prevention by WCRF/AICR. It includes information on weight status, physical activity, smoking, alcohol and high-risk diet (see *Scoring of diet and lifestyle based on WCRF/AICR recommendations*). The score ranged from a minimum of 0 and a maximum of 6, with 0 indicating absence of adherence to the recommendations and 6 indicating complete adherence.

No differences by treatment arm were observed for the diet/lifestyle score both at baseline (p=0.10; Wilcoxon rank-sum test) and at the end of the treatment (p=0.33; Wilcoxon rank-sum test). Moreover, no clinical characteristic of the patients was associated with the score, although females tended to be more adherent to the recommendations than men (t-test: p=0.04; Wilcoxon rank-sum test: p=0.12) (Figure 4.8).

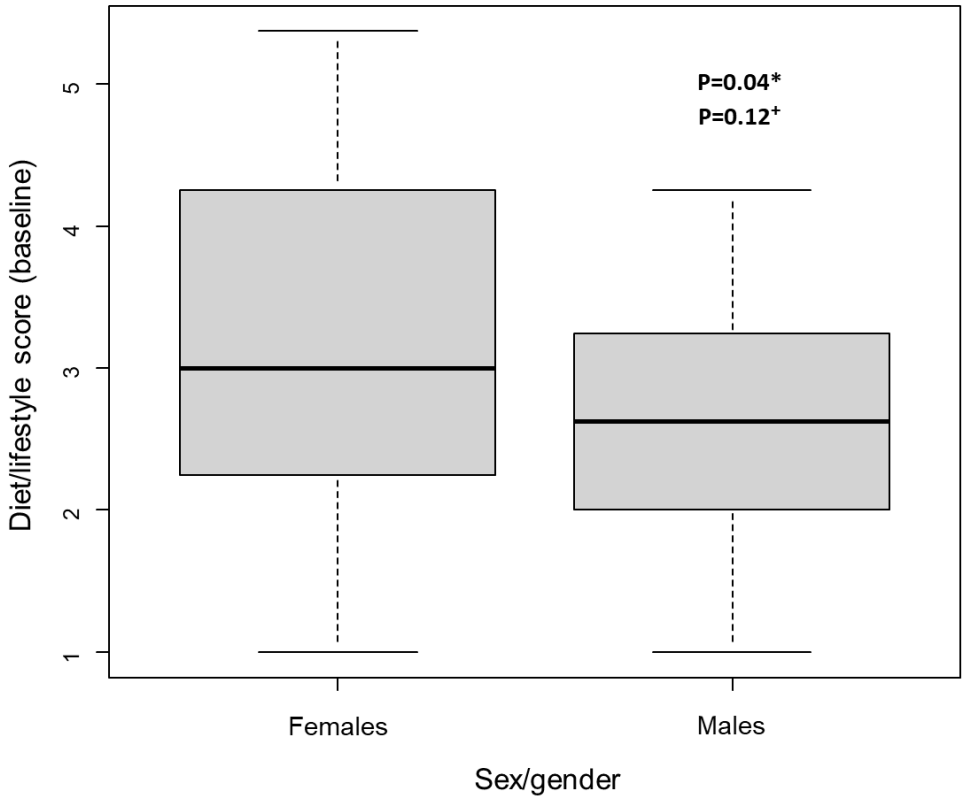


Figure 4.8. Distribution of the diet/lifestyle score at baseline by sex/gender.
*p-value derived from t-test on the means.
+p-value derived from Wilcoxon rank-sum test on the ranks.

At both timepoints, adherence to WCRF recommendations was inversely correlated with levels of leptin, TNF-alpha, Galectin-9 and Galectin-1. IL-10 was positively correlated with the score at baseline, while CD40 was inversely correlated with the post-treatment score. Interestingly, a significant positive correlation between 25(OH)D levels and higher levels of adherence to WCRF recommendations was also observed, but only at the end of the treatment (p=0.05) (Table 4.6).

Table 4.6 Correlations between the diet/lifestyle score and circulated biomarkers at both time points

	Diet/lifestyle score *(baseline)**		
Biomarker *(baseline)*	**Beta**	**95% CI[1]**	**p-value**
VDBP (µg/mL)	-4.07	[-17.5; 9.38]	0.55
25(OH)D (ng/mL)	0.33	[-1.54; 2.19]	0.73
Adiponectin (µg/mL)	0.51	[-0.951; 1.97]	0.49
Leptin (ng/mL)	-7.93	[-14.3; -1.54]	**0.02**
IL-10 (pg/mL)	0.59	[0.025; 1.16]	**0.04**
IL-6 (pg/mL)	-0.52	[-1.14; 0.097]	0.10
TNFα (pg/mL)	-0.82	[-1.43; -0.211]	**0.01**
CCL2/MCP1 (pg/mL)	0.88	[-98.0; 99.8]	0.99
CD27 (pg/mL)	-263.00	[-781; 256]	0.32
CD40 Ligand (pg/mL)	-425.00	[-992; 141]	0.14
CXCL6/GCP-2 (pg/mL)	-15.70	[-69.2, 37.8]	0.56
Galectin-3 (pg/mL)	-61.60	[-168; 44.7]	0.25
IL-8/CXCL8 (pg/mL)	-0.89	[-2.00; 0.219]	0.11
CD40 (pg/mL)	-18.80	[-38.2; 0.653]	0.06
CXCL2/GROβ (pg/mL)	27.40	[-37.9; 92.6]	0.41
Galectin-1 (pg/mL)	-1721.00	[-3,081; -361]	**0.01**
Galectin-9 (pg/mL)	-624.00	[-1,189; -59.0]	**0.03**
IL-7 (pg/mL)	0.12	[-1.12; 1.35]	0.85

	Diet/lifestyle score *(post)***		
Biomarker *(post)*	**Beta**	**95% CI[1]**	**p-value**
VDBP (µg/mL)	4.67	[-8.55; 17.9]	0.48
25(OH)D (ng/mL)	2.31	[0.017; 4.61]	**0.05**
Adiponectin (µg/mL)	1.31	[-0.220; 2.84]	0.09
Leptin (ng/mL)	-10.30	[-18.6; -1.95]	**0.02**
IL-10 (pg/mL)	0.39	[-0.317; 1.10]	0.27
IL-6 (pg/mL)	-1.36	[-3.91; 1.18]	0.29
TNFα (pg/mL)	-1.55	[-2.37; -0.722]	**<0.01**
CCL2/MCP1 (pg/mL)	-12.70	[-146; 121]	0.85
CD27 (pg/mL)	-483.00	[-1,007; 40.5]	0.07
CD40 Ligand (pg/mL)	-260.00	[-794; 274]	0.33
CXCL6/GCP-2 (pg/mL)	-25.50	[-68.9; 17.9]	0.24
Galectin-3 (pg/mL)	-45.00	[-154; 64.5]	0.42
IL-8/CXCL8 (pg/mL)	-1.19	[-2.82; 0.445]	0.15
CD40 (pg/mL)	-24.10	[-40.8; -7.52]	**0.01**
CXCL2/GROβ (pg/mL)	-1.78	[-73.4; 69.8]	0.96
Galectin-1 (pg/mL)	-2167.00	[-3,769; -566]	**0.01**
Galectin-9 (pg/mL)	-669.00	[-1,241; -97.2]	**0.02**
IL-7 (pg/mL)	-0.66	[-2.12; 0.799]	0.37

*Estimates derived from multivariable linear regression models including the diet/lifestyle score at baseline, adjusted for age and sex/gender and including the baseline levels of each circulating biomarker.

**Estimates derived from multivariable linear regression models including the diet/lifestyle score at the end of the treatment period, adjusted for age, sex/gender and treatment arm and including the post-treatment levels of each circulating biomarker

[1]CI = Confidence Interval.

4.5.3 Analysis of the gut microbiome

Out of the 74 patients enrolled in the trial, gut microbiome data was available for 65 patients at both time points. Five drop-out patients were excluded, resulting in 60 patients. Of these, 32 were in the placebo group and 28 were in the supplementation group. Relative abundances of 980 taxa were available at both timepoints for each patient.

4.5.3.1 Alpha diversity

Alpha diversity was calculated for each patient and for each timepoint using the Shannon index, which accounts for both richness and evenness of the species.

Overall, we observed no differences by treatment arm (p=0.66) in change of alpha diversity (calculated as post treatment – baseline alpha diversity), after adjusting for sex/gender, age, season of blood draw, baseline 25(OH)D levels and diet/lifestyle score. However, adherence to WCRF recommendations was positively correlated with the change in alpha diversity (p=0.04) (Table 4.7).

Table 4.7 Association between vitD supplementation and change in alpha diversity from multivariable regression analysis adjusted for confounders

	Shannon Index change		
Characteristic	**Beta**	**95% CI[1]**	**p-value**
Arm (Vitamin D vs Placebo)	-0.09	[-0.29; 0.10]	0.34
Diet/Lifestyle score (baseline)	0.11	[0.01; 0.21]	**0.03**
Sex/gender (Male vs Female)	0.06	[-0.16; 0.28]	0.60
Age	-0.01	[-0.02; 0.002]	0.13
Season of blood draw			
Autumn	—	—	
Summer	0.25	[-0.05; 0.55]	0.11
Winter	0.04	[-0.25; 0.32]	0.79
Spring	0.01	[-0.26; 0.28]	>0.9
Baseline 25(OH)D	-0.005	[-0.02; 0.01]	0.48

[1] CI = Confidence Interval

Estimates derived from multivariable linear regression model.

When looking at weight status, we observed a significant interaction between overweight status (BMI>25) and change in 25(OH)D on the change in alpha diversity, with increasing alpha diversity at increasing levels of 25(OH)D in non-overweight patients, and no relationship in overweight (p_{int}=0.01; pint=0.03 after excluding the two underweight patients) (Table 4.8; Figure 4.9).

Table 4.8 Interaction between vitD supplementation and overweight status on change in alpha diversity from multivariable regression analysis adjusted for confounders

	Shannon Index change		
Characteristic	**Beta**	**95% CI[1]**	**p-value**
25(OH)D change (ng/mL)	0.01	[0.001; 0.02]	**0.03**
Overweight (Yes vs No)	-0.16	[-0.38; 0.06]	0.14
Sex/gender (Male vs Female)	0.08	[-0.12; 0.27]	0.44
Age	-0.003	[-0.01; 0.01]	0.62
Season of blood draw			
Autumn	—	—	
Summer	0.14	[-0.14; 0.42]	0.31
Winter	0.09	[-0.16; 0.33]	0.48
Spring	0.11	[-0.13; 0.36]	0.34
25(OH)D change (ng/mL)*Overweight	-0.02	[-0.03; 0.00]	**0.01**

Estimates derived from multivariable linear regression model. Interaction between treatment arm overweight status was introduced in the model.

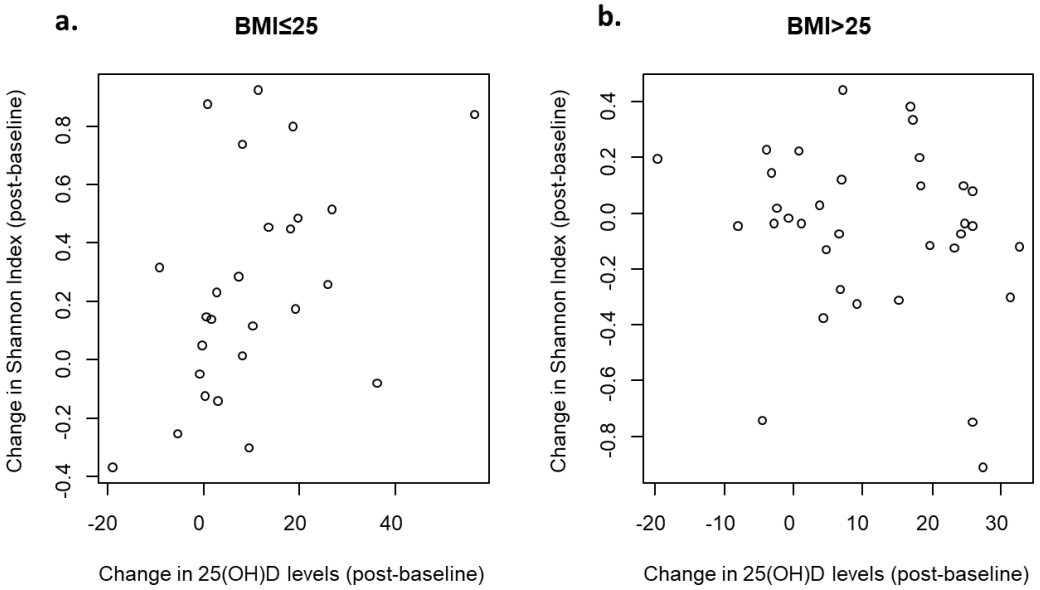


Figure 4.9. Scatterplots of change in 25(OH)D levels (post-treatment – baseline) and change in Shannon Index (post-treatment – baseline) by overweight status. (**a**) BMI≤25. (**b**) BMI>25.

In general, patients who had experienced a more advanced CRC had a lower alpha diversity at baseline. In fact, this decrease in diversity tended to be correlated with increasing pT and advanced tumour stages, suggesting a potential relationship between the severity of the cancer and the diversity of the microbial community. However, at the end of the treatment period, the alpha diversity of these patients increased significantly more than that of patients having experienced early-stage cancer (Table 4.9).

Table 4.9 Estimates of associations between clinical characteristics of patients and alpha diversity

	Shannon Index (baseline)			Shannon index (change)		
Characteristic	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**
Age	0.01	[0.00; 0.02]	0.104	-0.01	[-0.02; 0.004]	0.212
Sex/gender						
Female	—	—		—	—	
Male	0.12	[-0.07; 0.31]	0.217	-0.03	[-0.22; 0.16]	0.734
Stage						
0	—	—		—	—	
I	-0.4	[-0.85; 0.05]	0.085	0.45	[0.01; 0.89]	**0.048**
II	-0.44	[-0.90; 0.01]	0.063	0.61	[0.17; 1.1]	**0.009**
III	-0.42	[-0.88; 0.03]	0.073	0.52	[0.08; 0.96]	**0.024**
pT						
T0	—	—		—	—	
T1	-0.23	[-0.76; 0.29]	0.388	0.34	[-0.17; 0.85]	0.2
T2	-0.45	[-0.90; 0.00]	0.054	0.47	[0.04; 0.91]	**0.037**
T3	-0.44	[-0.88; 0.00]	0.055	0.6	[0.18; 1.0]	**0.008**
T4	-0.45	[-1.0; 0.10]	0.114	0.47	[-0.07; 1.0]	0.093
pN						
N0	—	—		—	—	
N1	0.05	[-0.18; 0.28]	0.688	-0.06	[-0.29; 0.17]	0.637
N2	-0.23	[-0.57; 0.12]	0.205	0.28	[-0.06; 0.62]	0.112
Tumour side						
Left colon	—	—		—	—	
Right colon	-0.1	[-0.32; 0.12]	0.373	0.1	[-0.13; 0.32]	0.404
Rectum	0.12	[-0.16; 0.40]	0.398	0	[-0.29; 0.29]	0.981
Histotype						
Adenocarcinoma	—	—		—	—	
Multiple histotypes	0.08	[-0.14; 0.30]	0.488	-0.07	[-0.30; 0.15]	0.514
Other	-0.1	[-0.85; 0.65]	0.787	-0.4	[-1.1; 0.35]	0.302
Grade						
G1	—	—		—	—	
G2	-0.33	[-0.88; 0.22]	0.24	0.41	[-0.13; 0.94]	0.14
G3	-0.27	[-0.84; 0.30]	0.362	0.29	[-0.26; 0.85]	0.303
Unknown	-0.19	[-0.78; 0.39]	0.518	0.25	[-0.32; 0.82]	0.386
Radiotherapy						
No	—	—		—	—	
Yes	0.02	[-0.17; 0.21]	0.815	-0.06	[-0.33; 0.21]	0.664
Chemotherapy						
No	—	—		—	—	
Yes	0.02	[-0.17; 0.21]	0.815	0.14	[-0.05; 0.32]	0.146

[1] CI = Confidence Interval

Beta regression estimates derived from linear regression models.

Alpha diversity at baseline was also significantly and positively correlated with adiponectin (p=0.029) and Galectin-1 (p=0.017), and inversely correlated with CCL2/MCP1 (p=0.001) and CXCL6/GCP2 (p=0.021). However, no significant correlation was observed between increasing levels of alpha diversity and increasing levels of circulating biomarkers during the study period, and between alpha diversity and the biomarkers at the end of the treatment (Table 4.10).

Table 4.10 Estimates of association between baseline, post-treatment and change in circulating biomarkers with baselin, post-treatment and change in alpha diversity

	Shannon Index (baseline)*			Shannon index (post)**			Shannon index (change)***		
Characteristic⁺	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**	**Beta**	**95% CI[1]**	**p-value**
VDBP (µg/mL)	-0.03	-0.13, 0.07	0.521	-0.05	-0.14, 0.04	0.278	-0.06	-0.17, 0.04	0.226
25(OH)D (ng/mL)	0.01	-0.09, 0.10	0.917	0.01	-0.12, 0.14	0.859	0.13	-0.02, 0.28	0.088
Adiponectin (µg/mL)	0.11	0.01, 0.22	**0.029**	0.07	-0.02, 0.17	0.127	0.02	-0.07, 0.12	0.644
Leptin (ng/mL)	0.05	-0.08, 0.18	0.415	-0.06	-0.18, 0.06	0.327	-0.06	-0.20, 0.07	0.371
IL-10 (pg/mL)	-0.05	-0.14, 0.04	0.319	0.05	-0.03, 0.14	0.229	0.01	-0.13, 0.14	0.935
IL-6 (pg/mL)	0.05	-0.05, 0.15	0.293	0	-0.08, 0.08	0.981	0.04	-0.06, 0.13	0.412
TNF-α (pg/mL)	0.05	-0.05, 0.15	0.279	-0.01	-0.10, 0.08	0.818	0.01	-0.10, 0.12	0.81
CCL2/MCP1 (pg/mL)	-0.14	-0.22, -0.06	**0.001**	-0.07	-0.15, 0.00	0.053	0.04	-0.05, 0.14	0.373
CD27 (pg/mL)	0.1	0.00, 0.20	**0.046**	0.03	-0.06, 0.12	0.506	0.02	-0.09, 0.12	0.756
CD40 Ligand (pg/mL)	-0.01	-0.11, 0.08	0.805	0.02	-0.06, 0.11	0.597	-0.03	-0.13, 0.08	0.613
CXCL6/GCP-2 (pg/mL)	-0.1	-0.19, -0.02	**0.021**	-0.05	-0.12, 0.03	0.236	-0.04	-0.14, 0.05	0.356
Galectin-3 (pg/mL)	-0.04	-0.13, 0.06	0.455	0.03	-0.05, 0.11	0.488	0.01	-0.09, 0.11	0.854
IL-8/CXCL8 (pg/mL)	-0.04	-0.15, 0.07	0.45	0.01	-0.07, 0.10	0.731	0.04	-0.06, 0.13	0.451
CD40 (pg/mL)	0.05	-0.06, 0.17	0.364	0.04	-0.06, 0.14	0.453	-0.02	-0.12, 0.07	0.647
CXCL2/GROβ (pg/mL)	-0.03	-0.13, 0.06	0.473	0.01	-0.07, 0.09	0.817	0.01	-0.10, 0.12	0.809
Galectin-1 (pg/mL)	0.12	0.02, 0.22	**0.017**	-0.02	-0.11, 0.07	0.595	-0.04	-0.14, 0.05	0.351
Galectin-9 (pg/mL)	0.04	-0.08, 0.16	0.508	-0.01	-0.11, 0.08	0.745	-0.02	-0.13, 0.09	0.719
IL-7 (pg/mL)	0.01	-0.09, 0.11	0.851	0.02	-0.06, 0.10	0.617	0.03	-0.07, 0.14	0.554

[1] CI = Confidence Interval

⁺ Biomarker values were scaled.

*Estimates were obtained from multivariable linear regression models including the scaled baseline values of each biomarker as a covariate and the baseline Shannon Index as outcome. Models were adjusted for sex/gender and age.

** Estimates were obtained from multivariable linear regression models including the scaled post-treatment values of each biomarker as a covariate and the post-treatment Shannon Index as outcome. Models were adjusted for sex/gender, age and treatment arm.

*** Estimates were obtained from multivariable linear regression models including the scaled change in values of each biomarker as a covariate and the change in Shannon Index as outcome. Models were adjusted for sex/gender, age and treatment arm.

4.5.3.2 Beta diversity

Beta diversity was first assessed using the Bray-Curtis distance between samples and tested between groups using PERMANOVA. Overall, no differences by treatment arm were observed both at baseline (p=0.84) ad at the end of the treatment (p=0.70).

Differences in baseline microbiota were also tested according to previous chemotherapy. However, no significant differences were observed in either alpha diversity (Shannon index: p=0.64) or beta diversity (p=0.18).

To account for the compositional structure of the microbiome data, we also looked at the Euclidean distance between the *clr*-transformed taxa abundances at follow-up, as suggested by recent literature[266,267]. Differences by groups were tested using PERMANOVA.

No differences in post-treatment microbiome by treatment arm (p=0.83) were found after adjusting for age, sex/gender, season of blood draw, diet/lifestyle score and baseline 25(OH)D levels (Figure 4.10). However, age and sex/gender were significantly associated with microbiome composition (p=0.01 and p=0.04, respectively).

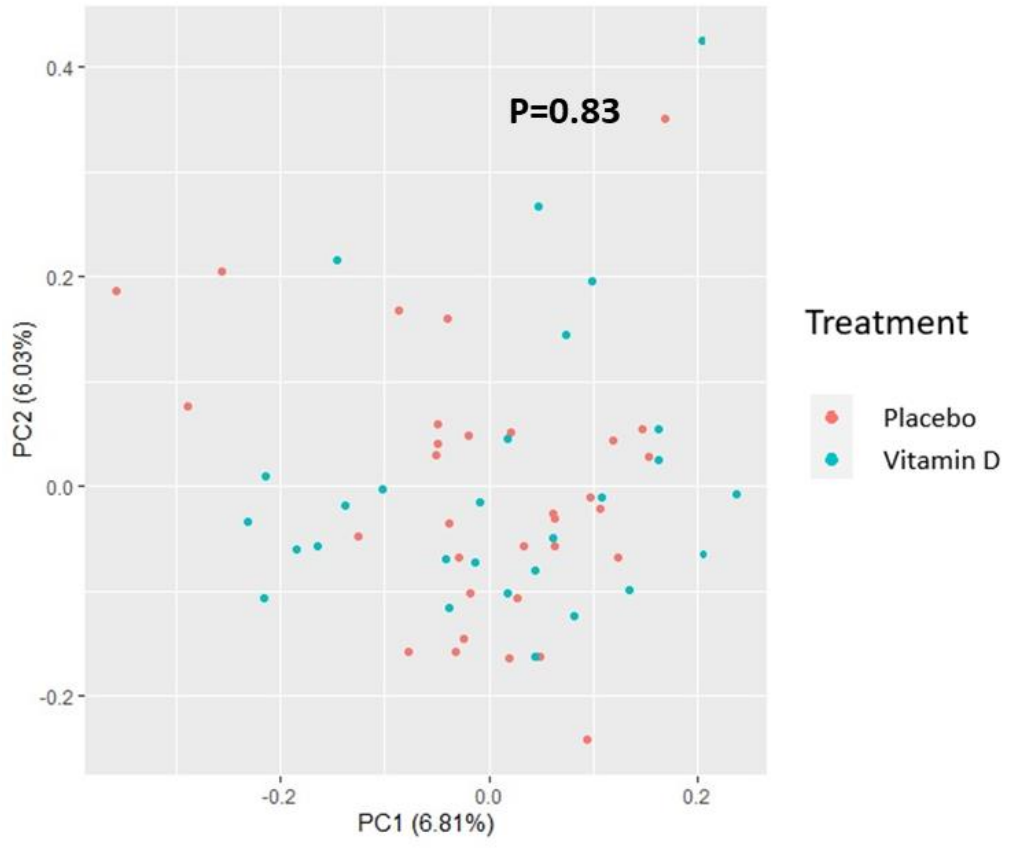


Figure 4.10. Plot of the scaled scores of the first two components from principal component analysis (PCA) run on the *clr*-transformed abundances at follow-up of all the 979 taxa. Data points were colored according to treatment arm (red for the placebo group; blue for the vitamin D supplementation group). P-value was obtained from PERMANOVA.

We then looked at the change in microbiome composition within each patient after the study period. To do this, we applied perturbation difference between the post-treatment and baseline taxa abundances, as described in *Perturbation and power operations in the simplex*. After closuring, we used the *clr*-transformation and calculated the Euclidean distance matrix between data points. As for the post-treatment composition, we did not find any difference in the beta diversity of the change in microbiome by treatment arm (p=0.80). However, we found a borderline significant interaction between the vitD supplementation and BMI categories (overweight vs no overweight: p=0.06, Figure 4.11).

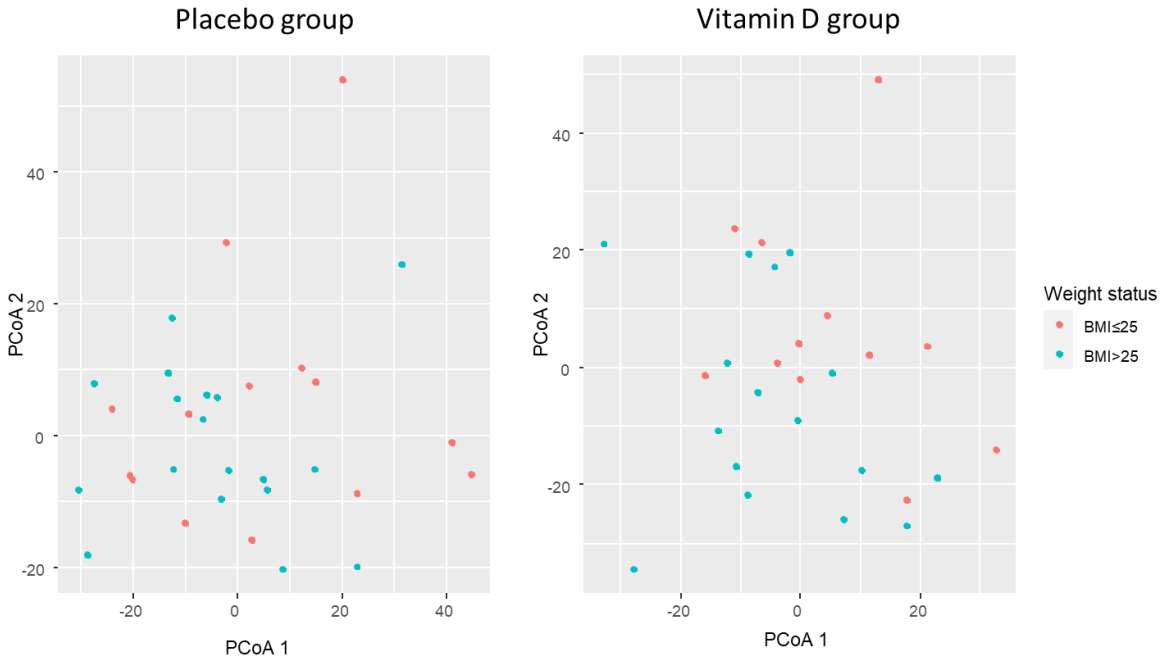


Figure 4.11. Plot of the scaled scores of the first two components from principal coordinates analysis (PCoA) run on the *clr*-transformed perturbation difference of abundances of taxa by treatment arm. Data points were colored according to overweight status (red for BMI≤25; blue for BMI>25).

4.5.3.3 Analysis of the gut microbiome at the end of the treatment

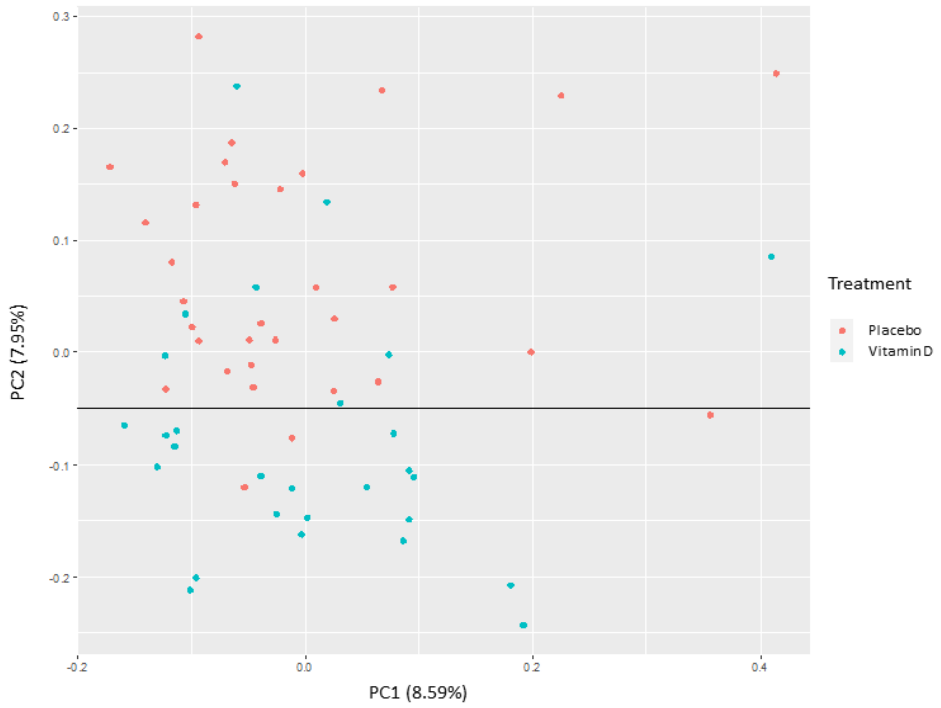
Because the primary endpoint of the study was to identify potential beneficial taxa that changed after one year of vitD supplementation, only the patients with available microbiota at both timepoints (n=65) were considered for analysis[273]. Five drop-out patients in the supplementation group were further excluded. As a result, the final sample consisted of 60 individuals, 28 in the supplemented group and 32 in the placebo group. Except for two patients in the placebo group, all were compliant with the treatment.

As in the overall population, 25(OH)D levels significantly increased in patients who received vitD supplementation, reaching a median post-treatment concentration of 40.4 ng/ml (IQR: 37.4-46.6 ng/ml). No significant change was observed in the placebo group (p = 0.432), although about 25% reached vitD sufficiency by the end of the study (Q3: 31.0 ng/ml).

Relative abundances of 980 taxa were available at both timepoints for each patient. They were *clr*-transformed after zero-value imputation. Out of the total 980 taxa, we first selected 75 whose abundance at follow-up varied significantly between the two treatment groups. Twelve of these taxa were subsequently excluded because they were already significantly unbalanced between the groups at baseline, leaving 63 taxa for statistical analysis. The selection was carried out using *coda-lasso*.

Principal component analysis (PCA) was performed on the clr-transformed abundances at follow-up of the 63 selected taxa. In Figure 4.12a, the scaled PCA scores of the first two components, which together explained about 17% of the total variance, are plotted. As shown in the figure, the second component (PC2) significantly discriminated (Wilcoxon rank-sum test, $p < 0.001$) vitD-supplemented patients from those in the placebo group. Specifically, most of the supplemented patients fell within the component's negative axis (82% of the group), while the majority of patients in placebo (72%) had positive PC2 scores. These values, although not directly interpretable, identified two different microbiome-based clusters of PC2 that well discriminated between the two treatment groups.

a)



b)

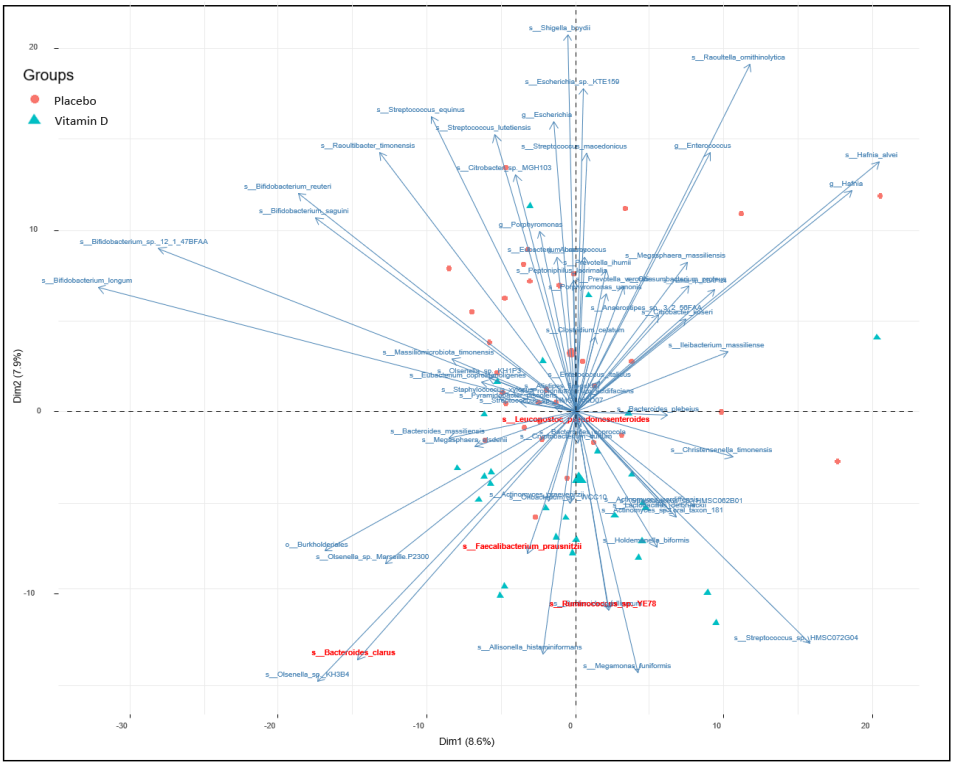


Figure 4.12. a. Plot of the scaled scores of the first two components from principal component analysis (PCA) run on the *clr*-transformed abundances at follow-up of the 63 selected taxa. Data points were colored according to treatment arm (red for the placebo group; blue for the vitamin D supplementation group). **b.** Biplot of the first two components. Vectors represent the 63 taxa. Points represent individual samples. In red are the taxa that significantly associated to either vitamin D supplementation or sufficiency in multivariable regression analysis shown in Figure 4.14.

The biplot in Figure 4.12b and the loadings barplot in Figure 4.13 show the contribution of each of the 63 taxa on PC2: among the taxa that were correlated with the negative side of PC2, i.e., the one characterizing the supplemented patients, we found several species from *Bacteroides* genus, *Faecalibacterium prausnitzii*, - which is a well-known probiotic highly abundant in the gut microbiota of healthy adults - and *Holdemanella biformis*. In contrast, *Shigella boydii* and *Raoultella ornithinolytic*, as well as several species from *Streptococcus* and *Escherichia* genera, were the most correlated with the positive side of PC2, which mostly characterized the placebo group.

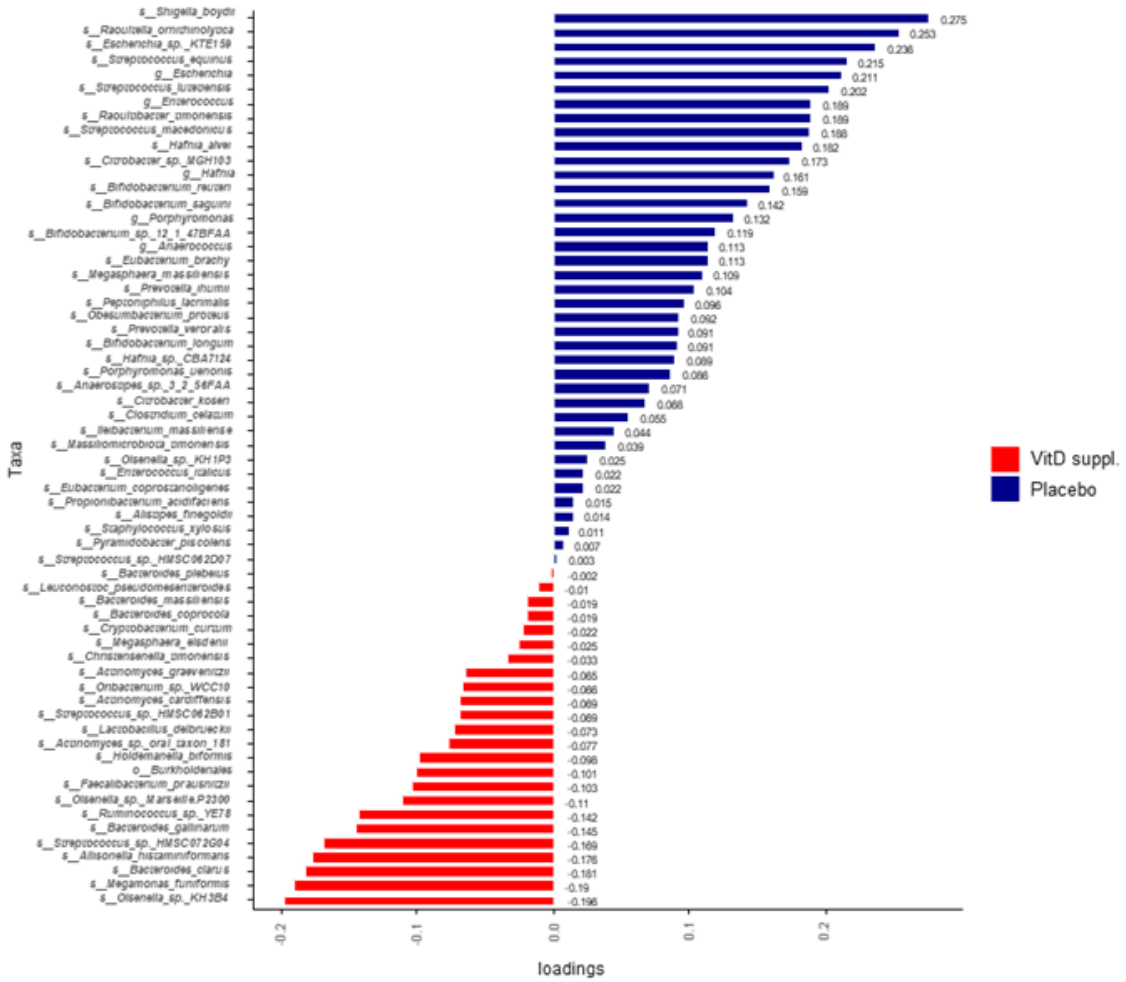


Figure 4.13. Bar plots of the loadings of the 63 taxa on PC2 (second component from PCA run on the post-treatment *clr*-transformed abundances). A positive loading indicates a positive correlation with the component, a negative loading indicates an inverse correlation with the component. Because most of the vitamin D-supplemented patients had negative PC2 scores and most of the patients in the placebo group had positive PC2 scores, a taxa with a negative loading is expected to be more abundant at follow-up in those that were supplemented with vitamin D, while a taxa with a positive loading is expected to be more abundant in those receiving placebo.

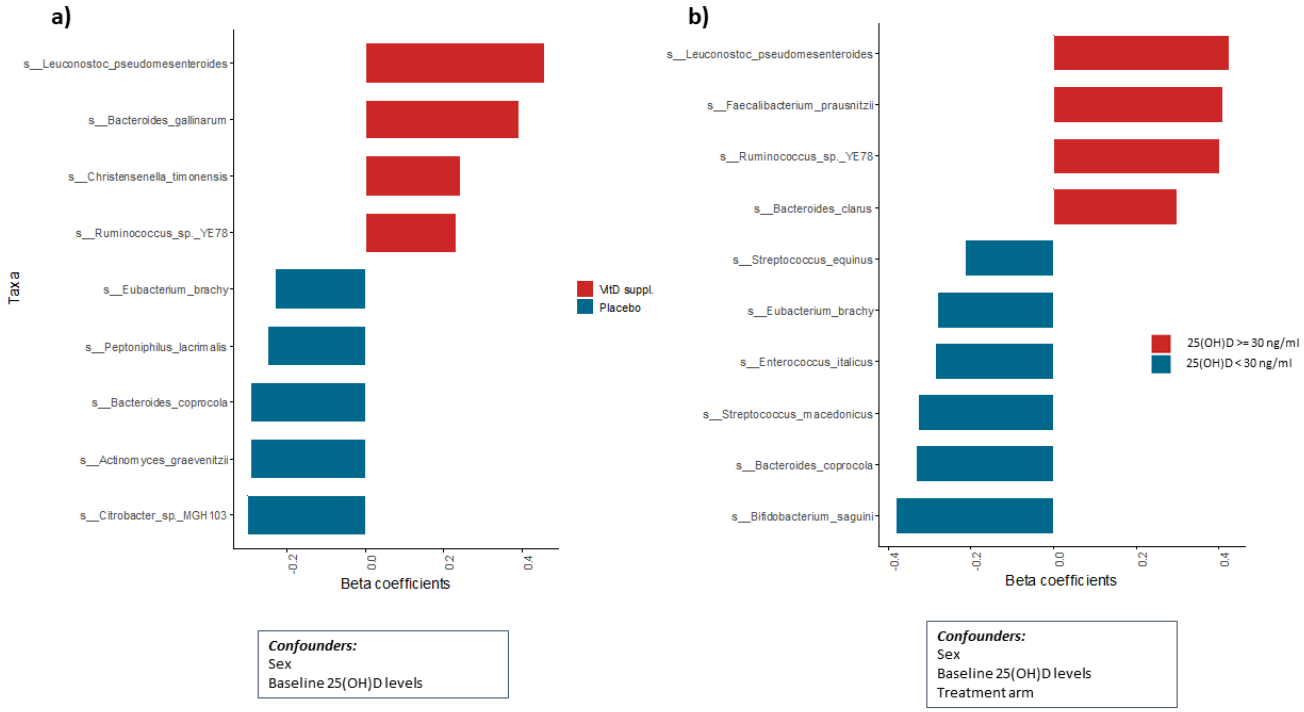


Figure 4.14. Taxa significantly associated with **a.** treatment arm **b.** post-treatment vitamin D sufficiency (25(OH)D ng/ml). For each taxon, results are obtained from a multivariable logistic model including the post-treatment *clr*-transformed abundance of the taxon as covariate and adjusted for confounders. The bar length indicates the significant beta-coefficient of the taxon ($p<0.05$). If positive, the taxon was significantly more abundant in patients **a.** supplemented with vitamin D **b.** reaching vitamin D sufficiency at the end of the treatment. If negative, the taxon was significantly more abundant in patients **a.** in the placebo group **b.** not reaching vitamin D sufficiency at the end of the treatment. 25(OH)D = 25-hydroxy vitamin D.

VitD-supplemented patients had significantly higher abundances of *Leuconostoc pseudomesenteroides*, *Bacteroides gallinarum*, *Christensenella timonensis* and *Ruminococcus YE78* (Figure 4.14a).

Comparing the patients with vitD sufficiency (25(OH)D >= 30 ng/ml, n=36) versus those deficient (n=23) at follow-up, we found that *Leuconostoc pseudomesenteroides* and *Ruminococcus YE78* were also significantly more abundant in vitD-sufficient patients, regardless of treatment arm, together with *Faecalibacterium prausnitzii* and *Bacteroides clarus* (Figure 4.14b). Conversely, *Eubacterium brachy* and *Bacteroides coprocola* were significantly more prevalent in placebo-treated patients and in those not reaching vitD sufficiency at the end of the study (Figure 4.14a-b).

4.5.3.3.1 Taxa-mediated effect of vitamin D supplementation on 25(OH)D levels

Since PCA analysis confirmed differences in the 63 taxa between supplemented and non-supplemented patients, we performed a mediation analysis to see if these taxa also mediated the effect of the supplementation on post-treatment 25(OH)D levels. To do this, we employed the counterfactual approach to mediation analysis, assuming an interaction between vitD supplementation (exposure) and the selected taxa (mediator) on 25(OH)D levels at follow-up (outcome). Both Natural Direct Effect (NDE) and Natural Indirect Effect (NIE) of vitD supplementation on post-treatment 25(OH)D levels were hypothesized and graphically represented using a DAG, with baseline 25(OH)D levels and sex/gender as confounders (Figure 4.15). Because PC2 was the component that best discriminated the supplemented from the non-supplemented, we used it as a proxy for the 63 taxa abundances.

We found that vitD supplementation significantly and directly affected the final 25(OH)D levels (NDE: p<0.0001), but part of its overall effect was significantly mediated by the modulation of the 63 taxa (NIE: p = 0.02) (Figure 4.15).

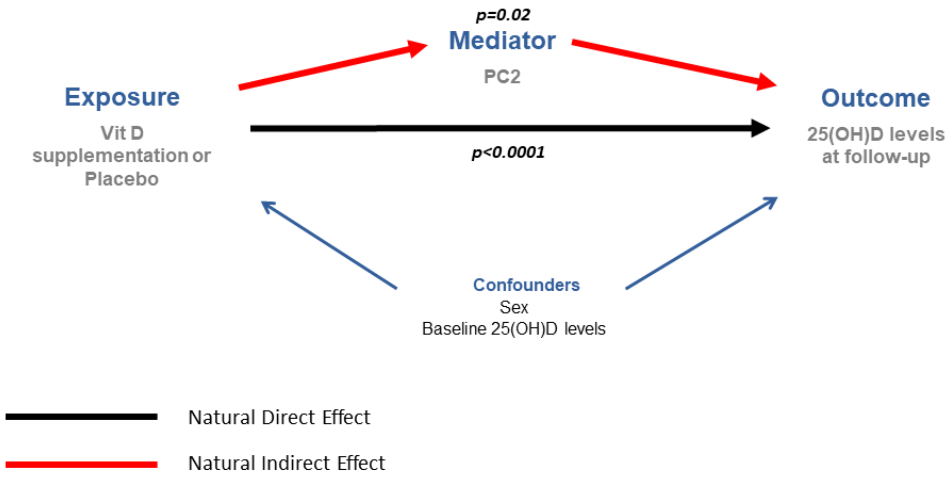


Figure 4.15. Direct acyclic graph (DAG**)** of mediation analyses. The 63 selected taxa (summarized with PC2) as mediator of the effect of vitamin D supplementation (exposure) on post-treatment 25(OH)D levels (outcome). In black, natural direct effect (NDE); in red, natural indirect effect (NDE); in blue, the effect of confounders on the exposure–outcome relationship. *p*-value obtained from mediation analysis. Significant direct effect of vitamin D supplementation on post-treatment 25(OH)D (p<0.0001). The 63 taxa significantly mediate the effect of supplementation on post-treatment 25(OH)D (p=0.02). 25(OH)D = 25-hydroxy vitamin D.

4.5.3.3.2 Functional pathways and vitamin D

We considered community-level pathway abundances for microbial function. At both timepoints, the abundances of 1465 pathways were computed for each patient and normalized using the counts per million (CPM) technique. Normalized abundances were clr-transformed after zero-imputing.

Only the pathways present in at least 10% of the analyzed patients (n=60) at the end of the treatment were considered (n=237 pathways). Using *coda-lasso*, we initially selected 40 pathways with post-treatment abundances significantly associated with the treatment arm. Of these, 8 were already significantly unbalanced at baseline. Consequently, 32 pathways were selected for investigation. One male patient in placebo was excluded from the statistical analysis because he had zero abundances for all the selected pathways.

Post-treatment abundances of the 32 pathways were investigated in relation to both vitD supplementation and vitD sufficiency. In multivariate analysis, we observed significantly increased *D-fructuronate degradation*, *superpathway of glycerol degradation to 1,3-propanediol*, *acetyl-CoA fermentation to butanoate II*, *superpathway of thiamin diphosphate biosynthesis II*, *guanosine nucleotides degradation II* in patients that were vitD supplemented, with *superpathway of glycerol degradation to 1,3-propanediol*, *superpathway of thiamin diphosphate biosynthesis II* and *guanosine nucleotides degradation II* significantly more abundant also in those with post-treatment 25(OH)D levels≥30 ng/ml. Conversely, *L-histidine biosynthesis* and *pyrimidine deoxyribonucleosides salvage* pathways were significantly more abundant in placebo patients, while the *pathway of L-ornithine de novo biosynthesis* was more abundant in those with vitD deficiency at follow-up (Figure 4.16a-**b**).

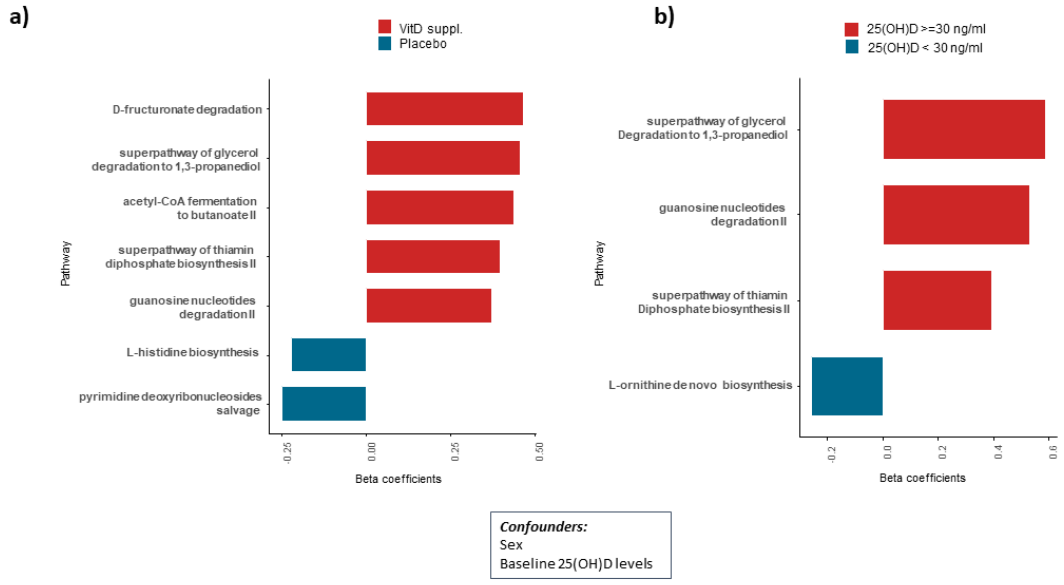


Figure 4.16. Pathways significantly associated with **a.** treatment arm **b.** post-treatment vitamin D sufficiency (25(OH)D ng/ml). For each pathway, results are obtained from a multivariable logistic model including the post-treatment *clr*-transformed abundance of the pathway as covariate and adjusted for confounders. The bar length indicates the significant beta-coefficient of the pathway ($p<0.05$). If positive, the pathway was significantly more abundant in patients **a.** supplemented with vitamin D **b.** reaching vitamin D sufficiency at the end of the treatment. If negative, the pathway was significantly more abundant in patients **a.** in the placebo group **b.** not reaching vitamin D sufficiency at the end of the treatment. 25(OH)D = 25-hydroxy vitamin D.

4.5.3.3.3 Vitamin D, microbiome and sex/gender

Looking at the distribution of 25(OH)D levels at both timepoints, we found that women in the supplementation group had lower vitD levels at baseline than men ($p_{baseline}$=0.04). However, the supplementation restored this gap, and by the end of the study both post-treatment levels and the change in 25(OH)D levels from baseline were comparable between men and women (p_{post}=0.70; p_{change}=0.95) (Figure 4.17d-f).

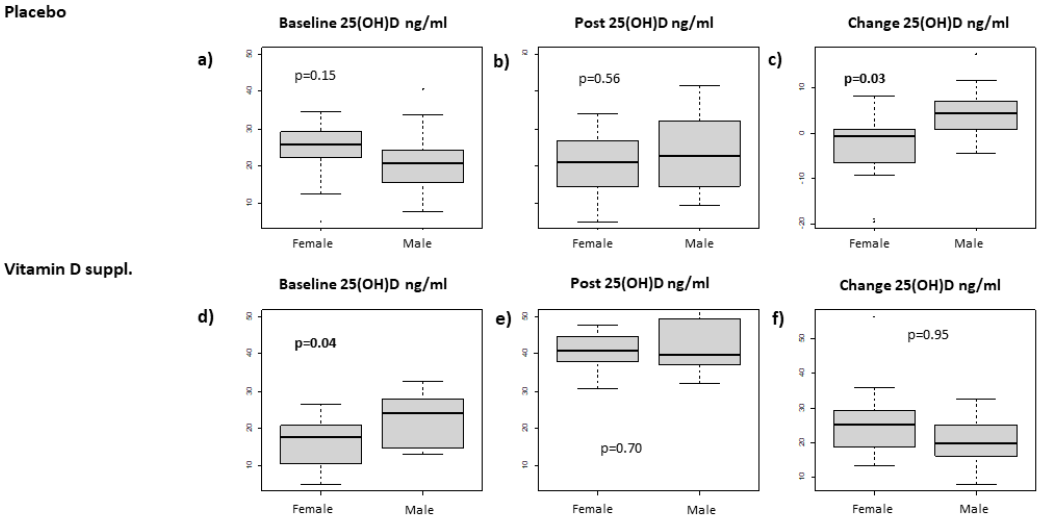


Figure 4.17. Boxplots of 25(OH)D levels and change of 25(OH)D from baseline by sex/gender at each timepoint and according to treatment arm. For the placebo group, the following 25(OH)D levels distributions are plotted: **a**. baseline; **b**. post-treatment; **c**. change from baseline. For the vitamin D supplementation group, the following 25(OH)D levels distributions are plotted: **d**. baseline; **e**. post-treatment; **f**. change from baseline.
In **a.** and **d.** p-values refer to the effect of the sex/gender covariate on baseline 25(OH)D levels and were derived from multivariable linear models adjusted for age, baseline Body Mass Index (BMI) and previous chemotherapy.
In **b.** and **e.** p-values refer to the effect of the sex/gender covariate on post-treatment 25(OH)D levels and were derived from multivariable linear models adjusted for age, baseline BMI and previous chemotherapy.
In **c.** and **f.** p-values refer to the effect of the sex/gender covariate on change of 25(OH)D levels from baseline and were derived from multivariable linear models adjusted for age, baseline BMI, previous chemotherapy and baseline 25(OH)D.

Conversely, no significant difference in 25(OH)D levels at baseline was found between women and men in the placebo group ($p_{baseline}$=0.15), although vitD levels increased significantly more in men throughout the year of treatment (p_{change}=0.03) (Figure 4.17a-c).

We also looked at a potential interaction between sex/gender and the treatment-associated taxa on post 25(OH)D levels, using PC2 as a proxy for taxa abundances. The predicted regression lines stratified by sex/gender are shown in Figure 4.18. We found a statistically significant interaction between sex/gender and PC2 on 25(OH)D levels ($p < 0.001$), suggesting that men and women had a different taxa composition at follow-up and that this difference also affected the final 25(OH)D levels (Figure 4.18).

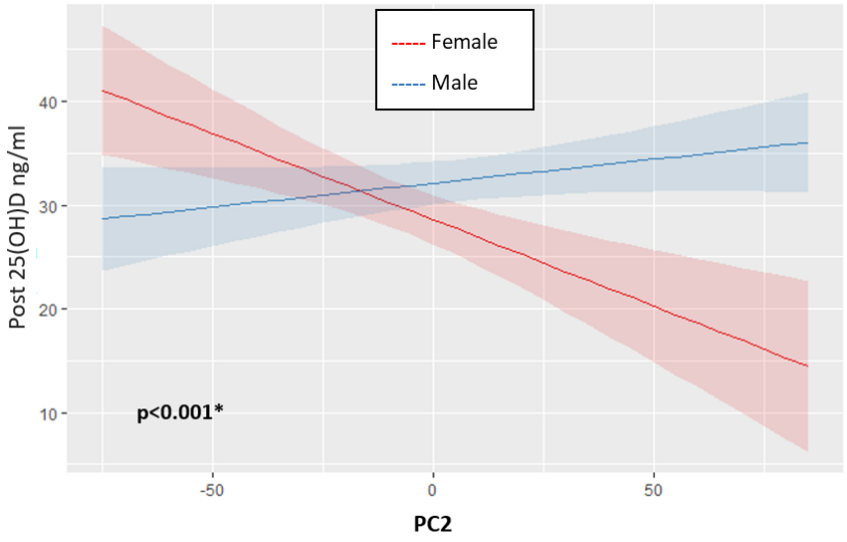


Figure 4.18. Regression lines obtained from a linear regression model run on post-treatment 25(OH)D levels, including the interaction between sex/gender and PC2 (second component from PCA run on the post-treatment *clr*-transformed abundances of the 63 selected taxa) and adjusted for confounders. In the plot, the p-value of the beta regression coefficient of the interaction between sex/gender and PC2 on post-treatment 25(OH)D levels is displayed.

As for taxonomic data, PCA was computed on the selected 32 *clr*-transformed pathways at follow-up. The first two components (Figure 4.19a-b), which explained about 30% of the total variance, did not differ between treatment groups. However, a difference between men and women could be detected in the first component (PC1), where the majority of men were distributed alongside the negative axis of PC1 (59% of men), whereas most of the women (74%) were in the positive side.

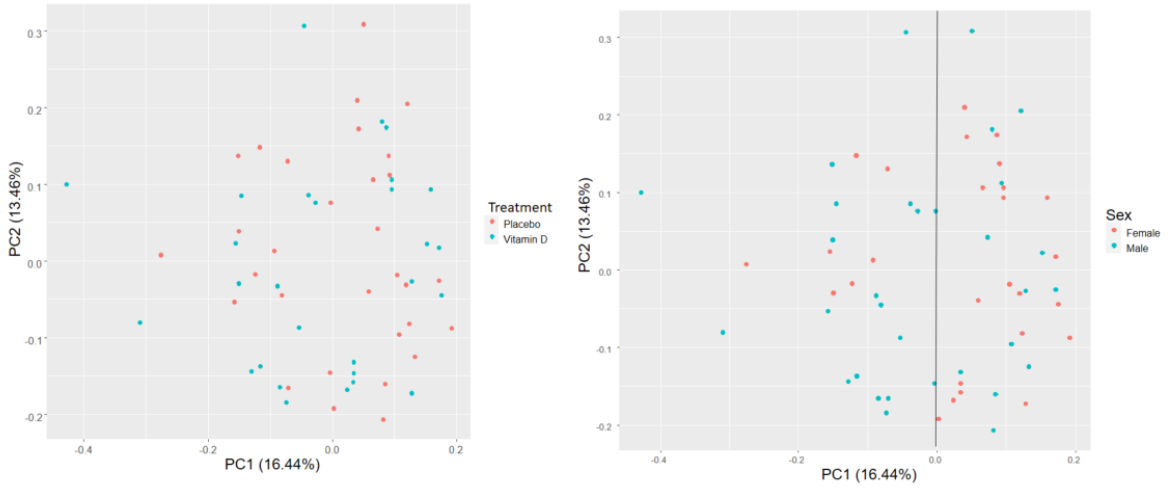


Figure 4.19. Plot of the scaled scores of the first two components from PCA run on the clr-transformed abundances at follow-up of the 32 selected pathways. Data points were colored according to **a.** treatment arm (red for the placebo group; blue for the vitamin D supplementation group) **b.** sex (red for female; blue for male).

This difference was further investigated in multivariate analysis, where an interaction between vitD supplementation and sex/gender was introduced. Results from the model showed that, while the abundances of the selected pathways summarized by PC1 was comparable between men and women in the placebo group, a significant difference was present between men and women after the supplementation ($p = 0.006$, Figure 4.20).

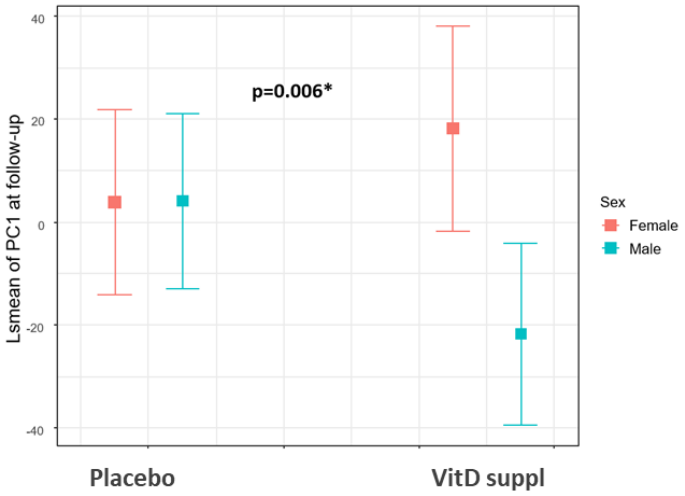


Figure 4.20. Least-square means obtained from the multivariable linear regression model run on PC1 (first component from PCA run on the post-treatment *clr*-transformed abundances of the 32 selected pathways), including the interaction between sex/gender and treatment arm and adjusted for confounders.
*p-value of interaction effect of treatment (vitamin D supplementation vs placebo) and sex on *PC1* at follow-up. The model was adjusted for: previous chemotherapy and baseline BMI.

The contribution of each pathway on PC1 was summarized in Figure 4.21 with their corresponding loading.

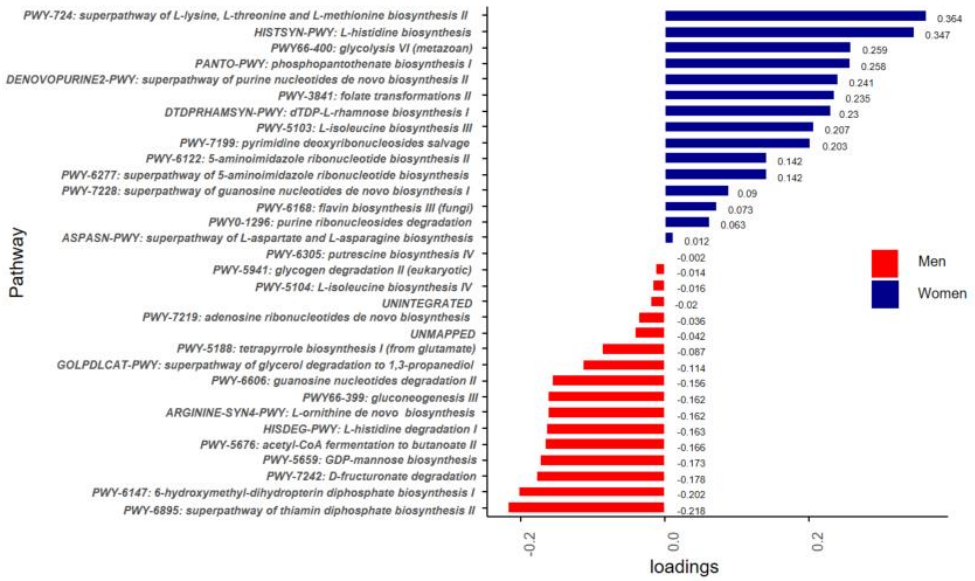


Figure 4.21. Bar plots of the loadings of the 32 pathways on PC1 (first component from PCA run on the post-treatment *clr*-transformed abundances of the 32 selected pathways). A positive loading indicates a positive correlation with the component, a negative loading indicates an inverse correlation with the component. Because most of the women had positive PC1 scores and most of the men had negative PC1 scores, a pathway with a positive loading is expected to be more abundant at follow-up in women, while a pathway with a negative loading is expected to be more abundant in men.

Because women had mostly positive PC1 scores, a pathway with a positive loading was expected to be more abundant in women. Conversely, a pathway with a negative loading was expected to be more prevalent in men. *Superpathway of L-lysine, L-threonine and L-methionine biosynthesis II and L-histidine biosynthesis* were the pathways with the two largest positive loadings. Both pathways involve the biosynthesis of essential amino acids. Multivariate regression analysis confirmed that *Superpathway of L-lysine, L-threonine and L-methionine biosynthesis II* was significantly more abundant in supplemented women compared to supplemented men (p = 0.002, Figure 4.22) while *L-histidine biosynthesis* was significantly less abundant in supplemented men than supplemented women (p = 0.002, Figure 4.23).

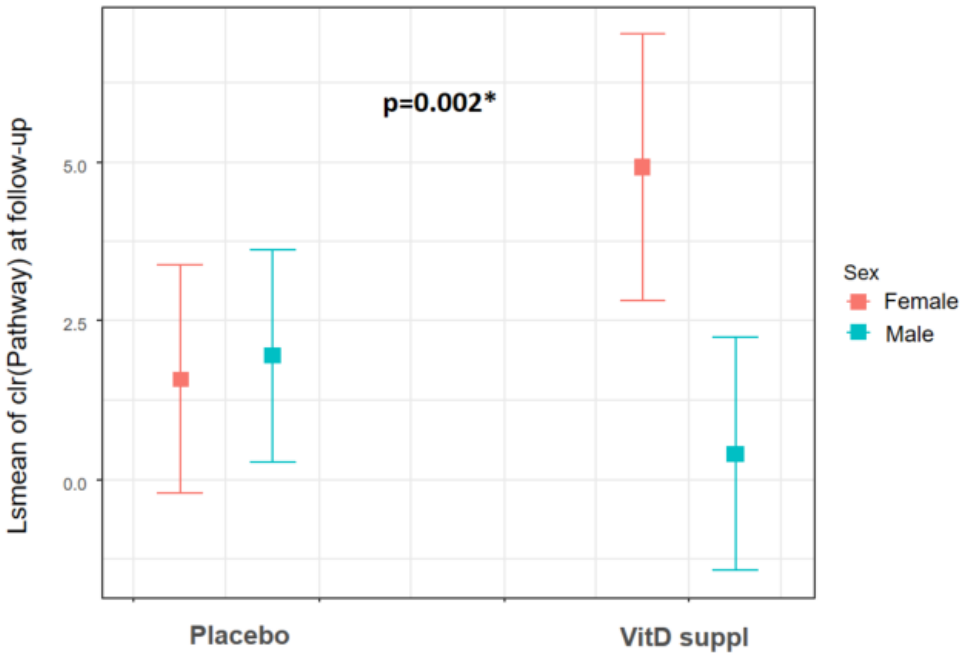


Figure 4.22. Least-square means obtained from the multivariable linear regression model run on *clr*-transformed abundances of *Superpathway of L-lysine, L-threonine and L-methionine biosynthesis II* at follow-up, including the interaction between sex/gender and treatment arm and adjusted for confounders.
*p-value of interaction effect of treatment (vitD supplementation vs Placebo) and sex/gender on *superpathway of L-lysine, L-threonine and L-methionine biosynthesis II* abundance (clr) at follow-up. The model was adjusted for: previous chemotherapy, baseline BMI and baseline 25(OH)D levels.

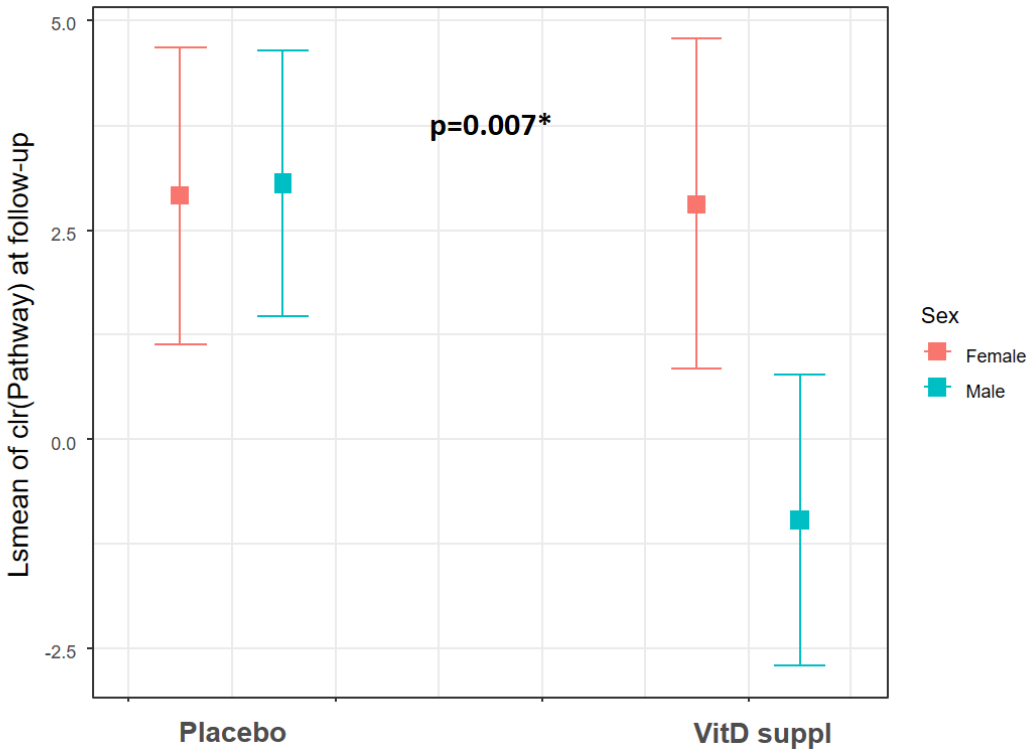


Figure 4.23. Least-square means obtained from the multivariable linear regression model run on *clr*-transformed abundances of *Pathway of L-histidine biosynthesis* at follow-up, including the interaction between sex/gender and treatment arm and adjusted for confounders.
*p-value of interaction effect of treatment (vitamin D supplementation vs placebo) and sex/gender on *L-histidine biosynthesis pathway* abundance (*clr*) at follow-up. The model was adjusted for baseline BMI and baseline 25(OH)D levels.

However, both pathways looked comparable among non-supplemented men and women.

Looking at the opposite side of the loadings barplot, *superpathway of thiamin diphosphate biosynthesis II* and *6-hydroxymethyl-dihydropterin diphosphate biosynthesis I* were the pathways with the largest negative loadings, so with the highest inverse contribution on PC1. In multivariable analysis, *6-hydroxymethyl-dihydropterin diphosphate biosynthesis I* was borderline significantly associated to treatment (p = 0.051), with an indication of decreasing levels in the supplementation group, but no significant association with sex/gender (p = 0.14) or interaction between vitD supplementation and sex/gender was observed (p = 0.09). *Superpathway of thiamin diphosphate biosynthesis II*, on the other hand, was also not significantly different by sex/gender, although, overall, it was significantly more abundant in supplemented patients (p = 0.001, **Figure 4.24**).

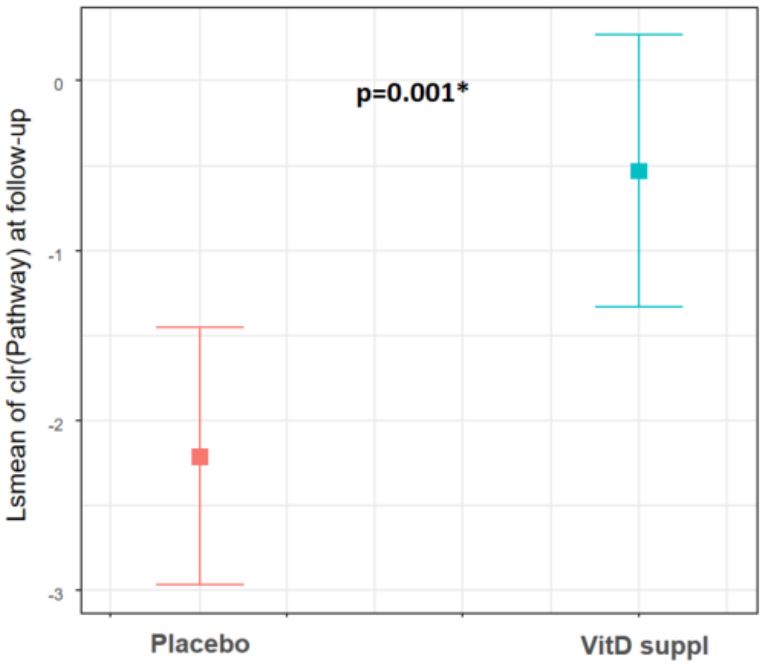


Figure 4.24. Least-square means obtained from the multivariable linear regression model run on *clr*-transformed abundances of *superpathway of thiamin diphosphate biosynthesis II* at follow-up, including the interaction between sex/gender and treatment arm and adjusted for confounders. *p-value of the treatment effect (vitamin D supplementation vs placebo) on *superpathway of thiamin diphosphate biosynthesis II* abundance (clr) at follow-up. The model was adjusted for sex/gender and baseline 25(OH)D levels.

4.5.3.4 Analysis of the change in microbiome

We used *coda-lasso* to select the taxa whose change was significantly different between the two treatment arms. The model was implemented after applying perturbation difference between the post-treatment and baseline taxa abundancies and closure.

We identified 21 taxa that differently changed during the study period according to the treatment: genus *Enterococcus*, genus *Sutterella*, *Bifidobacterium biavatii*, *Cellulomonas flavigena*, *Cryptobacterium curtum*, *Enterococcus durans*, *Enterorhabdus caecimuris*, *Eubacterium infirmum*, *Eubacterium pyruvativorans*, *s__Lachnospiraceae_bacterium_3-1*, *Lactobacillus parabuchneri*, *Lactobacillus sanfranciscensis*, *Prevotella buccalis*, *Pyramidobacter piscolens*, *Raoultella ornithinolytica*, *Streptococcus intermedius*, *s__Streptococcus_sp._263_SSPC*, *s__Streptococcus_sp._C150*, *s__Streptococcus_sp._HMSC062D07*, *s__Streptococcus_sp._I-P16*, s__*Victivallales_bacterium_CCUG_44730*.

Of these, genus *Sutterella*, *Enterorhabdus caecimuris*, *Cellulomonas flavigena*, *Prevotella buccalis*, *Eubacterium infirmum*, *Cryptobacterium curtum*, *s__Streptococcus_sp._263_SSPC*, *s__Streptococcus_sp._I-P16* significantly increased in patients supplemented with vitD.

To summarize the overall change of the gut microbiome accounting for the compositional structure of data, we built a score based on PCA applied on the *clr*-transformed perturbation of the abundances at the two time points. Scaling and centering were applied, and the final score of each patient was calculated as the linear combination of the score of each component multiplied by the square root of its eigenvalue:

$$Score_{microbiome_{change}}(j) = \sqrt{\lambda_1} * PC_{1j} + \sqrt{\lambda_2} * PC_{2j} + \cdots + \sqrt{\lambda_{60}} * PC_{60j}$$

where PC_{ij} is the score of the principle component *i* for the patient *j*, $\lambda_i$is the eigenvalue of the principal component *i*, with *i, j*=1,…,60.

Once again, we observed a significant interaction between vitD and weight status, with change in 25(OH)D levels being correlated with the score in normal-weight individuals (p=0.04), but not in overweight (p=0.31) (Figure 4.25).

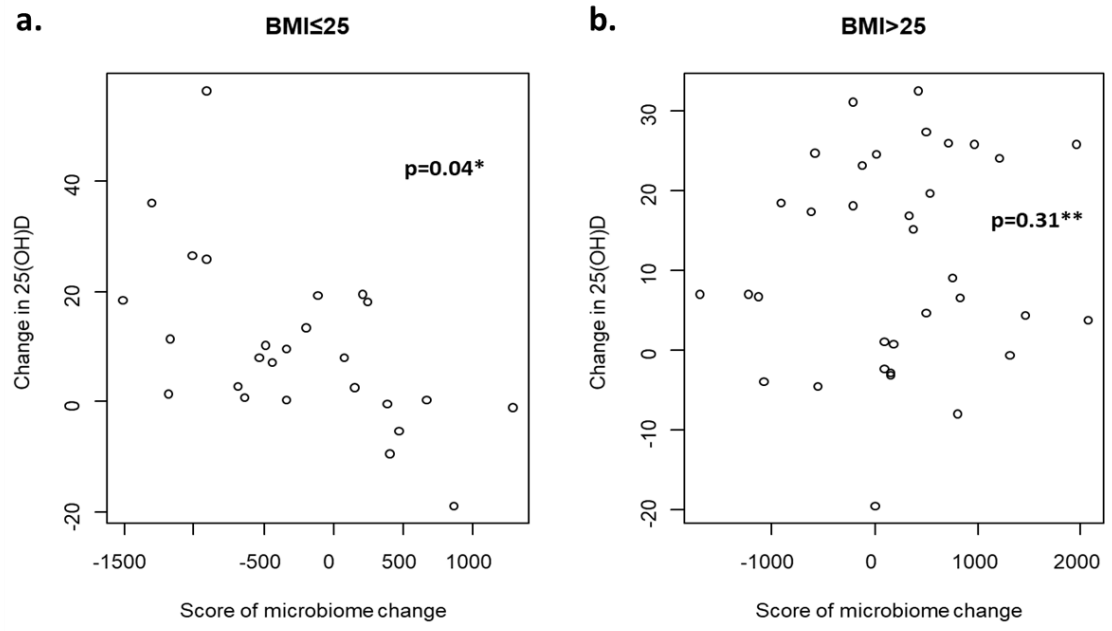


Figure 4.25. Scatterplots of change in microbiome score and change in 25(OH)D levels by overweight status. **a**. BMI≤25. **b**. BMI>25.
*p-valued derived from a multivariable linear regression model on the score of microbiome change, adjusted for age, sex/gender, season of blood draw and treatment arm, including individuals with BMI≤25.
**p-value derived from a multivariable linear regression model on the score of microbiome change, adjusted for age, sex/gender, season of blood draw and treatment arm, including individuals with BM>25.

4.5.3.4.1 Integrative Data Analysis of gut microbiome, circulating markers, diet and lifestyle, and weight status

To better comprehend the intricate interplay between the diversity of the gut microbiome, diet, lifestyle, and circulating markers, we estimated networks based on partial Spearman correlations to integrate all this information. To assure the robustness of the displayed relationships, L1 regularisation was applied.

We estimated two distinct networks, one for normal-weight individuals (BMI≤25) and one for overweight individuals (BMI>25), as the evidence shown so far seemed to suggest a different modulation of the microbiome and circulating markers based on the weight status of patients.

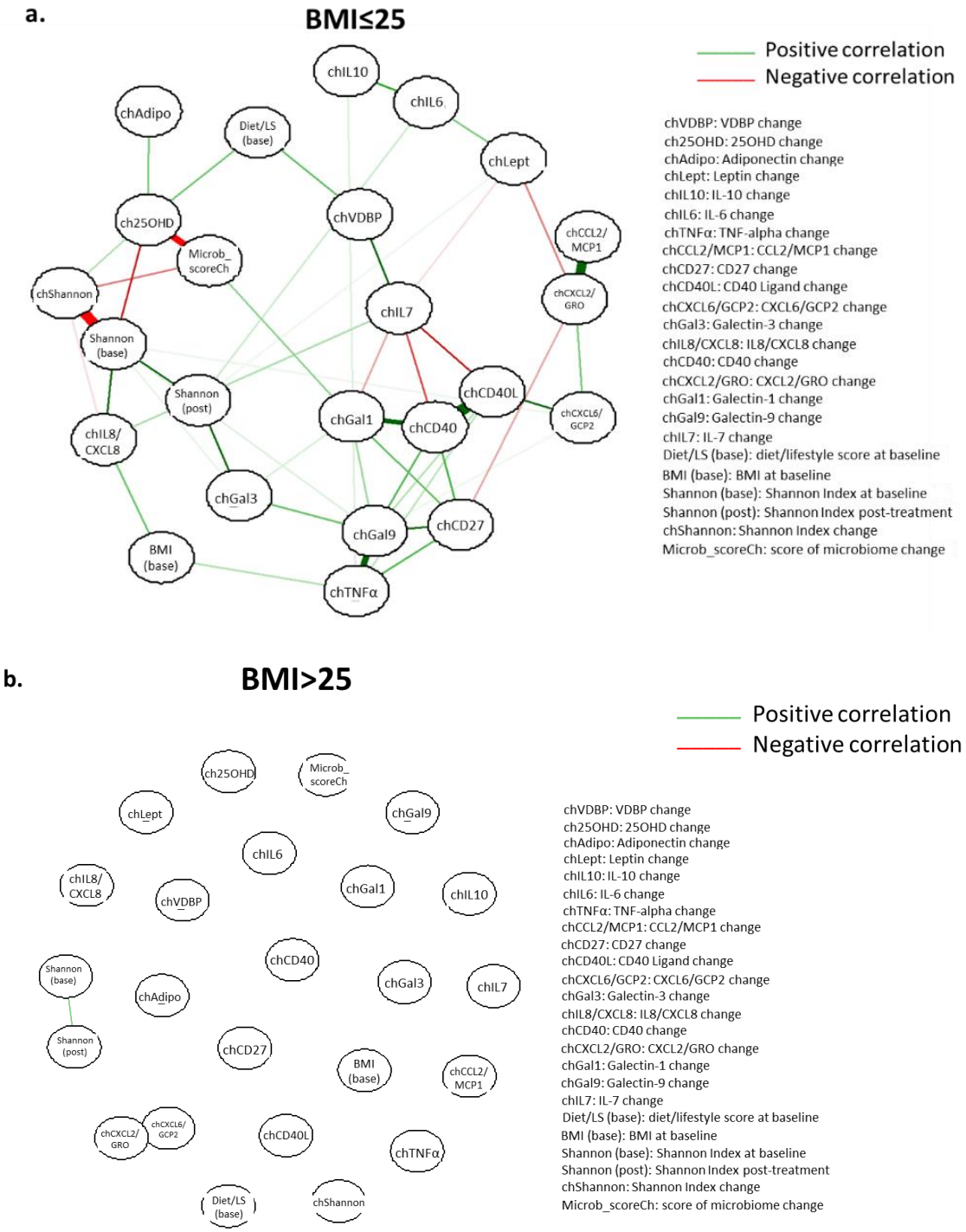


Figure 4.26. Network analysis based on Spearman partial correlations, stratified by overweight status. **a**. BMI≤25. **b**. BMI>25. The significant edges (indicating partial correlations) were retained using graphical LASSO. The line color indicates the direction of a correlation (green for positive and red for negative).

Network analysis confirmed a complex and intricate interconnection among the factors in normal-weight individuals.

The change in 25(OH)D levels was directly correlated with the microbiome change score, as well as with the baseline and change in alpha diversity. In particular, alpha diversity increased with increasing 25(OH)D levels, and 25(OH)D levels increased more in individuals with low alpha diversity at baseline. Interestingly, change in 25(OH)D was also positively and directly correlated with increasing levels of adiponectin and baseline adherence to WCRF recommendations, which, at the same time, appeared to mediate the relationship between change in 25(OH)D levels and change in VDBP (Figure 4.26).

In overweight individuals, no significant relationship between the investigated factors was identified, with the exception of the correlation between alpha diversity at baseline and at the end of the treatment (Figure 4.26).

The subsequent phase of data integration involved employing block sPLS-DA. Block sPLS-DA is a multivariate analysis method used to integrate and analyze data from multiple sources or 'blocks', to select relevant variables from each block that contribute to the discrimination of groups.

In our case, we used block sPLS-DA to identify the taxa that were significantly altered post vitD supplementation, while also accounting for the change in circulating biomarkers, dietary habits and lifestyle, and BMI. To do that, we created two blocks: the “microbiome” block, including the change in microbiome (obtained as the *clr*-transformation of the perturbation difference between the post-treatment and baseline relative abundances), and the “clinical” block, including the change in circulating biomarkers, and the baseline diet/lifestyle score and BMI.

The design matrix for the model was chosen based on the observed correlation structure between the blocks, determined by executing PLS on each block pair. We considered the first two PLS components provided by the model and identified the optimal number of variables in each block for each component using 5x50 cross-validation.

We selected ten taxa for each component of the "microbiome" block, and two and fourteen variables for the first and second component of the biomarker/lifestyle block, respectively.

Overall, both blocks of data effectively discriminated between the vitD supplementation group and the placebo group. In particular, the first component showed a high capacity for discrimination between the two groups, with positive scores mainly identifying the placebo arm and the negative scores primarily identifying the vitD supplementation group (Figure 4.27).

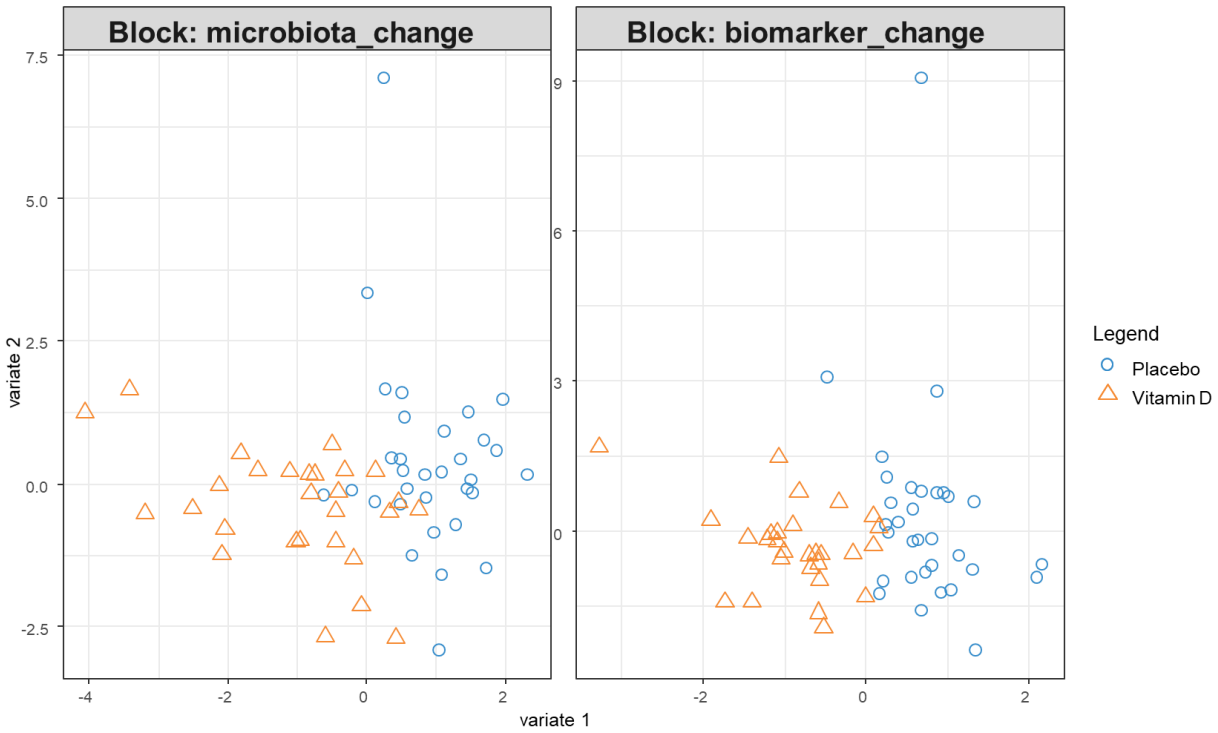


Figure 4.27. Scatter plot of the scores of the first two components estimated for each "block" of data (the "microbiome" block including the *clr*-transformed change in taxa abundances and the "biomarker/lifestyle" block including the change in circulating biomarkers, baseline diet/lifestyle score and baseline BMI) through block sparse Partial Least Square-Discriminant Analysis (block sPLS-DA), including the treatment arm as outcome. A step of variable selection was first employed on the components of each block through L1 (LASSO) penalization. The optimal number of variables to select from each component of each block was estimated using 5x50 CV.

As expected, the two biomarkers that characterized the first component of the biomarkers and lifestyle block were 25(OH)D and VDBP. Their variations during the treatment period effectively differentiated between the supplementation and placebo groups. Conversely, for the microbiome block, the first component was defined by changes in ten taxa, namely: *Dehalococcoides mccartyi*, genus *Peptoniphilus*, *Hallella seregens*, genus *Sutterella*, *s__Lachnospiraceae_bacterium_3-1*, *s__Streptococcus_sp._I-P16*, *Cryptobacterium curtum*, *Porphyromonas uenonis*, *s__Streptococcus_sp._263_SSPC*, *Mitsuokella multacida*.

Of these, *Dehalococcoides mccartyi*, *Hallella seregens*, genus *Sutterella*, *s__Streptococcus_sp._I-P16*, *Cryptobacterium curtum*, *s__Streptococcus_sp._263_SSPC*, *Mitsuokella multacida* increased in the supplementation group (**Figure 4.28**). Genus *Sutterella*, *Cryptobacterium curtum*, *s__Lachnospiraceae_bacterium_3-1*, *s__Streptococcus_sp._263_SSPC*, *s__Streptococcus_sp._I-P16* were also selected by the coda-lasso model.

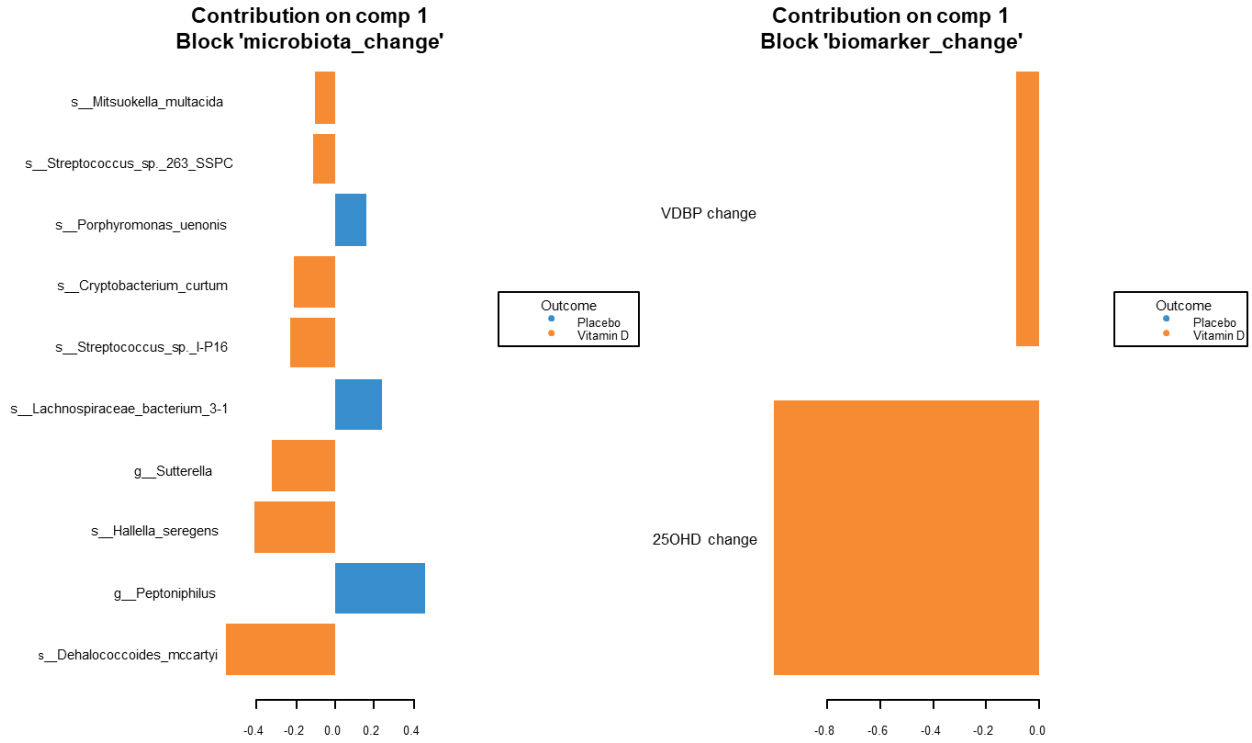


Figure 4.28. Barplots represent the loadings of the variables selected from the first component of each block in the block sPLS-DA model, with the treatment arm included as the outcome. In the "microbiome" block, 10 taxa were retained. In the "biomarker/lifestyle" block, change in 25(OH)D and VDBP were selected. The loading for each selected variable is allocated to the arm where that variable exhibits a higher median value.

The hierarchical clustering-based clustered image map in Figure 4.29, which visualizes the scaled values of the variables selected from the first component of both data blocks, shows that the change in the 10 chosen taxa, as well as in the levels of 25(OH)D and VDBP, clearly distinguish between the two treatment arms.

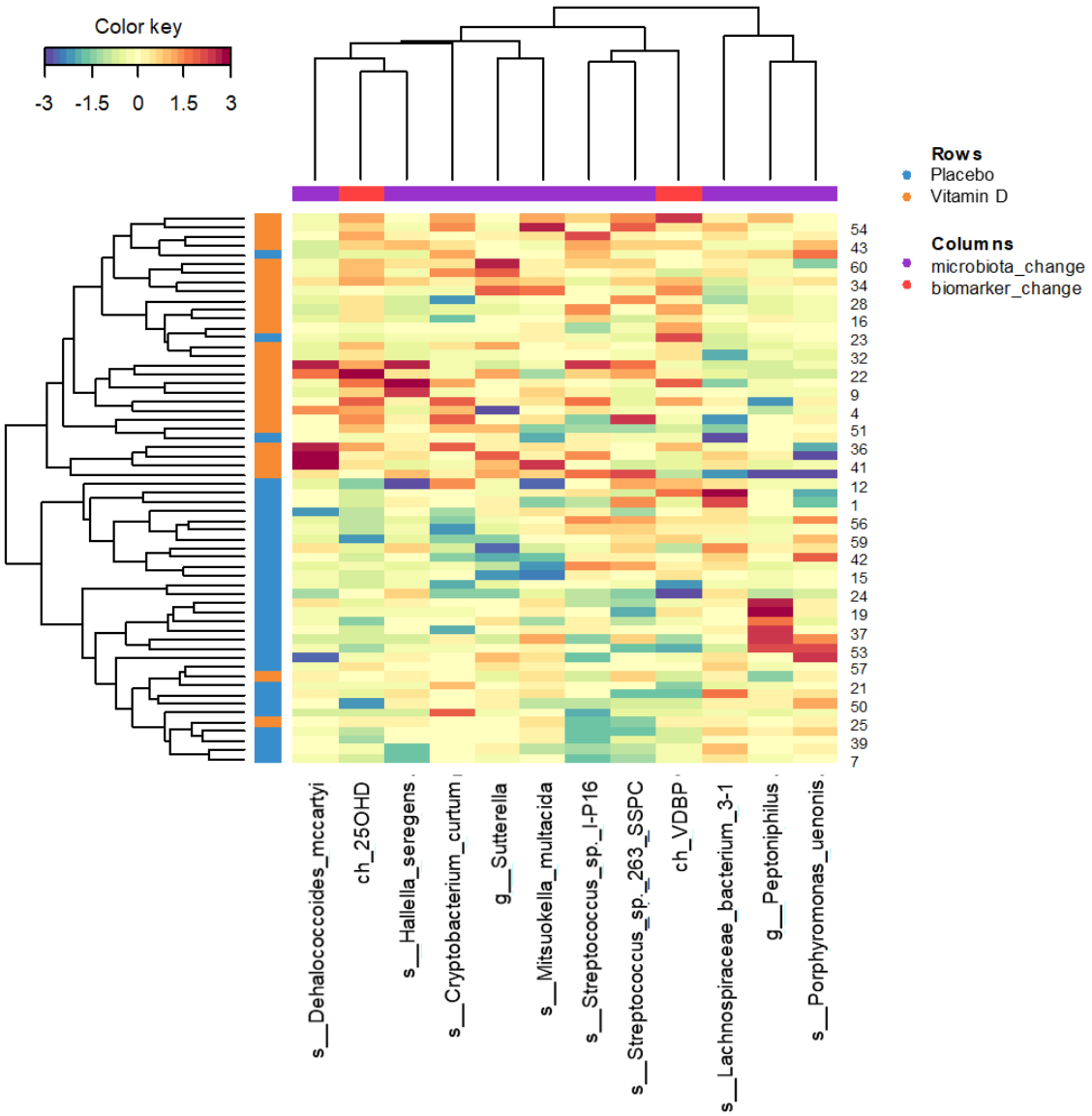


Figure 4.29. Clustered Image Map for the variables selected on component 1 by block sPLS-DA including the treatment arm as outcome. Hierarchical clustering of the scaled values of the selected variables was employed based on Euclidian distance and the complete linkage method. The color scale ranges from blue indicating low values to red indicating high values. Patients were displayed in rows while the selected variables were displayed in columns. The two data blocks in the model were: the "microbiome" block, which included the perturbation difference between the taxa abundance in the two time following *clr*-transformation, and the "biomarker/lifestyle" block, including the changes in circulating biomarkers, baseline diet/lifestyle score and baseline BMI. ch_VDBP = change in VDBP levels. ch_25(OH)D = change in 25(OH)D levels.

4.5.4 Analysis of Gene Expression

We performed a clustering analysis on the expression profiles of the immune-related 395 genes from the OIRRA panel, evaluated in the tumour tissue. High-quality reads were generated for 46 patients, meeting the quality standards of the laboratory. For two patients, gene expression (GE) profiling was performed on two distinct samples due to the heterogeneous characteristics of the tumor in each sample.

After filtering (retaining only genes expressed in a minimum of 10% of the samples, leading to the selection of 371 genes) and zero imputation, we *clr*-transformed the normalized GE data to address its compositional nature.

We used consensus clustering on the *clr*-transformed data to group patients into distinct clusters, aiming to define a sort of "transcriptomic signature". We employed every possible combination of clustering algorithms (*hierarchical*, *paritioning around medoids*, *k-means*) and distance metrics (*pearson [1 - Pearson correlation]*, *spearman [1 - Spearman correlation]*, *Euclidean*, *maximum*, *Canberra* and *minkowski*) available in the *ConsensusClusterPlus* package in R, to identify the optimal clustering partition of patients based on their GE.

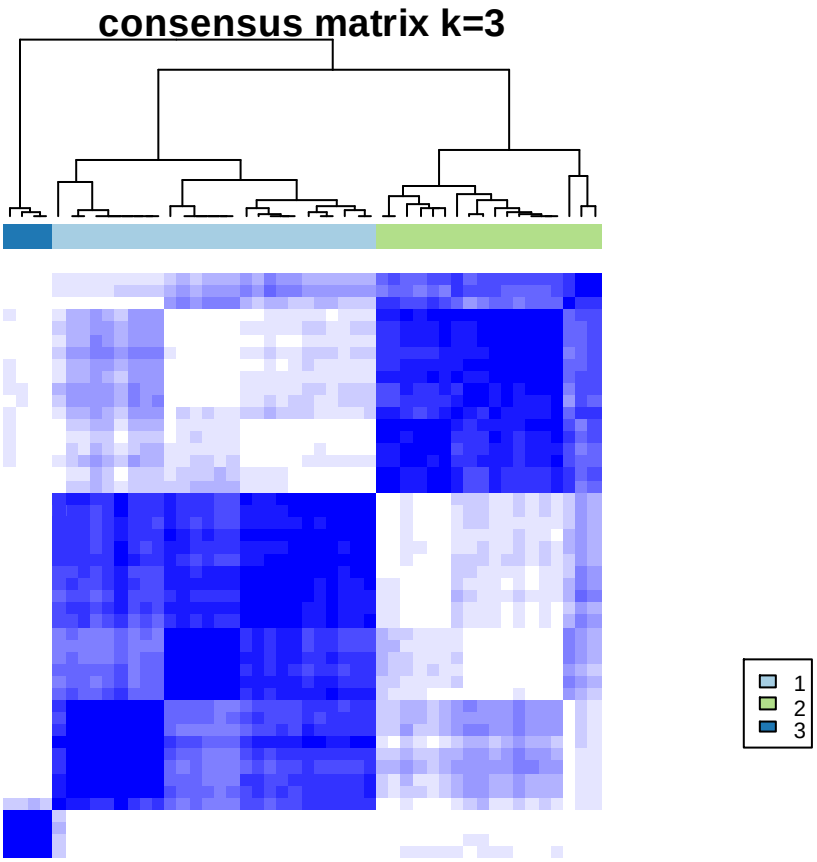


Figure 4.30. Hierarchical clustering of the consensus matrix which included the 46 samples of GE after *clr*-transformation and considering k=3 clusters. The clustering was carried using the partitioning around medoids algorithm based on the Spearman distance metric (1-Spearman correlation). High intensities of blue indicate high consensus.

Eventually, three robust GE clusters of patients were identified (Figure 4.30). They were determined using the partitioning around medoids algorithm, based on the Spearman distance. These clusters exhibited strong agreement across subpartitions and remained consistent even in the sensitivity analysis including the normalized GE data, without the *clr*-transformation. Moreover, the same clusters were obtained when we included, one at a time, the GE profiles of the patients who had two separate GE analyses on distinct tumour tissue samples.

Cluster 1 included 26 patients, *Cluster 2* included 16 patients and *Cluster 3* included 4 patients. After correction for multiple testing, the *clr*-transformed expression of 50 genes and the normalized expression of 96 genes (without *clr*-transformation) was significantly different across the three clusters (p_{adj}<0.05). Of these 96 genes, 42 were included in the set of 50 genes identified in the main analysis based on the *clr* transformation (see Supplementary Table S12; Figure 4.31).

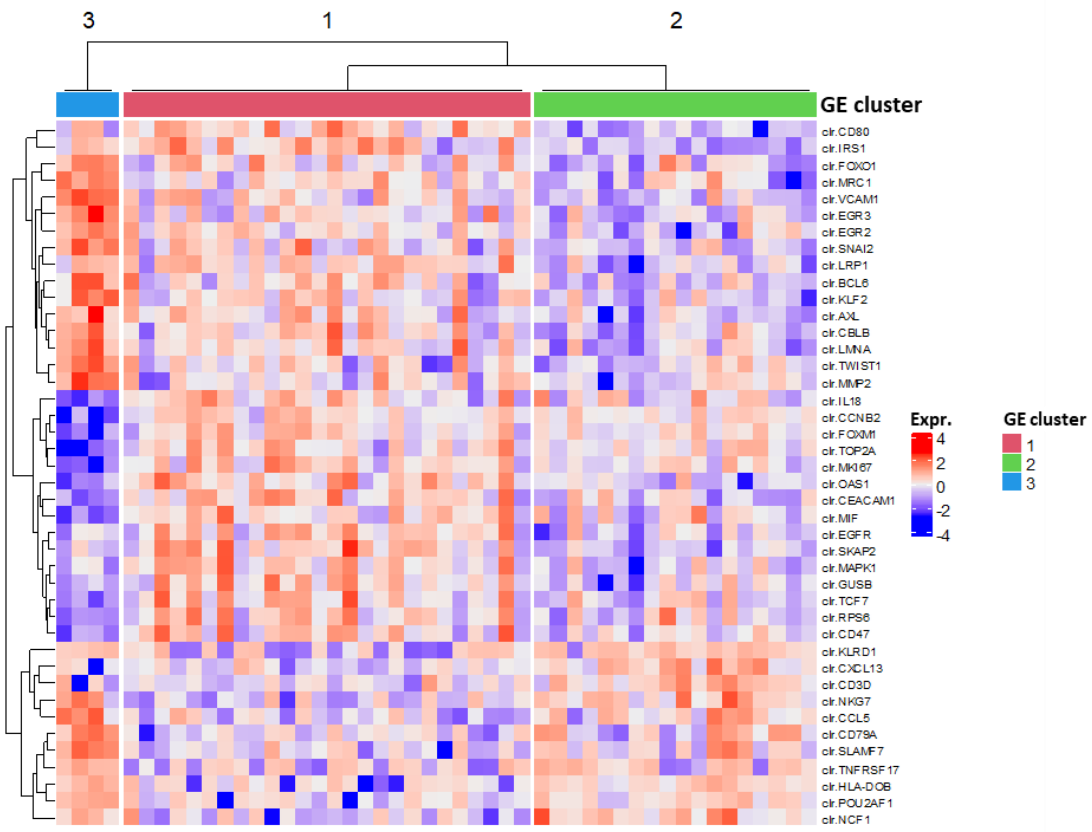


Figure 4.31. The heatmap includes the scaled *clr*-transformed expression of the 42 genes that were differently expressed between the three GE clusters. The color scale ranges from blue indicating low expression to red indicating high expression. Columns were grouped based on on the Gene Expression cluster. Genes in the heatmap were hierarchically clustered using the Spearman distance metric (1-Spearman correlation) and the "complete" linkage method. GE=Gene Expression. Expr.=scaled *clr*-transformed expression of genes.

Cluster 2 was associated with a higher risk of colorectal events compared to *Cluster 1* and *Cluster 3* (log-rank test for colorectal events: p=0.023; Figure 4.32), with 9 patients out of 16 experiencing at least one colorectal event during the follow-up period (Figure 4.33). It was characterized by high expression levels of 11 of the 42 reproducible selected genes (*KLRD1*, *CXCL13*, *CD3D*, *NKG7*, *CCL5*, *CD79A*, *SLAMF7*, *TNFRSF-17*, *HLA-DOB*, *POU2AF1*, *NXF1*) which conversely, were found to be underexpressed in *Cluster 1*. *Cluster 3* also exhibited increased expression of these genes, along with another set of genes that were also highly expressed in *Cluster 1*.

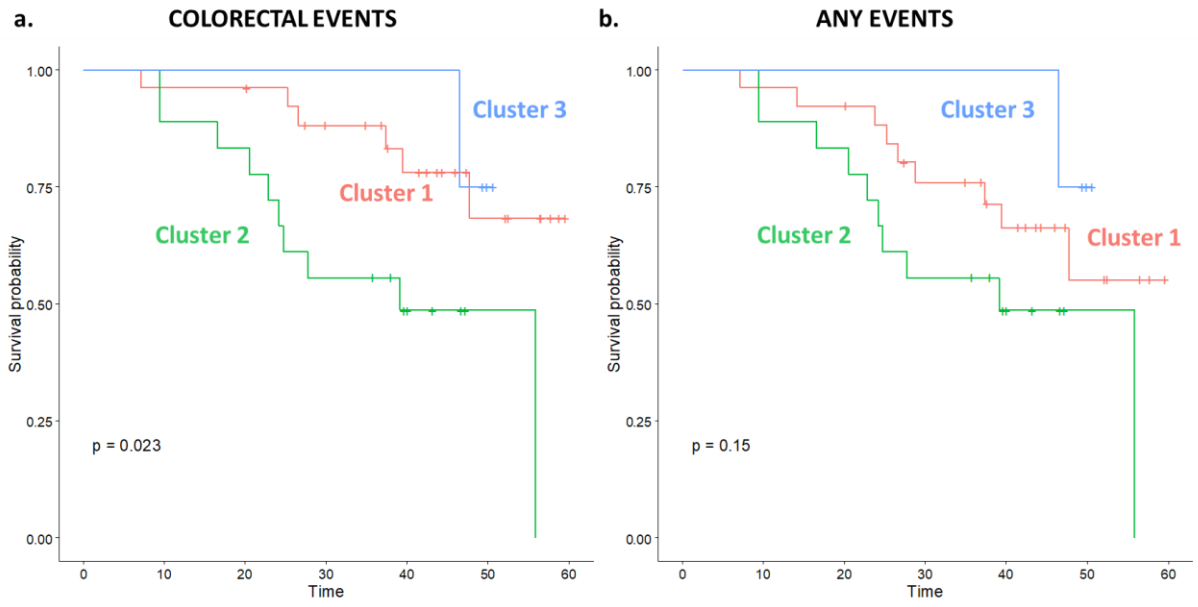


Figure 4.32. Kaplan-Meier curves for event-free survival (EFS) and Log-Rank test according to the three GE clusters. In **a**. are the Kaplan-Meier curves for colorectal events, which include tumour relapse, death but also colorectal adenomas and polyps. In **b**. are the Kaplan-Meier curves for any clinical events, which include tumour relapse, death, colorectal adenomas and polyps, and other tumours.

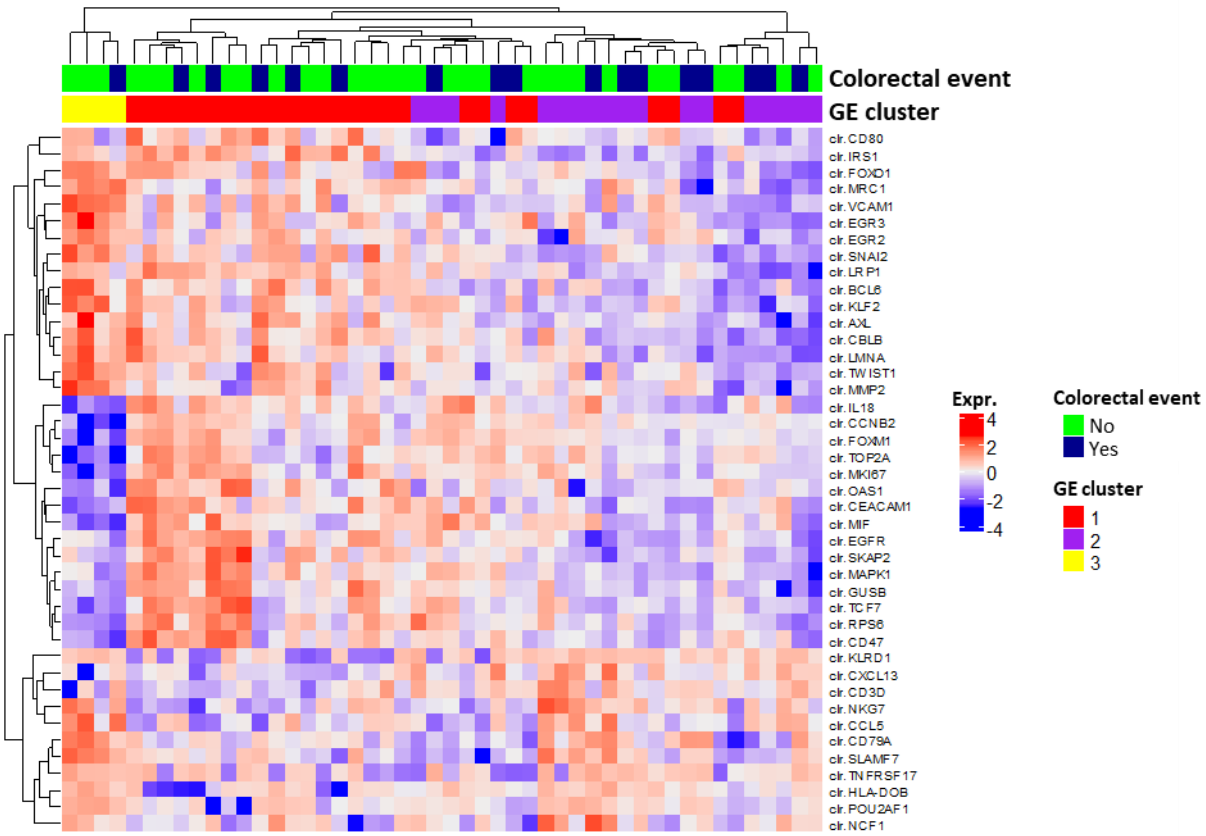


Figure 4.33. The heatmap includes the scaled *clr*-transformed expression of the 42 genes that were differently expressed between the three GE clusters. The color scale ranges from blue indicating low expression to red indicating high expression. Annotations for the GE clusters and occurrence of colorectal events were done. Genes in the heatmap were hierarchically clustered using the Spearman distance metric (1-Spearman correlation) and the "complete" linkage method. GE=Gene Expression. Expr.=scaled *clr*-transformed expression of genes.

Of the four individuals in *Cluster 3*, three were males and one was female. When compared to patients in *Clusters 1* and *Cluster 2*, all individuals in *Cluster 3* had undergone both radiotherapy and chemotherapy and, notably, exhibited higher levels of 25(OH)D both at baseline and at the end of the treatment period (Figure 4.34). Additionally, they had reduced levels of inflammation, higher baseline IL-7 and IL-8/CXCL8 levels, and a lower BMI (Figure 4.35).

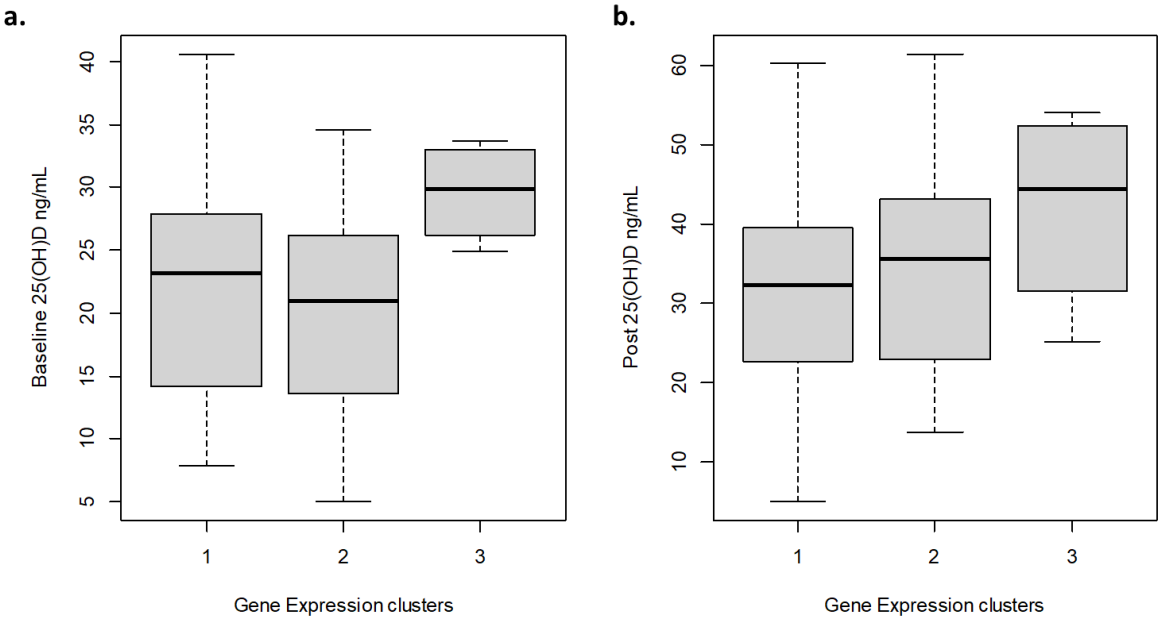


Figure 4.34 Boxplot of the distribution of 25(OH)D levels **a**. at baseline **b**. at the end of the treatment period by Gene Expression cluster.

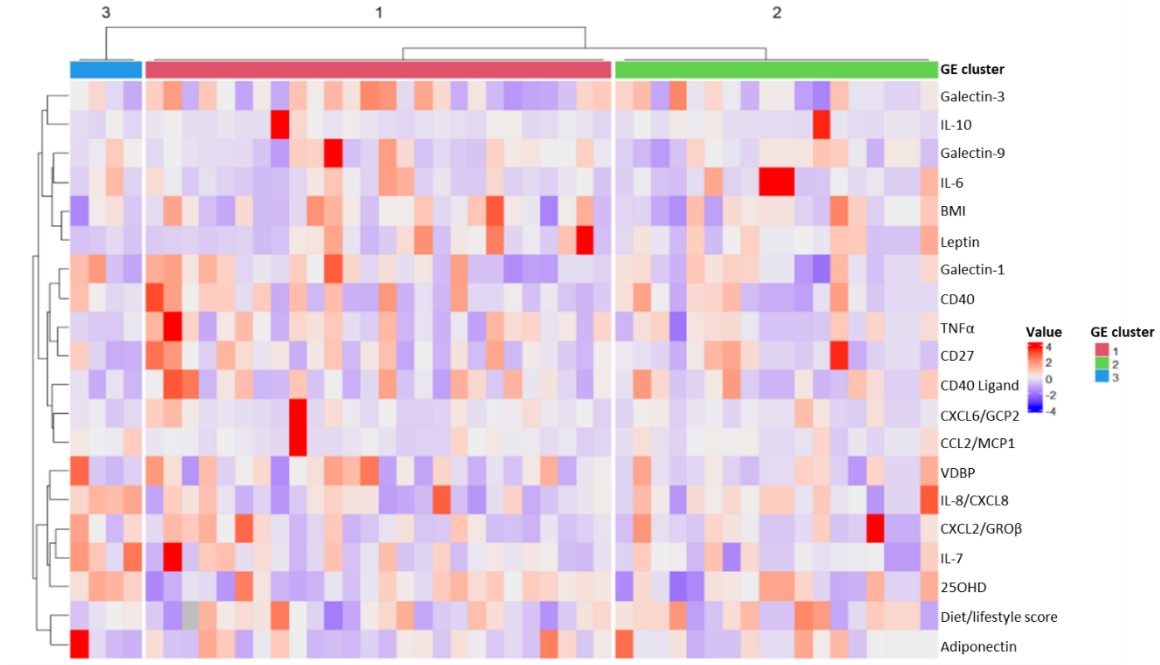


Figure 4.35. The heatmap includes the scaled values of the circulating biomarker, BMI and diet/lifestyle score at baseline. The color scale ranges from blue indicating low values to red indicating high values. Columns were grouped based on the Gene Expression cluster. The included variables in the heatmap were hierarchically clustered using the Spearman distance metric (1-Spearman correlation) and the "complete" linkage method. GE=Gene Expression. Value=scaled values of the included variables. BMI=Body Mass Index.

4.5.5 Analysis of Colorectal and Clinical events

4.5.5.1 Integrative Data Analysis of gut microbiome, gene expression, circulating markers, diet and lifestyle, and weight status

In total, 40 patients had the GE profiling and microbiome evaluation at both time points. We included them in an integrative analysis of GE data, microbiome, circulating biomarkers, diet/lifestyle score and BMI at the end of the treatment, in order to identify patterns that could differentiate between patients who experienced at least one colorectal event during the follow-up period and those who did not.

To achieve this, we employed block sPLS-DA including three distinct data blocks: "post-treatment microbiome," "post-treatment circulating markers, diet/lifestyle score, BMI," and "gene expression". Microbiome and GE data were *clr*-transformed.

We considered the first two PLS components of the model, estimating the optimal number of variables for each block via 5x50-fold cross-validation. The final model was obtained excluding one outlier patient, which accounted for most of the variability of the first component.

Overall, the integration of the tree blocks of data appeared to effectively distinguished patients who have experienced at least one colorectal event during follow-up from those who have not (Figure 4.36). Notably, the microbiome block, through the combination of the first two components, and the GE block, primarily through its first component, emerged as the most discriminative. In contrast, the circulating biomarkers, diet/lifestyle and BMI block provided the least differentiation between the two groups.

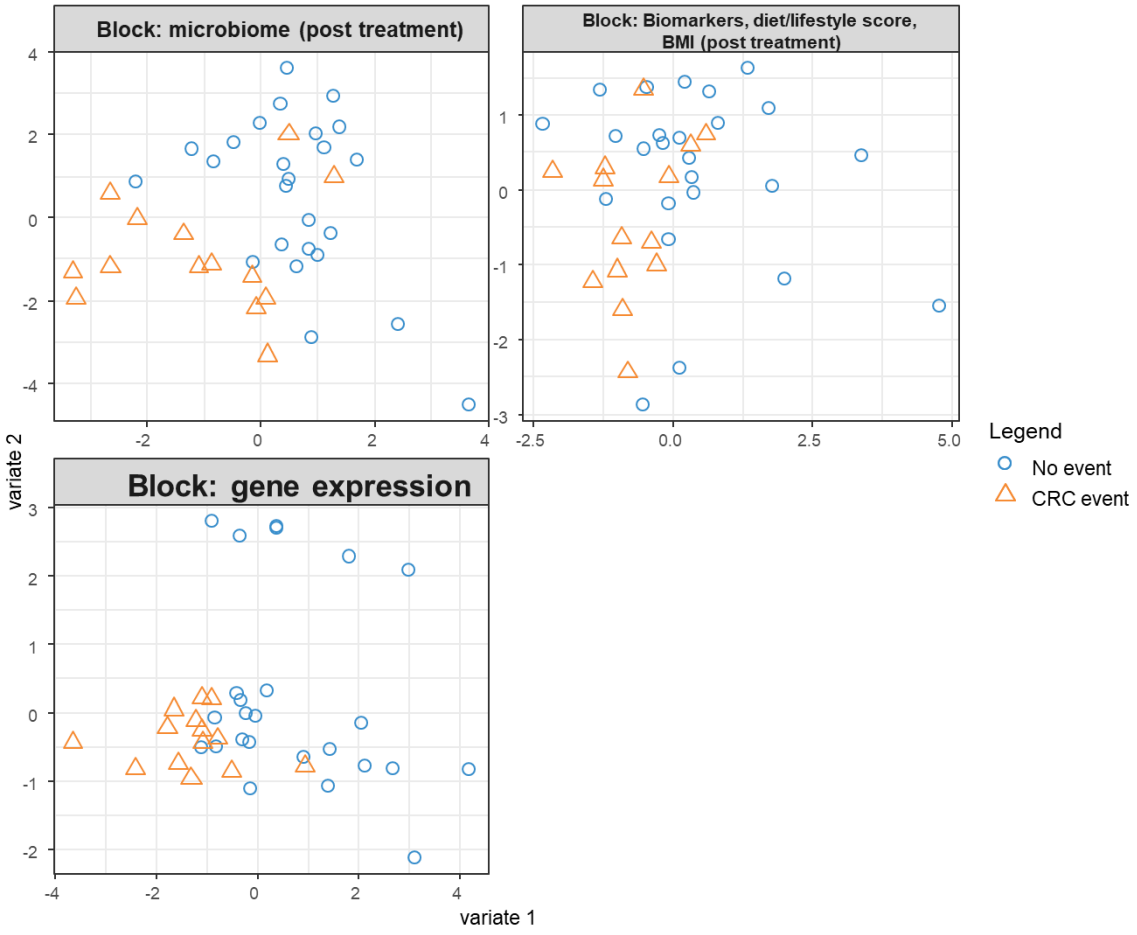


Figure 4.36. Scatter plot of the scores of the first two components estimated for each "block" of data (the "microbiome" block including the *clr*-transformed abundances of taxa at the end of the treatment, the "biomarker, diet/lifestyle score, BMI" including the post-treatment values of the circulating biomarkers, the diet/lifestyle score and BMI, and the "gene expression" block including the *clr*-transformed expression of the genes). The components were estimated through block sparse Partial Least Square-Discriminant Analysis (block sPLS-DA), including the occurrence of at least one colorectal event as outcome. A step of variable selection was first employed on the components of each block through L1 (LASSO) penalization. The optimal number of variables to select from each component of each block was estimated using 5x50 CV. One outlier patient was excluded from the analysis.

For the microbiome, we selected 10 taxa from each component. For GE, we selected 10 genes from the first component and 5 genes from the second component. From the circulating markers, diet/lifestyle and BMI block, we selected 4 variables from the first component and 2 from the second (Figure 4.37).

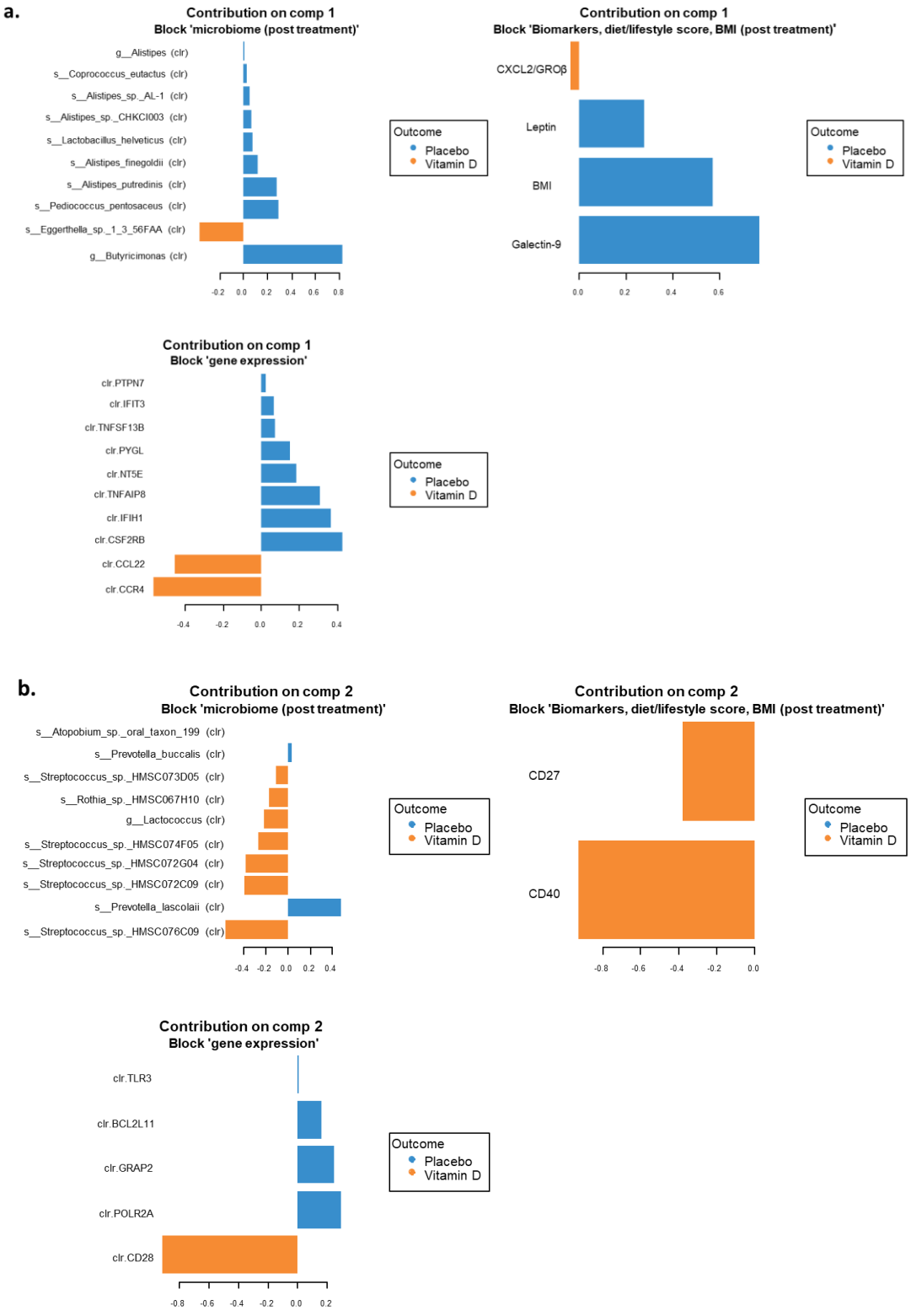


Figure 4.37. Barplots represent the loadings of the variables selected from **a**. the first component and **b**. the second component of each block in the block sPLS-DA model, with the occurrence of at least one colorectal event as outcome. The "microbiome" block includes the *clr*-transformed abundances of taxa at the end of the treatment; the "biomarker, diet/lifestyle score, BMI" includes the post-treatment values of the circulating biomarkers, the diet/lifestyle score and BMI; the "gene expression" block including the *clr*-transformed expression of the genes. The loading for each selected variable is allocated to the arm where that variable exhibits a higher median value.

The heatmap displayed in Figure 4.38 shows that the features selected from the first two components of each block define a cluster which identifies most of the patients who experienced at least one colorectal event. Specifically, this group of patients exhibited elevated levels of CD27 and CD40, along with higher abundances of genus *Lactococcus*, *s__Rothia_sp._HMSC067H10*, *s__Atopobium_sp._oral_taxon_199*, *s__Streptococcus_sp._HMSC073D05*, *s__Streptococcus_sp._HMSC074F05*, *s__Streptococcus_sp._HMSC076C09*, *s__Streptococcus_sp._HMSC072C09*, *s__Streptococcus_sp._HMSC072G04*, *s__Eggerthella_sp._1_3_56FAA* by the end of the treatment. Moreover, the genes *CD28*, *CCR4* and *CCL22* were upregulated in this subset of patients. Interestingly, *CCL22* was found to be significantly upregulated in *F. nucleatum*-infected CRC cell lines, suggesting a role in *F. nucleatum*-related colorectal tumorigenesis[274].

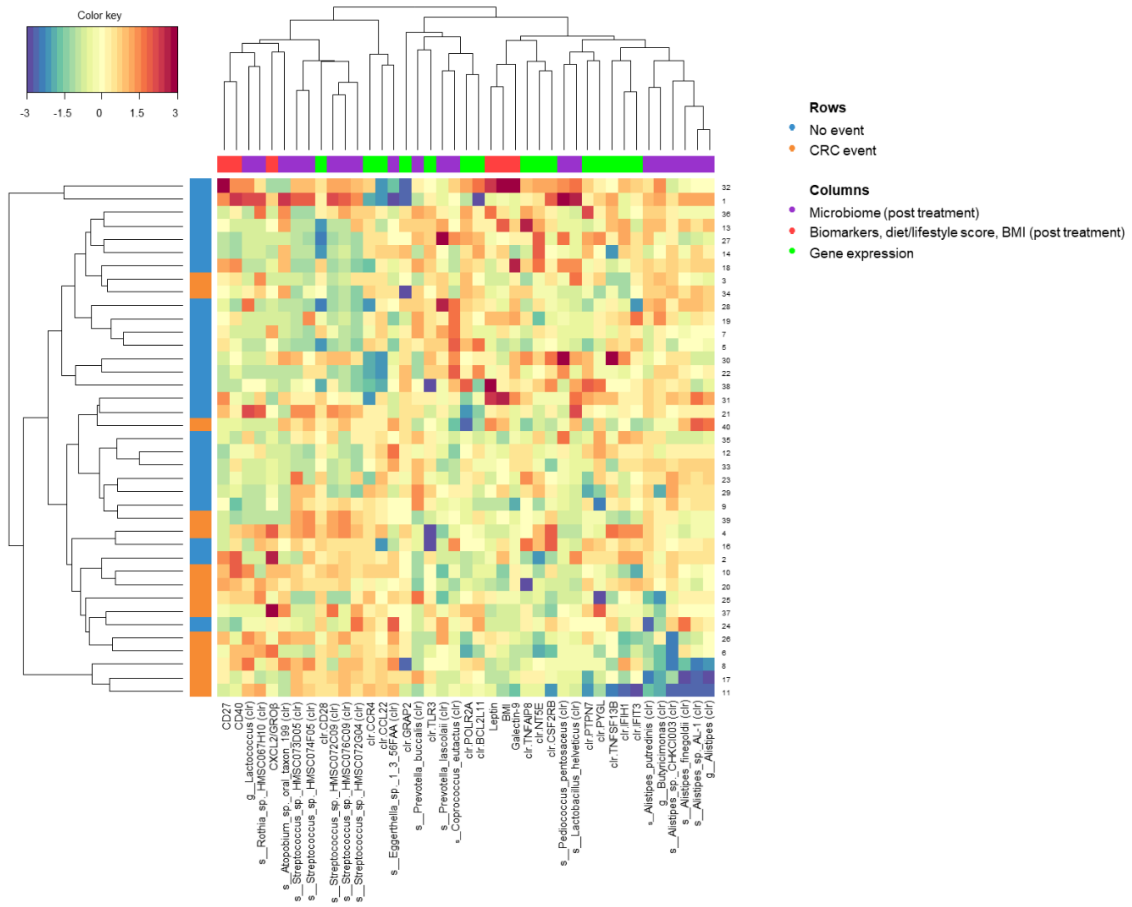


Figure 4.38. Clustered Image Map for the variables selected on component 1 by block sPLS-DA including the occurrence of at least one colorectal event as outcome. Hierarchical clustering of the scaled values of the selected variables was employed based on Euclidian distance and the complete linkage method. The color scale ranges from blue indicating low values to red indicating high values. Patients were displayed in rows while the selected variables were displayed in columns. The "microbiome" block includes the *clr*-transformed abundances of taxa at the end of the treatment; the "biomarker, diet/lifestyle score, BMI" includes the post-treatment values of the circulating biomarkers, the diet/lifestyle score and BMI; the "gene expression" block including the *clr*-transformed expression of the genes.

4.5.5.2 Event-Free Survival analysis

As a final step in our investigation, we included the time dimension through multivariable Cox proportional hazard model and integrated all the findings presented so far to determine if any of the examined factors was associated with a higher risk of developing a colorectal event or experiencing any clinical event (Table 4.11).

Table 4.11 Summary estimates from multivariable Cox Proportional-Hazards models including GE, diet/lifestyle, vitamin D status and Galectin-9

	Any event			Colorectal event		
Characteristic	**HR[1]**	**95% CI[1]**	**p-value**	**HR[1]**	**95% CI[1]**	**p-value**
Gene Expression cluster						
1	—	—		—	—	
2	3.97	1.27, 12.4	**0.018**	4.92	1.48, 16.4	**0.009**
3	2.03	0.19, 21.9	0.56	2.22	0.20, 24.2	0.514
NA	2.48	0.65, 9.46	0.185	3.43	0.86, 13.6	0.08
Diet and lifestyle score (baseline)	0.75	0.47, 1.20	0.228	0.91	0.57, 1.44	0.678
Vitamin D sufficiency (post), Yes vs No	0.28	0.09, 0.89	**0.03**	0.31	0.10, 0.98	**0.046**
Shannon Index (baseline)	0.16	0.04, 0.63	**0.009**	0.18	0.04, 0.73	**0.017**
Season of blood draw						
Winter	—	—		—	—	
Autumn	0.42	0.10, 1.78	0.238	0.92	0.20, 4.27	0.915
Summer	1.78	0.44, 7.26	0.422	3.75	0.80, 17.5	0.092
Spring	0.99	0.25, 3.82	0.984	2.17	0.48, 9.81	0.312
Galectin-9 (baseline) *per 100 pg/mL increase**	0.95	0.92, 0.95	**0.002**	0.96	0.93, 0.99	**0.011**

[1] HR = Hazard Ratio, CI = Confidence Interval

*significant associations also for Galectin-9 at the end of the treatment. Age, sex/gender, tumour treatment and characteristics were not significantly associated with neither outcome. Colorectal events include tumour recurrence, death, adenomas and polyps. Any events include colorectal events and other tumours.

Age, sex/gender, and the clinical characteristics of the tumour were not significantly associated with the risk of colorectal and any clinical event, and were therefore excluded from the models. The same was for BMI, overweight status (BMI > 25) and obesity status (BMI > 30).

Even though the diet/lifestyle score did not reveal a significant association with the risk of clinical and colorectal events (p=0.23 and p=0.68, respectively; Table 4.11), it was kept in the model to account for the lifestyle habits of patients.

A pivotal finding was that achieving vitD sufficiency (25(OH)D > 30) by the end of the treatment period acted as a significant protective factor for both outcomes, resulting in an approximate 70% reduction in risk for both colorectal and any clinical events (Hazard Ratio (HR) = 0.28, 95%CI: 0.09-

0.89; p = 0.03 for clinical events; HR = 0.31, 95%CI: 0.10-0.98; p = 0.046 for colorectal events; Table 4.11).

In line with what already observed in univariate analysis, individuals in *Cluster 2* had a significantly higher risk of both outcomes compared to *Cluster 1* (HR = 3.97, 95%CI: 1.27-12.4; p = 0.02 for clinical events; HR = 4.92, 95%CI: 1.48-16.4; p = 0.01 for colorectal events; Table 4.11). No differences were observed between *Cluster 1* and *Cluster 3*, and between *Cluster 1* and those without the GE evaluation.

Alpha diversity at baseline was also strongly and inversely correlated with the two outcomes, with individuals with low diversity at baseline being at greater risk (HR = 0.16, 95%CI: 0.04-0.63; p = 0.01 for clinical events; HR = 0.18, 95%CI: 0.04-0.73; p = 0.02 for colorectal events; Table 4.11).

While none of the changes in biomarker levels were associated with the risk of either outcome, Galectin-9 was significantly and inversely correlated with the risk of colorectal events and clinical events, both pre- and post-treatment (HR = 0.95, 95%CI: 0.92-0.95; p = 0.002 for clinical events; HR = 0.96, 95%CI: 0.93-0.99; p = 0.01 for colorectal events; Table 4.11). Galectin-9 is a member of the galectin family of proteins and is involved in several biological processes, including the modulation of immune responses, cell-to-cell adhesion, and apoptosis. Several studies have shown a close relationship between Galectin-9 and CRC[275], as it is downregulated in colon tumour tissues[276], and high levels of Galectin-9 were found to be associated with improved overall survival in CRC[276,277].

In Figure 4.39 are the Kaplan-Meier curves including the factors associated with both outcomes, stratified according to their median value.

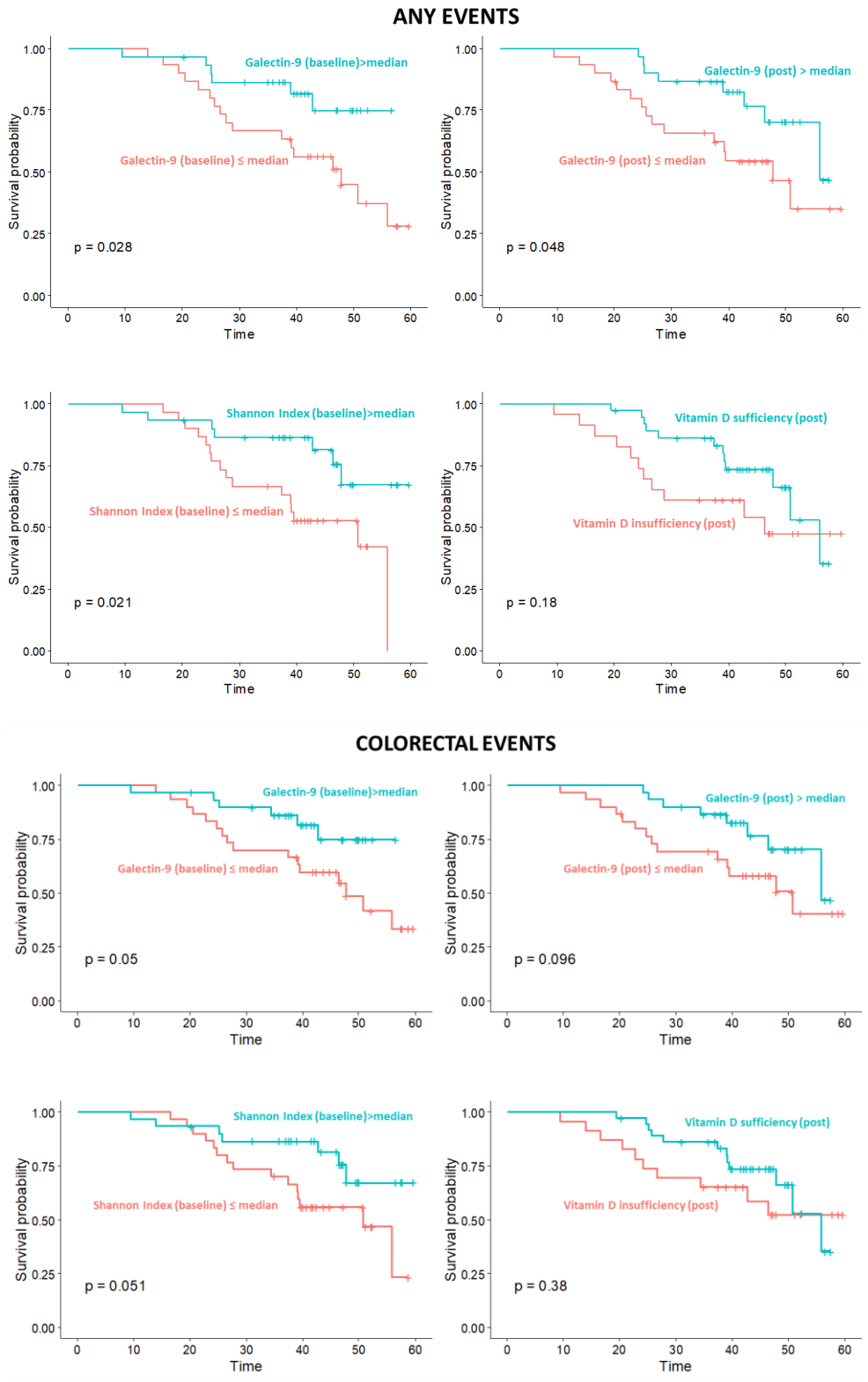


Figure 4.39. Kaplan-Meier curves for event-free survival and colorectal event-free survival and Log-Rank tests according to the median of baseline and post-treatment Galectin-9, the median of the Shannon Index at baseline and vitamin D sufficiency (25(OH)D>60 ng/mL) at the end of the treatment period.
Colorectal events include tumour recurrence, death, adenomas and polyps. Any events include colorectal events and other tumours.

4.5.5.3 Fusobacterium nucleatum, vitamin D and colorectal events

Fusobacterium nucleatum is a common bacterium in the oral cavity known to be significantly associated with CRC and oral diseases. Data on the bacterium prevalence at both timepoints is provided in Figure 4.40, according to treatment arm and colorectal event (median follow-up = 3.7 years).

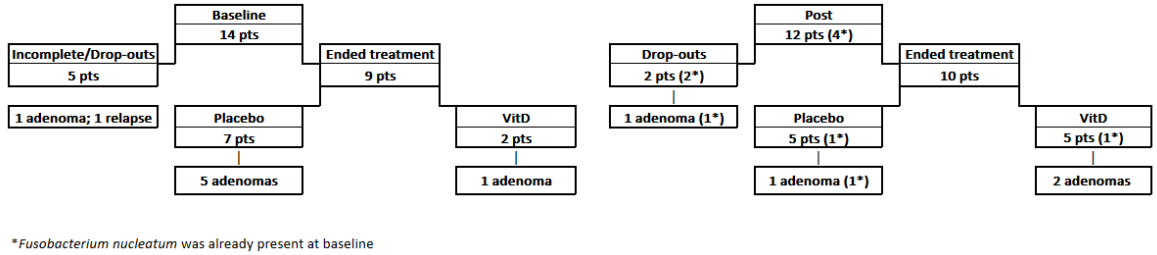


Figure 4.40. Data on prevalence of *Fusobacterium nucleatum* at both time points and according to treatment arm and study compliance. pts = patients.

Due to the short follow-up period, we considered a colorectal event not only death and cancer relapse but also colorectal adenomas and polyps.

In univariate analysis, the colorectal EFS of patients with *F. nucleatum* only at baseline was significantly worse (p = 0.047) (Figure 4.41).

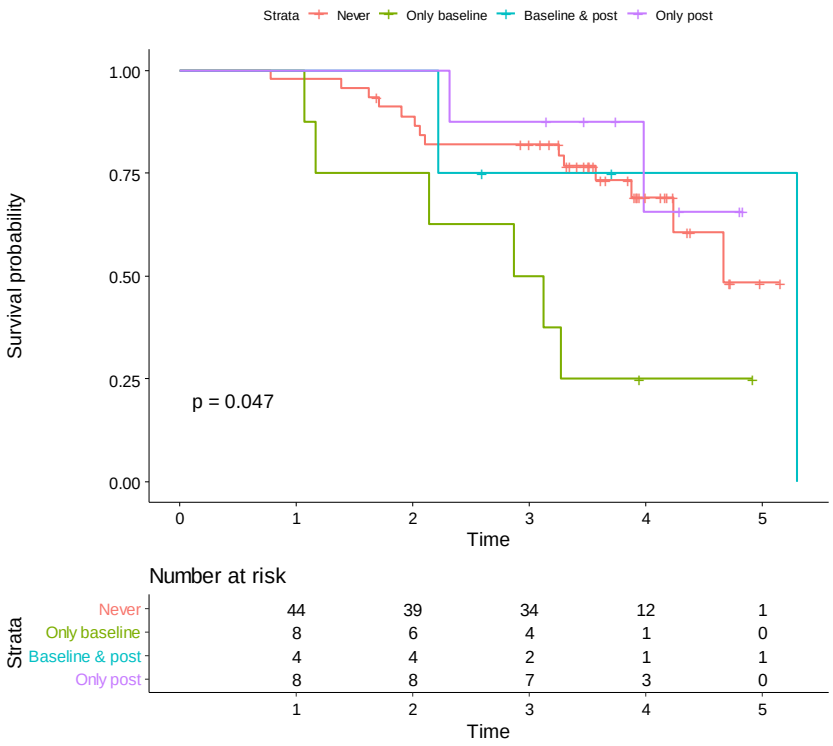


Figure 4.41. Kaplan-Meier curves for event-free survival (EFS) and Log-Rank test according to the presence of *Fusobacterium nucleatum* at either time point.

However, after adjusting for baseline 25(OH)D and post vitD sufficiency (which, in this instance, was a proxy for the treatment effect, having also included drop-outs), the association between *Fusobacterium nucleatum* at baseline and an increased risk of event was significant, regardless of the presence of the bacterium post-treatment (Cox Proportional-Hazards model: HR yes versus no: 3.19; 95%CI: 1.21-8.35; p = 0.019).

However, no significant association was found between the post-treatment presence of the bacterium and risk of colorectal event.

Post-treatment abundances in those with the bacterium were significantly and inversely correlated with age (beta: -0.14; 95%CI: -0.21; -0.08; p = 0.001), significantly higher in those carrying it from baseline (beta: 2.8; 95%CI: 1.3-4.2; p = 0.003) and borderline significantly lower in those reaching vitD sufficiency at the end of the treatment (beta: -1.3; 95%CI: -2.7-0.02; p = 0.05) (Table 4.12; Figure 4.42a).

Table 4.12 Results from multivariate linear regression on log(*F. Nucleatum* relative abundance) at follow-up in those who had it

Characteristic	Beta	95% CI[1]	p-value
Age	-0.14	[-0.21; -0.08]	**0.001**
Vitamin D sufficiency at f.u. (yes vs no)	-1.3	[-2.7; 0.02]	**0.052**
F. nucleatum at baseline (yes vs no)	2.8	[1.3; 4.2]	**0.003**

[1] CI = Confidence Interval; f.u = follow-up; *F. nucleatum = Fusobacterium nucleatum*.
Beta regression coefficients were estimated with a multivariable linear regression model including the log-transformed relative abundance of *Fusobacterium nucleatum* at the end of the treatment period. The model included only the patients with the bacterium at this time point.

In addition, an inverse correlation between *F. nucleatum* and post-treatment 25(OH)D levels was observed, with abundances decreasing as vitD levels increased (Figure 4.42b).

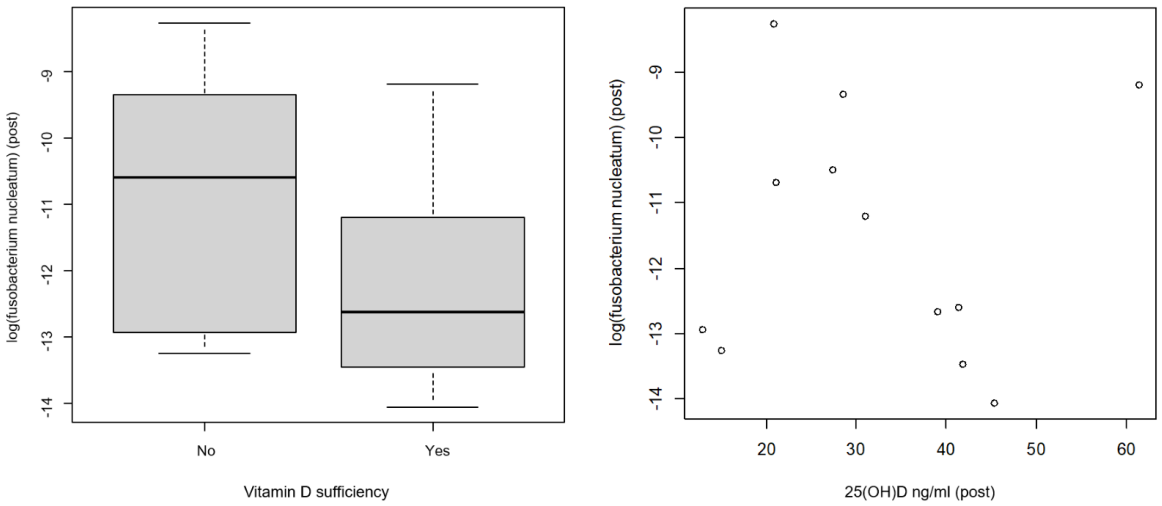


Figure 4.42 a. Boxplots of log-transformed relative abundances of *Fusobacterium nucleatum* in patients who have it at the end of the treatment, according to vitamin D sufficiency status (25(OH)D≥30 ng/ml) at follow-up. p-value comes from the multivariable linear regression model summarized in Table 4.12 **b.** Scatterplot displaying post-treatment 25(OH)D levels in x-axis and log-transformed relative abundances of *Fusobacterium nucleatum* in patients who had it at the end of the treatment in y-axis.

5. DISCUSSION

Overall, our results confirm a relationship between gut microbiome and vitD, and suggest an interplay between them and diet and other risk factors in the CRC setting.

In the case-control study, we identified several taxa significantly more abundant in cases compared to controls, such as *Parvimonas micra*, *F. nucleatum*, and *Bacteroides fragilis*. We also observed that subjects who did not follow WCRF recommendations for cancer prevention had a significantly higher risk of CRC, and that a high-risk diet was associated with a higher inflammatory status and with higher abundances of several species, particularly *F. nucleatum* and *Clostridium ramosum*. Interestingly, we also found an inverse association between CRC risk and high fatty fish consumption but not with other types of fish. Fatty fish is a source of dietary vitD_3, and high consumption can increase serum 25(OH)D. Moreover, the results from the mediation analysis suggested that the microbiota, through its modulation, can mediate the effect of a high risk diet on CRC risk. Altogether, these results showed that the integration of lifestyle risk factors, circulating biomarkers, and microbiome could significantly improve our ability to discriminate healthy subjects from CRC patients. However, the small sample size and the large number of variables considered in the analysis, together with the lack of external validation, are strong limitations of the study. Furthermore, all the variables were measured at a single timepoint, therefore our estimates could be affected by reverse causation bias.

The next step in our research was to conduct a systematic review of the literature on vitD and the microbiota in humans, to determine if the current scientific evidence suggested a possible modulatory effect of vitD on the microbial diversity and composition.

Overall, we found a relationship between vitD and microbiota composition, despite the substantial heterogeneity of the collected studies. Regarding alpha and beta diversity, a large dietary intake of vitD appeared to induce a change in the composition of the bacterial community in some studies, as well as an impact on species richness. At the phylum level, *Firmicutes*, *Actinobacteria* and *Bacteroidetes* were reported to be the most affected, either increasing or decreasing in relation to both vitD supplementation and serum levels.

However, most of the studies identified were observational studies, with only 7 being RCTs. While RCTs are generally less subject to biases, the studies we reviewed often included small sample sizes, with only 4 having over 500 participants. Many observational studies did not account for potential confounders like environment, lifestyle, diet, or sampling times, with the risk of introducing bias in the identified associations between vitD and the microbiota. Additionally, the diverse characteristics and health statuses of participants in these studies sometimes resulted in inconsistent findings.

For this reason, the next step in our research was to design a RCT involving vitD supplementation for CRC survivors with vitD deficiency (25(OH)D<30 ng/mL), to investigate if and

how the supplementation modulated the species known to be beneficial for human health. For vitD supplementation, a daily dose regimen was adopted. In a recent systematic review and individual patient data meta-analysis of RCTs, a daily dosing of vitD supplementation was found to be significantly associated with a 12% reduction in cancer mortality, whereas no effect was observed with high doses vitD supplementation at longer intervals[119].

By the end of the treatment, we observed increased abundances of several taxa in the group of patients supplemented with vitD. Several of them belonged to the *Bacteroides* genus, such as *Bacteroides clarus* and *Bacteroides gallinarum*. Species from Bacteroides are known to play an important role in modulating the human immune system by metabolizing polysaccharides and oligosaccharides, thus supplying the host with nutrition and vitamins[278]. These results are also consistent with what we observed in our systematic review of the literature, where Bacteroidetes emerged as one of the most recurrent phyla increasing following vitD supplementation. *Holdemanella biformis* was also more abundant in supplemented patients. This species was shown to have an anti-tumorigenic effect by producing fatty acids that control tumor cell proliferation[279]. Additionally, *Faecalibacterium prausnitzii*, highly present in the human gut and one of the major gut's butyrate producer, was found in higher abundances in both supplemented and vitD sufficient groups of patients. Known for its anti-inflammatory properties[280], especially in IBD conditions[281], recent studies also suggest a potential protective role of *F. prausnitzii* on both the initiation and progression of CRC[282].

The mediation analysis suggested that vitD supplementation modulated a specific group of taxa, and that this modulation significantly mediated the effect of the supplementation on 25(OH)D levels. These findings are consistent with those observed in our case-control study, in which we found that a high consumption of fatty fish – that is, a vitD-rich diet – significantly increased the levels of *Bifidobacteria/Escheria* ratio (an indicator of "good" intestinal health), thereby decreasing the risk of CRC.

Functional analysis also revealed differences by treatment arm, with *superpathway of glycerol degradation to 1,3-propanediol*, *superpathway of thiamin diphosphate biosynthesis II* (with thiamin diphosphate also known as vitamin B1) and *pathway of guanosine nucleotides degradation II* significantly more abundant in both vitD supplemented and vitD sufficient individuals.

Regarding the change in microbiome throughout the study period, we identified a group of taxa that increased following vitD supplementation, including several strains of *Streptococcus*, *Cryptobacterium curtum*, and the genus *Sutterella*.

Overall, the supplementation did not change the microbial diversity. However, the weight status of patients deeply modulated the effect of the supplementation on both microbiome diversity and serum 25(OH)D levels. Indeed, we observed a significant interaction between vitD supplementation and BMI, with 25(OH)D levels increasing less in the supplemented group at increasing BMI. These results are consistent with a secondary analysis of the VITAL trial that was recently published in

JAMA Network Open[131]. With over 16,000 participants, VITAL is one of the largest randomised controlled trials involving vitD supplementation for cancer and cardiovascular prevention. In the study, the authors investigated the change in serum vitD levels in a sub-cohort of 2,742 participants with a blood sample available at the 2-year follow-up. The authors discovered that supplementation significantly increased all circulating markers related to vitD. However, these increases decreased substantially as BMI categories increased. The same trial also revealed that only normal-weight individuals (BMI <25) in the supplementation group had a lower incidence of invasive cancer events compared to the placebo group, whereas there were no differences in the overweight and obese groups[124].

In our study, we found that weight also modulated the effect of vitD supplementation on the change in alpha diversity, which was significantly and positively correlated with the change in 25(OH)D levels only in the normal-weight individuals. In this subgroup, the change in 25(OH)D levels was also significantly and positively correlated with increasing adiponectin levels and a diet and lifestyle in accordance with WCRF cancer prevention recommendations. On the other hand, in the subgroup of overweight patients, no significant relationship – estimated in terms of partial correlations – was identified among vitD levels, diet/lifestyle, the microbial diversity and any of the investigated circulating markers.

The mechanisms underlying the inverse relationship between BMI and 25(OH)D levels are intricate and remain to be fully elucidated. A prevailing theory posits that, given the fat-soluble nature of vitD, the increased amount of adipose tissue in obese individuals serves as a reservoir, entrapping a greater quantity of vitD and subsequently diminishing its availability in the bloodstream[283,284]. An alternative theory suggests potential modifications in vit D metabolism associated with obesity. This condition may induce alterations in the enzymatic activity within the liver and kidneys responsible for vitD metabolism, impacting the transformation of vitD to its active form[131].

In the era of precision medicine, another pivotal point in research is to consider and account for sex and gender differences throughout investigations. In our study, we identified sex/gender differences in both vitD levels and gut microbiome, especially in the supplemented group. Regarding vitD metabolism, it has been shown that women absorb less vitD and have a different fatty acid metabolism following vitD supplementation compared to men[285]. Moreover, emerging evidence has shown that both biological sex and gender significantly affect gut microbiota. This appears to be attribute not only to sex hormones, but also to host metabolism, gut-brain communication, diet and environmental factors[285–289].

In our trial, we found a significant interaction between gut microbiota and sex/gender on 25(OH)D levels at follow-up. The abundances of pathways related to the biosynthesis of essential amino acids were also significantly different between males and females, but only if supplemented. Sex/gender-specific associations in short-chain acylcarnitines and branched-chain amino acid metabolites were also found in a metabolomics cohort study of critically-ill patients supplemented with high doses of

vitD3[290]. In addition, a mouse study investigating the relationship between dietary vitamin B6 supplementation and colon luminal environment identified significant differences by sex on colonic free amino acids such as *threonine, ornithine, aspargine/aspartate ratio and glutamine/glutamate* ratio[291].

For a subset of 45 patients, we could also assess the GE profile of 395 immune-related genes evaluated in the tumor tissue. Three cluster of patients were identified solely based on GE data. One of these clusters, which we named *Cluster 2*, was significantly associated with a higher risk of experiencing a clinical event during the follow-up period, especially colorectal. The patients experiencing at least one colorectal event were characterized by an overexpression of *CD28*, *CCR4* and *CCL22* genes. Interestingly, the expression of *CCL22* was found to be significantly upregulated in *Fusobacterium nucleatum*-infected CRC cell lines, suggesting a role of *CCL22* in *F. nucleatum*-related colorectal tumorigenesis[274].

Fusobacterium nucleatum is a proinflammatory[292] bacterium of the oral cavity that is highly abundant in CRC patients[293]. However, it is still unclear whether this relationship is just an association or implies a causal involvement of the bacterium in CRC prognosis and progression. In our study, *Fusobacterium nucleatum* was present in 14 patients at baseline and in 12 patients post-treatment. However, looking at preliminary data on clinical events, we found that only patients with *F. nucleatum* at baseline had worse EFS, whereas no association between the bacterium after the treatment and events was observed. This result could indicate that the bacterium is only an indicator of the patient's health status rather than a promoter of tumor carcinogenesis. Moreover, post-treatment abundances of *Fusobacterium nucleatum* were lower in those reaching vitD sufficiency, probably confirming the anti-inflammatory effect of vitD on tumorigenesis[294]. Alpha diversity was also significantly and inversely correlated with risk of both colorectal and any clinical events, with individuals with lower diversity being at higher risk. Low diversity was probably an indicator of a poor health condition in the individuals, as it tended to be lower in those with low baseline adiponectin levels and who had experienced a more advanced CRC. However, it was in these patients that the diversity increased more through the 1-year study period.

Regarding the effect of the supplementation on EFS, we found a significant protective effect of vitD on the risk of onset of clinical events, including colorectal. Indeed, we found that those who reached vitD sufficiency status by the end of the treatment period had a significant reduction in risk of both clinical and colorectal events. While the literature on a protective effect of vitD supplementation on overall and cancer-related mortality is wider and supported by large cohort studies and meta-analyses of RCTs, the evidence on the effect of vitD on cancer risk and progression – including colorectal – is conflicting. In our study, we observed a strong protective effect of vitD supplementation, with about 70% reduction in risk in those reaching 25(OH)D>30 ng/mL at follow-up. These findings can probably be explained by the characteristics of the study population enrolled, which consisted of individuals with vitD deficiency status.

Galectin-9 was also significantly and inversely correlated with the risk of events. Interestingly, a relationship between galectin-9 and vitD levels was observed, with levels of Galectin-9 increasing at increasing levels of 25(OH)D. Galectin-9 is a protein that belongs to the galectin family, which is characterized by the ability to bind beta-galactoside sugars. Galectin-9 is involved in multiple cellular processes, such as cell-cell interactions, cell adhesion, and intracellular signaling. It is also involved in immune modulation, primarily in regulating T-cell responses, and has thus been studied in the context of a variety of diseases, including cancer. Regarding galectin-9 and CRC, low levels of galactin-9 expression were observed in colon tumor tissues, and such low expression was correlated with unfavorable histological grades and the occurrence of lymph node metastasis[276]. Conversely, a high expression of galectin-9 was correlated with improved overall survival in colon cancer patients[276,277]. Moreover, galectin-9 was found to inhibit the growth of CRC cell lines both *in vitro* and *in vivo*. This inhibitory action is believed to arise from the induction of apoptosis through changes in miRNA[295].

Overall, in our trial, we found that the transcriptomic profile, the microbial diversity and composition, vitD status and circulating marker levels of individuals were significantly associated with EFS. However, due to the short follow-up period (median follow-up: 3.7 years) and the resulting small number of advanced tumour recurrences observed to date, we carried out EFS analysis considering as clinical events not only tumour relapses and death, but also colorectal adenomas and polyps, which are mostly benign types of tumours. Therefore, a longer follow-up is necessary to identify robust results and establish causal relationships.

As for the case-control study, the main limitation of the trial is the small sample size, especially when compared to the high number of variables analyzed. Moreover, without a validation set, we could not assess the reproducibility of the results, even though the randomization procedure and the multivariate statistical approaches allowing for confounders adjustment guaranteed a certain degree of estimates reliability.

A consistent and challenging part of our research was dedicated to the study of the literature, in order to figure out the most appropriate computational and statistical methodologies to answer our research questions, taking into consideration the specifics of our data (in particular, the compositional and high-dimensional nature of the microbiome and GE data). With mediation analysis, we were able to go beyond statistical associations and observe a mediating effect of the microbiota on the effect of diet on CRC risk in the observational setting of the case-control study and on the effect of vitD supplementation on 25(OH)D levels in the prospective, interventional setting of the RCT. However, a crucial step of mediation analysis is the definition of the causal pathways linking the data, which needs to be supported by the evidence from the literature, as the formulation of implausible assumptions leads to unreliable results.

In conclusion, our research has provided further evidence on the intricate relationship between modifiable risk factors for CRC, with a specific focus on gut microbiome, vitD status/levels,

diet, lifestyle and circulating biomarkers, especially those related to inflammation and adipokines. Additionally, we explored the role of the transcriptomic profile of immune-related genes evaluated in the tumour tissue on CRC progression.

We identified a fundamental role of both excess weight and sex/gender on the outcomes of vitD supplementation, which resulted in alteration in 25(OH)D levels, microbiome diversity and function, and in the modulation of the investigated markers. These findings are in agreement with emerging evidence from the recent literature and, in the era of precision medicine, highlight the necessity of taking into account these aspects in our investigations. By identifying and comprehending the complex risk factors and their interactions, we can pave the way for more informed and effective strategies in the CRC setting, ultimately leading to enhanced patient care and better health outcomes.

6. SUPPLEMENTARY TABLES

Supplementary Table S1. Diet questionnaire administered to the patients enrolled in the case-control study and in the trial

(Translated from Italian)

We would like to ask you a few questions about your usual (current) diet. In general, how often do you consume a portion of the following foods? ***If you eat a very small or very large portion (compared to the reference portion), halve or double the frequency of consumption.***

FOODS		**FREQUENCY**				
	Portion of reference	Rarely (never/ 1-2 times a month)	Once a week	2-3 times a week	Every day	More than once a day
SCORE		1	2	3	4	5
Q1. Cow's milk (whole, partly skimmed or skimmed)	125 gr	❑	❑	❑	❑	❑
Q2. Yoghurt (all types) (no soya yoghurt)	1 portion	❑	❑	❑	❑	❑
Q3. Pasta or rice (dry)	80 gr	❑	❑	❑	❑	❑
Q4. Soups	1 portion	❑	❑	❑	❑	❑
Q5. Bread (white, whole grain or seasoned)	50gr	❑	❑	❑	❑	❑
Q6. Crackers, breadsticks or rusks	1 packet 30gr	❑	❑	❑	❑	❑
Q7. Pizza	150 gr	❑	❑	❑	❑	❑
Q8. Grated cheese, on pasta dishes and soups	1 teaspoon	❑	❑	❑	❑	❑
Q9. Meat (all types)	100 gr	❑	❑	❑	❑	❑
Q10. Liver, all animals	100 gr	❑	❑	❑	❑	❑
Q11. Processed meat (ham, salami, bresaola, sausages etc.)	50 gr	❑	❑	❑	❑	❑
Q12. Fish (salmon, herring, mackerel)	150 gr	❑	❑	❑	❑	❑
Q13. Fish (other types)	150 gr	❑	❑	❑	❑	❑
Q14. Eggs (hard-boiled, omelette, soft-boiled)	n. 1	❑	❑	❑	❑	❑
Q15. Fresh cheeses (mozzarella, ricotta, robiola)	100 gr	❑	❑	❑	❑	❑
Q16. Matured cheeses (emmenthal, provolone, caciotta etc.)	50 gr	❑	❑	❑	❑	❑
Q17. Soy-based products	100gr	❑	❑	❑	❑	❑

Q18. Vegetables (all types), raw, cooked, including salad	250 gr	❑	❑	❑	❑	❑
Q19. Boiled, mashed, roasted, fried potatoes	200 gr	❑	❑	❑	❑	❑
Q20. Fresh fruit (all types)	150 gr	❑	❑	❑	❑	❑
Q21. Dried and shelled fruit	30 gr	❑	❑	❑	❑	❑
Q22. Ice cream (no ice lollies and sorbets)	100 gr	❑	❑	❑	❑	❑
Q23. Chocolate	8 gr	❑	❑	❑	❑	❑
Q24. Other sweets (cake, brioche, snacks)	50 gr	❑	❑	❑	❑	❑

Supplementary Table S2. Phylogenetic reconstruction of taxa that significantly decreased after vitamin D supplementation (Supplementation group)

Author, PY	Health Status	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Bashir, 2016	Healthy	Biopsy	Upper GI: GC (n paired = 13)	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	NA	Escherichia/Shigella
				Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	
				Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus	
				Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Variovorax	
				Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Enterobacteriaceae unclass	
				Proteobacteria	Gammaproteobacteria				
			Upper GI: GA (n paired = 13)	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	NA	Escherichia/Shigella
				Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia	
				Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	
				Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	
				Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	Enterobacteriaceae unclass	
				Proteobacteria	Gammaproteobacteria				
			Upper GI: DD (n paired = 13)	Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae	NA	Escherichia/Shigella
				Actinobacteria	Actinomycetia	Micrococcales	Microbacteriaceae	Leucobacter	
				Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	
			Lower GI: TI (n paired = 11)	Firmicutes	Clostridia	Eubacteriales	Peptostreptococcaceae	Peptostreptococcus	
			Lower GI: AO (n paired = 11)	Firmicutes	Clostridia			Clostridia unclass.	
			Lower GI: SC (n paired = 11)						
			Lower GI: AC (n paired = 12)						
		Stool	Stool (n paired = 8)	Proteobacteria	Betaproteobacteria				
Bosman,2019	Healthy (female)	Stool							
Cantarel, 2015	Healthy+MS (female)	Stool		Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
	Healthy (female)								
	MS (female)		Untreated MS vs HC or treated MS	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae		
			Treated vs HC or treated MS	Firmicutes	Clostridia	Eubacteriales	Eubacteriaceae	Eubacterium	
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
Charoenngam, 2020	Healthy	Stool		Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Faecalibacterium	
				Firmicutes	Clostridia	Eubacteriales	Ruminococcaceae		
				Firmicutes	Clostridia				

PY = Publication Year; NA= Not Available; GI = gastrointestinal; GC = gastric corpus; GA = gastric antrum; DD = duodenum; TI = terminal ileum; AO = appendiceal orifice; AC = ascending colon; SC = sigmoid colon; MS = Multiple Sclerosis; HC = Healthy Controls; CD = Crohn disease; Q3 = upper quartile; Q1 = lower quartile.

Supplementary Table S3. Phylogenetic reconstruction of taxa that significantly decreased after vitamin D supplementation (Supplementation group)

Author, PY	Health Status	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Ciubotaru, 2015	Prediabetes (males)	Stool	25(OH)D (Q3 vs Q1)	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Roseburia	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae		
			Delta 25(OH)D (Q3 vs Q1)	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Roseburia	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Dorea	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae		
Drall, 2020	Pregnancy (infants)	Stool	Infant vit D supplementation	Firmicutes	Negativicutes	Selenomonadales	Selenomonadaceae	Megamonas	
				Firmicutes	Negativicutes	Veillonellales	Veillonellaceae		
			Maternal prenatal or postnatalvit D suppl	Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Bilophila (only breastfed)	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Other (only brestfed)	
Garg, 2018		Stool							
Hjelmsø, 2020	Pregnancy	Infant stool							
Kanhere, 2018	Cystic fibrosis	Stool	Stool: vit D sufficient vs vit D insufficient at baseline	Proteobacteria	Gammaproteobacteria				
		Stool	Stool: change in microbiotaafter supplementation	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Anaerotruncus	
				Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Veillonella	
				Firmicutes	Clostridia	Eubacteriales	Clostridiaceae		
				Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae		
Missailidis,2019	HIV	Biopsy							
Naderpoor, 2018	Obesity	Stool	Vit D suppl. vs Placebo at follow-up	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	
			25(OH)D>75 nmol/L vs 25(OH)D<50 nmol/L at follow-up	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
				Firmicutes	Clostridia	Eubacteriales	Clostridiaceae		
Schaffler, 2018	Crohn disease; Healthy	Stool	CD: Week 4						
			HC						
Singh,2020	Healthy (female)	Stool	Main analysis	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Roseburia	
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Faecalibacterium	
				Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	
				Firmicutes					

PY = Publication Year; NA= Not Available; 25(OH)D = 25 hydroxyvitamin D; HC = Healthy Controls; CD = Crohn Disease; Q3 = upper quartile; Q1 = lower quartile.

Supplementary Table S4. Phylogenetic reconstruction of taxa that significantly decreased after vitamin D supplementation (Supplementation group)

Author, PY	Health Status	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Singh,2020	Healthy (female)	Stool	Responders (>20 ng/ml) vs non-responders (<20 ng/ml)						
			Responders	Firmicutes					
			Non-responders	Proteobacteria					
Sordillo,2016	Healthy	Stool		Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus	
Tabatabaeizadeh, 2019	Healthy (female, adolescents)	Stool		Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	
				Bacteroidetes					
Talsness, 2017	Pregnancy (infants)	Stool	Vit D supplementation (none, <10mg, >=10mg)	Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium sp
			25(OH) levels (quintiles)	Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	Bifidobacterium sp
				Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroides fragilis
			Infant vit D suppl. (yes vsno)						

PY = Publication Year; NA= Not Available; 25(OH)D = 25 hydroxyvitamin D; vit D = vitamin D.

Supplementary Table S5. Phylogenetic reconstruction of taxa that significantly increased after vitamin D supplementation (Supplementation group)

Author, PY	Health Status	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Bashir, 2016	Healthy	Biopsy	Upper GI: GC (n paired = 13)	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Bradyrhizobiaceae	Bradyrhizobium	
				Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Sulfurospirillum	
				Actinobacteria	Actinomycetia	Actinomycetales	Actinomycetaceae	Actinomyces	
			Upper GI: GA (n paired = 13)	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	Alkalibacterium	
				Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Bradyrhizobiaceae	Bradyrhizobium	
				Proteobacteria	Alphaproteobacteria				
			Upper GI: DD (n paired = 13)	Proteobacteria	Alphaproteobacteria	Hyphomicrobiales	Bradyrhizobiaceae	Bradyrhizobium	
				Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Janthinobacterium	
				Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Halomonas	
				Bacteroidetes				Bacteroidetes unclass.	
			Lower GI: TI (n paired = 11)	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Roseburia	
			Lower GI: AO (n paired = 11)						
			Lower GI: SC (n paired = 11)						
			Lower GI: AC (n paired = 12)						
		Stool	Stool (n paired = 8)	Actinobacteria	Actinomycetia	Actinomycetales	Actinomycetaceae	Actinomyces	
Bosman,2019	Healthy (female)	Stool		Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Lachnospira	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Fusicatenibacter	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae		
Cantarel, 2015	Healthy+MS (female)	Stool	Untreated MS vs HC or treated MS	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Faecalibacterium	
				Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae		
	Healthy (female)								
	Multiple Sclerosys (female)			Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Akkermansiaceae	Akkermansia	
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Faecalibacterium	
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Coprococcus	
			Treated vs HC or treated MS	Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	Janthinobacterium	
Charoenngam, 2020	Healthy	Stool		Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	
				Bacteroidetes	Bacteroidia	Bacteroidales	Tannerellaceae	Parabacteroides	
Ciubotaru, 2015	Prediabetes (males)	Stool	25(OH)D (Q4 vs Q1)						
			Delta 25(OH)D (Q4 vs Q1)						
Drall, 2020	Pregnancy (infants)	Stool	Infant vit D suppl.						
			Maternal prenatal or postnatal vit D suppl	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus (only breastfed)	

PY = Publication Year; NA= Not Available; GI = gastrointestinal; GC = gastric corpus; GA = gastric antrum; DD = duodenum; TI = terminal ileum; AO = appendiceal orifice; AC = ascending colon; SC = sigmoid colon; MS = Multiple Sclerosis; HC = Healthy Controls; CD = Crohn disease; Q4 = upper quartile; Q1 = lower quartile.

Supplementary Table S6. Phylogenetic reconstruction of taxa that significantly increased after vitamin D supplementation (Supplementation group)

Author, PY	Health Status	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Garg, 2018		Stool		Firmicutes	Clostridia	Eubacteriales	Clostridiaceae	Clostridium	Clostridium colinae
				Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae		
Hjelmsø, 2020	Pregnancy	Infant stool							
Kanhere, 2018	Cystic fibrosis	Stool	Stool: vit D sufficient vs vit D insufficient at baseline	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	
				Bacteroidetes	Bacteroidia	Bacteroidales	Tannerellaceae	Parabacteroides	
				Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae		
				Bacteroidetes	Bacteroidia	Bacteroidales	Porphyromonadaceae		
		Stool	Stool: change in microbiotaafter suppl.	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus	
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
				Firmicutes	Negativicutes	Acidaminococcales	Acidaminococcaceae	Acidaminococcus	
				Firmicutes	Negativicutes	Acidaminococcales	Acidaminococcaceae	Phascolarctobacterium	
				Bacteroidetes	Bacteroidia	Bacteroidales	Odoribacteraceae		
				Bacteroidetes	Bacteroidia	Bacteroidales	Paraprevotellaceae		
Missailidis,2019	HIV	Biopsy							
Naderpoor, 2018	Obesity	Stool	Vit D suppl. vs Placebo at follow-up	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Lachnospira	
			25(OH)D>75 nmol/L vs 25(OH)D<50 nmol/L at fullow-up	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Coprococcus	Coprococcus eutactus
				Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Coprococcus	
Schaffler, 2018	Crohn disease; Healthy	Stool	CD: Week 4	Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	
				Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Megasphaera	
			HC						
Singh,2020	Healthy (female)	Stool	Main analysis	Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	
				Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Akkermansiaceae	Akkermansia	
				Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	
				Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	
				Bacteroidetes	Bacteroidia	Bacteroidales	Tannerellaceae	Parabacteroides	
				Bacteroidetes					
			Responders (>20 ng/ml) vs non-responders (<20 ng/ml)	Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroides acidifaciens
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	Ruminococcus bromii
				Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroides eggerthii
				Bacteroidetes	Bacteroidia	Bacteroidales	Barnesiellaceae	Barnesiella	Barnesiella intestinihominis

PY = Publication Year; 25(OH)D = 25 hydroxyvitamin D; HC = Healthy Controls; CD = Crohn Disease; vit D = vitamin D.

Supplementary Table S7. Phylogenetic reconstruction of taxa that significantly increased after vitamin D supplementation (Supplementation group)

Author, PY	Health Status	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Singh,2020	Healthy (female)	Stool	Responders	Bacteroidetes					
				Actinobacteria					
				Proteobacteria					
				Lentisphaeraea					
			Non-responders	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Roseburia	Roseburia faecis
				Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	Bacteroides eggerthii
				Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	Prevotella copri
				Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Oscillospira	Oscillospira guilliermondii
				Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	Alistipes	Alistipes finegoldii
Sordillo,2016	Healthy	Stool		Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Lachnobacterium	
Tabatabaeizadeh, 2019	Healthy (female, adolescents)	Stool		Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Enterococcus	
				Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	
				Firmicutes					
Talsness, 2017	Pregnancy (infants)	Stool	Vit D supplementation (none, <10mg, >=10mg)						
			25(OH) levels (quintiles)						
			Infant vit D suppl. (yes vs no)						

PY = Publication Year; 25(OH)D = 25 hydroxyvitamin D; vit D = vitamin D.

Supplementary Table S8. Phylogenetic reconstruction of taxa that were significantly and negatively associated with either vitamin D serum concentrations or intake (Non-supplementation group)

Author, PY	Health Status	Vit D	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Kassem, 2020	Pregnancy	Prenatal maternal 25[OH]Dand cord 25[OH]D	Stool	Prenatal maternal 25(OH)D	Firmicutes	Tissierellia	Tissierellales	Peptoniphilaceae	Anaerococcus	
					Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	
				Cord 25(OH)D	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Mediterraneibacter	Ruminococcus gnavus
Luthold, 2017	Healthy	Dietary vit D intake	Stool	Dietary Vit D intake tertiles	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Veillonella	
					Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	
Luthold, 2017	Healthy	25(OH)D	Stool	25(OH)D concentrations tertiles	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Veillonella	
					Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Coprococcus	
					Actinobacteria	Actinomycetia	Bifidobacteriales	Bifidobacteriaceae	Bifidobacterium	
Mandal, 2016	Pregnancy	Dietary vit D intake	Stool	Maternal microbiota	Bacteroidetes					
Seura, 2017	Healthy (female)	Dietary vit D intake	Stool							
Soltys, 2020	Ulcerative Colitis	Serum Vit D levels	Stool	Stool						
			Biopsy	Biopsy: sigma inflamed	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	Haemophilus parainfluenzae
					Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	
					Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	
					Firmicutes	Bacilli	Lactobacillales	Streptococcaceae		
					Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae		
					Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae		
					Proteobacteria	Gammaproteobacteria	Pasteurellales			
					Fusobacteria	Fusobacteriia	Fusobacteriales			
					Fusobacteria					
				Biopsy: sigma non-inflamed	Actinobacteria	Coriobacteriia	Coriobacteriales	Coriobacteriaceae	Collinsella	Collinsella aerofaciens
					Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	
					Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae		
					Fusobacteria	Fusobacteriia	Fusobacteriales			
					Actinobacteria					
					Fusobacteria					

PY = Publication Year; 25(OH)D = 25 hydroxyvitamin D; vit D = vitamin D.

Supplementary Table S9. Phylogenetic reconstruction of taxa that were significantly and negatively associated with either vitamin D serum concentrations or intake (Non-supplementation group)

Author, PY	Health Status	Vit D	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Soltys, 2020	Crohn disease	25(OH)D	Stool	Stool						
			Biopsy	Biopsy: sigma inflamed	Firmicutes					
				Biopsy: sigma non-inflamed						
				Biopsy: terminal ileum inflamed						
				Biopsy: terminal ileum non-inflamed						
Thomas, 2020	Healthy (male, older)	25(OH)D; 1,25(OH)$_2$D; 24,25(OH)$_2$D; activation ratio (1,25(OH)$_2$D/25(OH)D) and catabolism ratio	Stool	1,25(OH)$_2$D	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Oscillospira	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	
					Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Anaerotruncus	
				Activation ratio	Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Oscillospira	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	
Weng, 2019	Ulcerative Colitis; Healthycontrols	Dietary vit D intake	Biopsy		Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Dorea	
					Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus 2	
			Stool		Firmicutes	Clostridia	Eubacteriales	Clostridiaceae	Clostridium	Clostridium clostridioforme CAG:132
	Crohn disease; Healthy controls	Dietary vit D intake	Biopsy		Firmicutes	Bacilli	Lactobacillales	Lactobacillaceae	Lactobacillus	
					Actinobacteria	Actinomycetia	Micrococcales	Intrasporangiaceae	Janibacter	
					Proteobacteria	Hydrogenophilalia	Hydrogenophilales	Hydrogenophilaceae	Hydrogenophilus	
			Stool							
Wu, 2011	Healthy	Dietary Vit D intakes	Stool		Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Dialister	

PY = Publication Year; 25(OH)D = 25 hydroxyvitamin D; 1,25(OH)2D = 1,25 hydroxyvitamin D2; 24,25(OH)2D = 24,25 hydroxyvitamin D2; vit D = vitamin D.

Supplementary Table S10. Phylogenetic reconstruction of taxa that were significantly and positively associated with either vitamin D serum concentrations or intake (Non-supplementation group)

Author, PY	Health Status	Vit D	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Kassem, 2020	Pregnancy	Prenatal maternal 25(OH)D andcord 25(OH)D	Stool	Prenatal maternal 25(OH)D	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Mediterraneibacter	Ruminococcus gnavus
					Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	
					Actinobacteria	Actinomycetia	Corynebacteriales	Corynebacteriaceae	Corynebacterium	
					Firmicutes	Clostridia	Eubacteriales	Clostridiaceae		
				Cord 25(OH)D	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	Acinetobacter rhizosphaerae
					Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	
					Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Bulleidia	
					Actinobacteria	Actinomycetia	Corynebacteriales	Corynebacteriaceae	Corynebacterium	
					Firmicutes	Tissierellia	Tissierellales	Peptoniphilaceae	Finegoldia	
					Firmicutes	Tissierellia	Tissierellales	Peptoniphilaceae	Peptoniphilus	
					Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus	
					Firmicutes	Clostridia	Eubacteriales	Clostridiaceae		
					Proteobacteria	Gammaproteobacteria	Enterobacterales	Enterobacteriaceae		
Luthold, 2017	Healthy	Dietary vit D intake	Stool	Dietary Vit D intake tertiles	Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	
Luthold, 2017	Healthy	25(OH)D	Stool	25(OH)D concentrations tertiles	Firmicutes	Negativicutes	Veillonellales	Veillonellaceae	Megasphaera	
Mandal, 2016	Pregnancy	Dietary vit D intake	Stool	Maternal microbiota	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus	
Seura, 2017	Healthy (female)	Dietary vit D intake	Stool/Biopsy							
Soltys, 2020	Ulcerative Colitis	25(OH)D	Stool	Stool						
			Biopsy	Biopsy: sigma inflamed						
				Biopsy: sigma non-inflamed						
Soltys, 2020	Crohn disease	25(OH)D	Stool	Stool	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	
					Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae		
					Proteobacteria	Gammaproteobacteria	Pasteurellales			
			Biopsy	Biopsy: sigma inflamed						
				Biopsy: sigma non-inflamed						
				Biopsy: terminal ileum inflamed						
				Biopsy: terminal ileum non-inflamed						

PY = Publication Year; 25(OH)D = 25 hydroxyvitamin D; vit D = vitamin D intake.

Supplementary Table S11. Phylogenetic reconstruction of taxa that were significantly and positively associated with either vitamin D serum concentrations or intake (Non-supplementation group)

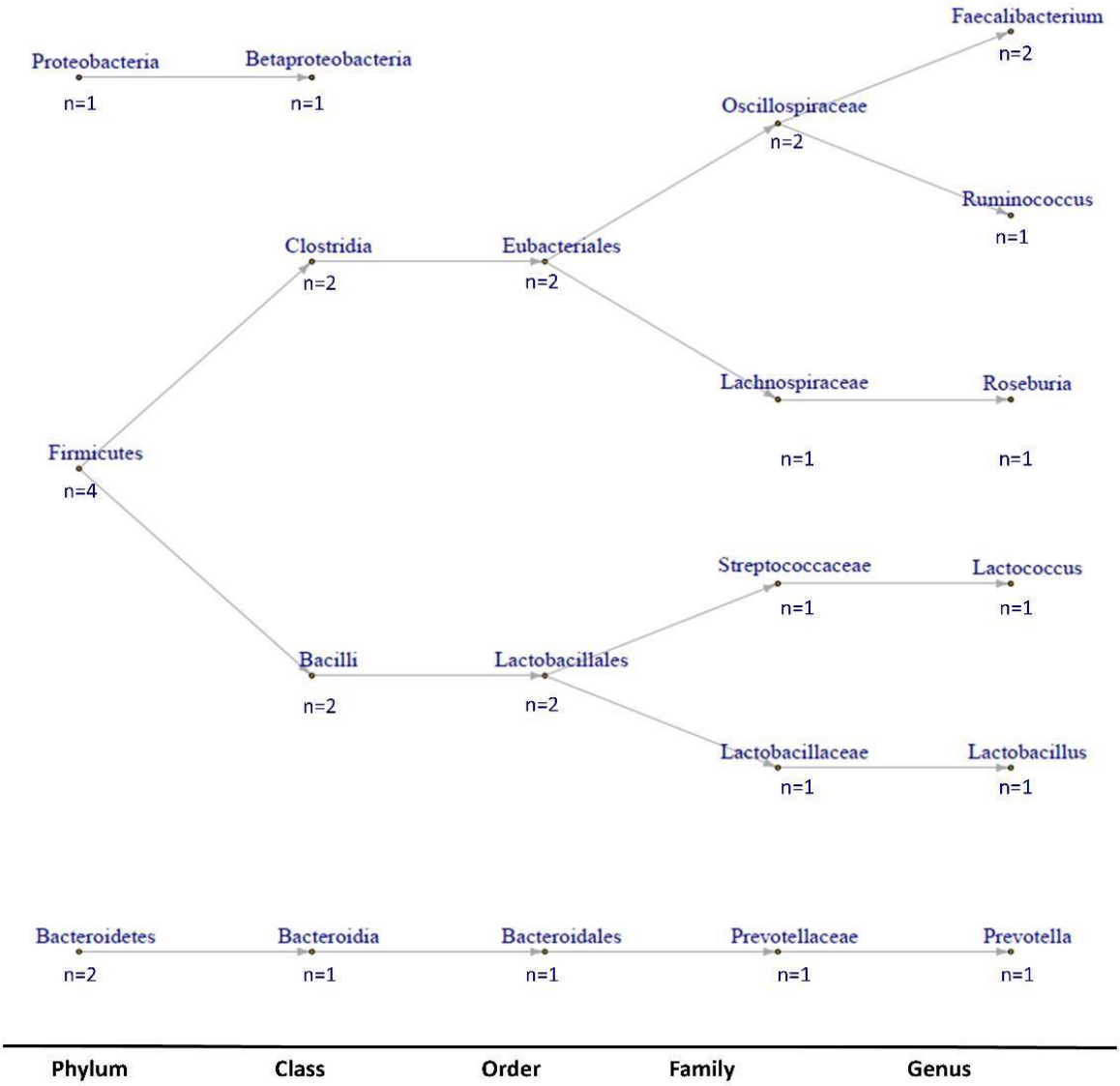
Author, PY	Health Status	Vit D	Sample	Stratification	Phylum	Class	Order	Family	Genus	Species
Thomas, 2020	Healthy (male, older)	25(OH)D; 1,25(OH)2D; 24,25(OH)2D; activation ratio (1,25(OH)2D/25(OH)D) and catabolism ratio	Stool	1,25(OH)2D	Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Coprococcus	Coprococcus catus
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	Blautia Obeum
				Activation ratio	Firmicutes	Clostridia	Eubacteriales	Eubacteriales Family XIII. Incertae Sedis	Mogibacterium	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Coprococcus	
					Firmicutes	Clostridia	Eubacteriales	Ruminococcaceae		
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae		
					Lentisphaerae	Lentisphaeria	Victivallales	Victivallaceae		
					Firmicutes	Clostridia	Eubacteriales			
Weng, 2019	Ulcerative Colitis; Healthy controls	Dietary vit D intake	Biopsy		Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Bilophila	
			Stool		Proteobacteria	Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	Desulfovibrio	
					Bacteroidetes	Bacteroidia	Bacteroidales	Barnesiellaceae	Barnesiella	
	Crohn disease; Healthy controls	Dietary vit D intake	Biopsy		Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Fusicatenibacter	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Blautia	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Lachnospiracea incertae sedis	
			Stool		Firmicutes	Clostridia	Eubacteriales	Oscillospiraceae	Ruminococcus	
					Firmicutes	Clostridia	Eubacteriales	Lachnospiraceae	Fusicatenibacter	
					Proteobacteria	Oligoflexia	Bdellovibrionales	Bdellovibrionaceae	Bdellovibrio	
					Bacteroidetes	Bacteroidia	Bacteroidales	Barnesiellaceae	Barnesiella	
Wu, 2011	Healthy	Dietary Vit D intakes	Stool		Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	

PY = Publication Year; 25(OH)D = 25 hydroxyvitamin D; 1,25(OH)$_2$D = 1,25 hydroxyvitamin D2; 24,25(OH)$_2$D = 24,25 hydroxyvitamin D$_2$; vit D = vitamin D

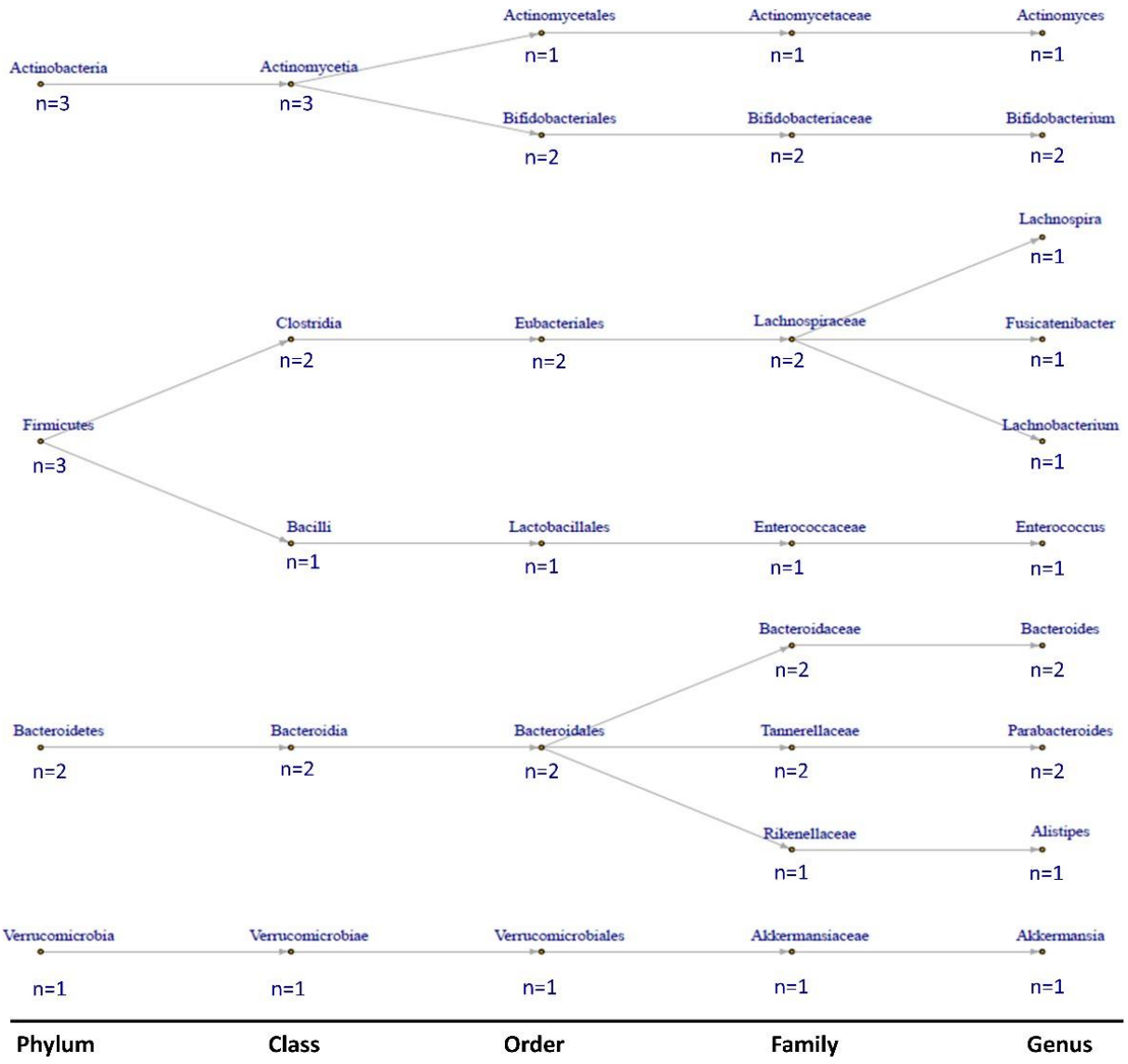
Supplementary Table S12. List of genes significantly expressed in the 3 GE cluster

List of genes	
AXL	KLRD1
BCL6	LMNA
CBLB	LRP1
CCL5	MAPK1
CCNB2	MIF
CD3D	MKI67
CD47	MMP2
CD79A	MRC1
CD80	NCF1
CEACAM1	NKG7
CXCL13	OAS1
EGFR	POU2AF1
EGR2	RPS6
EGR3	SKAP2
FOXM1	SLAMF7
FOXO1	SNAI2
GUSB	TCF7
HLA-DOB	TNFRSF17
IL18	TOP2A
IRS1	TWIST1
KLF2	VCAM1

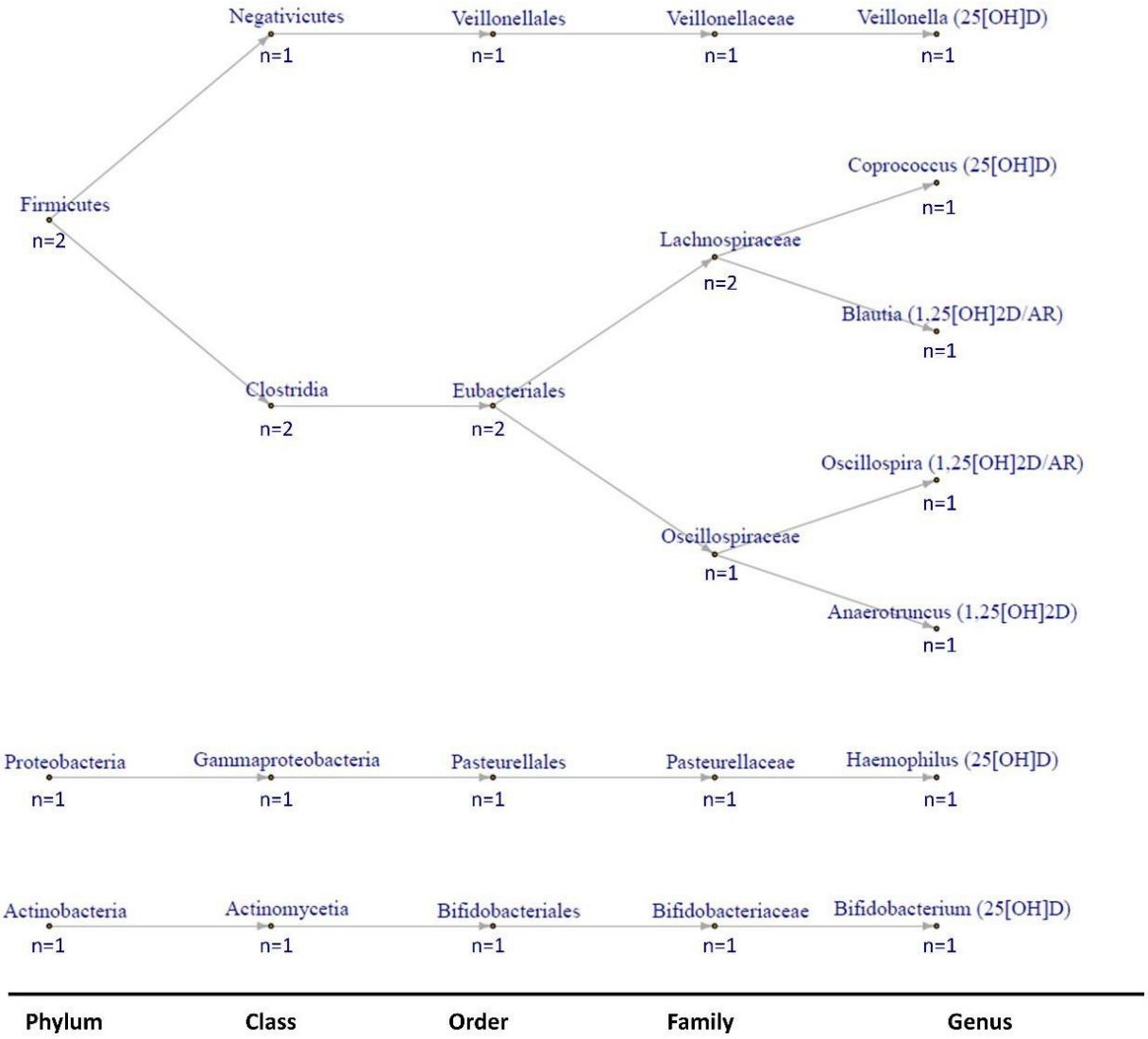
7. SUPPLEMENTARY FIGURES



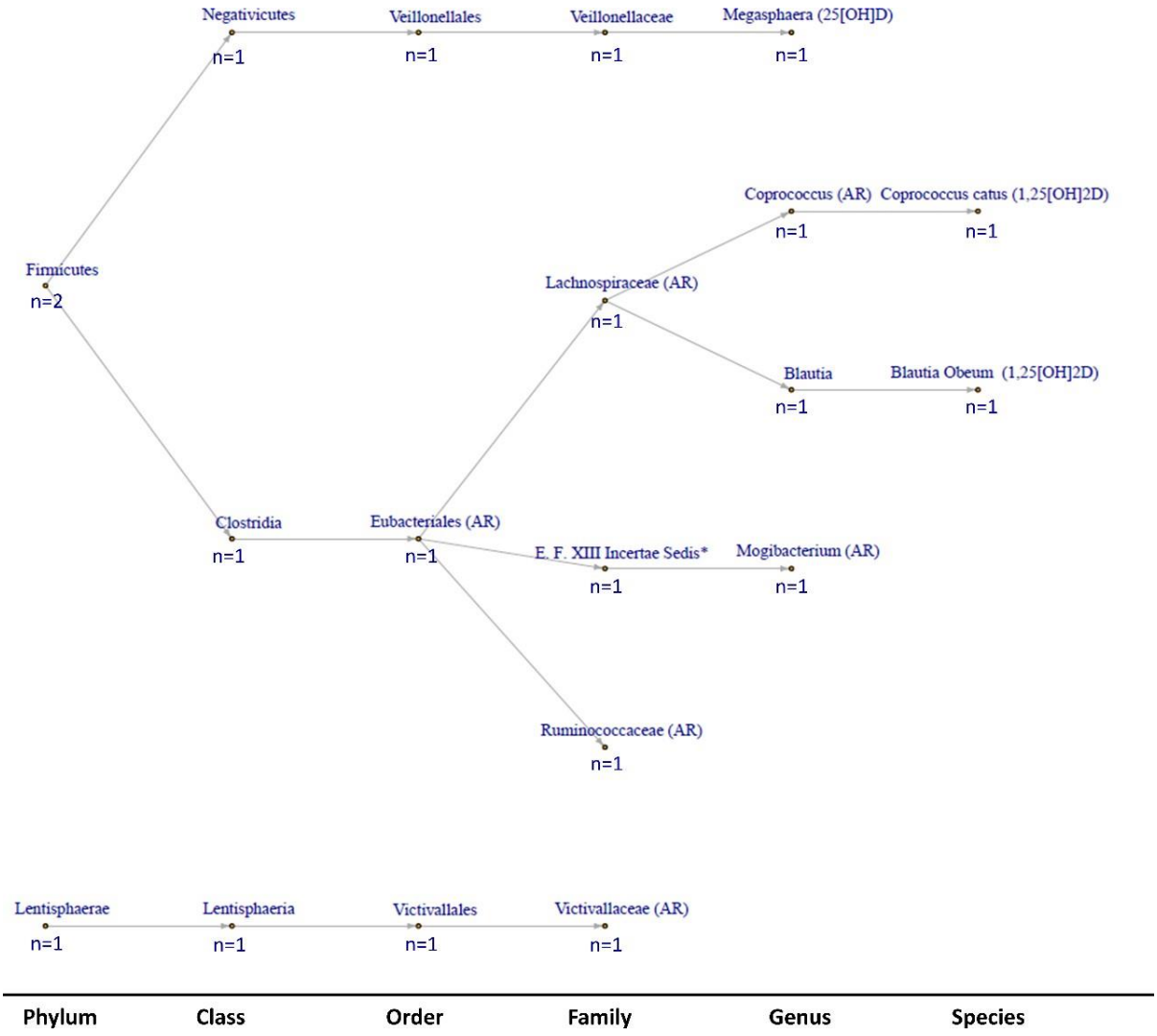
Supplementary Figure S1. Phylogenetic tree of taxa that significantly decreased after vitamin D supplementation (supplementation group).



Supplementary Figure S2. Phylogenetic tree of taxa that significantly increased after vitamin D supplementation (supplementation group).



Supplementary Figure S3. Phylogenetic tree of taxa that were significantly and negatively associated with either vitamin D serum concentrations or intake (non-supplementation group).
AR = Activation ratio of vitamin D, defined as 1,25(OH)2D/25(OH)D; 25(OH)D = 25hydroxyvitamin D; 1,25(OH)2D = 1,25 hydroxyvitamin D2.



Supplementary Figure S4. Phylogenetic tree of taxa that were significantly and positively associated with either vitamin D serum concentrations or intake (Non-supplementation group).
AR= Activation ratio of vitamin D, defined as 1,25(OH)2D/25(OH)D; 25(OH)D = 25 hydroxyvitamin D; 1,25(OH)2D = 1,25 hydroxyvitamin D2. *Eubacteriales Family XIII. Incertae Sedis.

ACKNOWLEDGMENTS

Over the past four years, I have had the privilege to undergo incredible experiences, to visit beautiful places, and to take part in projects I once only dreamed of. I am truly grateful for all the opportunities and for all the knowledge imparted by the professionals I have worked and engaged with. Everything I know has been shaped by their guidance, and for this, I want to express my heartfelt thanks to everyone who has shared their time and expertise with me during this time.

Even though a PhD is supposed to be a milestone in one's professional career, the historical moment in which it took place added layers of emotional complexity to the experience. For this reason, I would like to use this section to extend my gratitude to all the people who have offered unwavering support and guidance throughout this journey:

To my wonderful supervisor, Sara Gandini, who has given me the two most invaluable things I could get from this experience: a chance and complete trust, in the most selfless way. I cannot even begin to explain how much she has taught me, both professionally and humanly. Every day I learn from her, I am challenged by her and I feel grateful to have met her. She is nothing but an inspiration for me and I will always look up to her. So thank you Sara, for everything. I will always be in your debt.

To Sara Raimondi, for the patience, the kindness and the understanding she has always had for me. I have learned a lot from her and I am truly happy to have had the opportunity to work together. I hope that the future will bring you everything you deserve.

To my colleagues, Oriana D'Ecclesiis, Elisa Tomezzoli, Aurora Gaeta, Sofia Netti. In a world often marked by competition and individualism, you have always shown me the true essence of a group. Thank you, from the bottom of my heart, for all the support, the friendship and the sage advice without which I could have never made it. There is so much of you in this.

To my mother, who has always given me the trust to be whatever and whoever I thought I could be. I hope this repays you, at least in part, for all the sacrifices you have made for us. Things would be so different if you were not my mom, and I will always be grateful.

To Laura and Alessio. I must have done something amazing in another life to get to have you in this one. I can't even count how many times I have wanted to give up during these four years, feeling out of place and not up to the task. And every time – every single time – you have always put everything back into perspective. So, this is all for you and because of you.

To my grandparents. You taught me family, you showed me love and you gave me a happy place to be. Nothing has made sense since you have been gone and we are so lost. I would pay a million dollars to see you cry in this moment, grandpa.

To Rinaldo, for his heart and all the good that it does. You changed everything with a message and with your essence, and I will always be grateful to the madness of life for giving me the chance to meet you.

To my dearest teacher, Gioia Rustichelli. I will never forget your invaluable teachings and unwavering support throughout the years.

To Maurizio, Paolo, Isabella, Federico, Riccardo, Valentina. You are special people and this would not have happened without you.

To the friends who have not given up on me, even though I have been the worst version of myself. Every little act of love from you has been so important.

A Roma, 'pe' quella panchina rotta ch'ha dato i sogni a 'sta pischella'. The first real chance for me came from you, and you will always be home.

BIBLIOGRAPHY

1. International Agency for Research on Cancer. World Health Organization. Global Cancer Observatory. Available from: https://gco.iarc.fr/
2. Morgan E, Arnold M, Gini A, Lorenzoni V, Cabasag CJ, Laversanne M, et al. Global burden of colorectal cancer in 2020 and 2040: incidence and mortality estimates from GLOBOCAN. Gut. 2023;72(2):338–44. Available from: https://pubmed.ncbi.nlm.nih.gov/36604116/
3. Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. Cancer Epidemiol Biomarkers Prev. 2009 Jun;18(6):1688–94. Available from: https://pubmed.ncbi.nlm.nih.gov/19505900/
4. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011 Mar;61(2):69–90. Available from: https://pubmed.ncbi.nlm.nih.gov/21296855/
5. Klabunde CN, Cronin KA, Breen N, Waldron WR, Ambs AH, Nadel MR. Trends in colorectal cancer test use among vulnerable populations in the United States. Cancer Epidemiol Biomarkers Prev. 2011 Aug;20(8):1611–21. Available from: https://pubmed.ncbi.nlm.nih.gov/21653643/
6. Doubeni CA, Major JM, Laiyemo AO, Schootman M, Zauber AG, Hollenbeck AR, et al. Contribution of behavioral risk factors and obesity to socioeconomic differences in colorectal cancer incidence. J Natl Cancer Inst. 2012 Sep 19;104(18):1353–62. Available from: https://pubmed.ncbi.nlm.nih.gov/22952311/
7. Doubeni CA, Laiyemo AO, Major JM, Schootman M, Lian M, Park Y, et al. Socioeconomic status and the risk of colorectal cancer: an analysis of more than a half million adults in the National Institutes of Health-AARP Diet and Health Study. Cancer. 2012 Jul 15;118(14):3636–44. Available from: https://pubmed.ncbi.nlm.nih.gov/22898918/
8. Street W. Colorectal Cancer Facts & Figures 2020–2022. Am Cancer Soc Atlanta, GA, USA. 2020;48.
9. Gornick D, Kadakuntla A, Trovato A, Stetzer R, Tadros M. Practical considerations for colorectal cancer screening in older adults. World J Gastrointest Oncol. 2022 Jun 6;14(6):1086. Available from: /pmc/articles/PMC9244986/
10. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017 Apr 1;66(4):683–91. Available from: https://pubmed.ncbi.nlm.nih.gov/26818619/
11. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Przegląd Gastroenterol. 2019;14(2):89. Available from: /pmc/articles/PMC6791134/

12. Dahmus JD, Kotler DL, Kastenberg DM, Kistler CA. The gut microbiome and colorectal cancer: a review of bacterial pathogenesis. J Gastrointest Oncol. 2018 Aug 1;9(4):769–77. Available from: https://jgo.amegroups.org/article/view/21032/html
13. Rebersek M. Gut microbiome and its role in colorectal cancer. BMC Cancer. 2021 Dec 1;21(1).
14. Sobhani I, Tap J, Roudot-Thoraval F, Roperch JP, Letulle S, Langella P, et al. Microbial Dysbiosis in Colorectal Cancer (CRC) Patients. PLoS One. 2011;6(1):e16393. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0016393
15. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M, et al. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe. 2013 Aug 14;14(2):207–15. Available from: https://pubmed.ncbi.nlm.nih.gov/23954159/
16. Routy B, Lenehan JG, Miller Jr WH, Jamal R, Messaoudene M, Daisley BA, et al. Fecal microbiota transplantation plus anti-PD-1 immunotherapy in advanced melanoma: a phase I trial. Nat Med. 2023;1–12.
17. Kim J, Lee HK. Potential Role of the Gut Microbiome In Colorectal Cancer Progression. Front Immunol. 2022 Jan 7;12:807648.
18. Centers for Disease Control and Prevention (CDC). What Are the Risk Factors for Colorectal Cancer?.. Available from: https://www.cdc.gov/cancer/colorectal/basic_info/risk_factors.htm
19. American Cancer Society. Colorectal Cancer Risk Factors .. Available from: https://www.cancer.org/cancer/types/colon-rectal-cancer/causes-risks-prevention/risk-factors.html
20. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Gastroenterol Rev Gastroenterol. 2019;14(2):89–103. Available from: https://doi.org/10.5114/pg.2018.81072
21. Willauer AN, Liu Y, Pereira AAL, Lam M, Morris JS, Raghav KPS, et al. Clinical and molecular characterization of early-onset colorectal cancer. Cancer. 2019 Jun 15;125(12):2002–10. Available from: https://pubmed.ncbi.nlm.nih.gov/30854646/
22. Dozois EJ, Boardman LA, Suwanthanma W, Limburg PJ, Cima RR, Bakken JL, et al. Young-onset colorectal cancer in patients with no known genetic predisposition: can we increase early recognition and improve outcome? Medicine (Baltimore). 2008 Sep;87(5):259–63. Available from: https://pubmed.ncbi.nlm.nih.gov/18794708/
23. Meester RGS, Mannalithara A, Lansdorp-Vogelaar I, Ladabaum U. Trends in Incidence and Stage at Diagnosis of Colorectal Cancer in Adults Aged 40 Through 49 Years, 1975-2015. JAMA. 2019 May 21;321(19):1933–4. Available from:

https://pubmed.ncbi.nlm.nih.gov/31112249/

24. Murphy N, Campbell PT, Gunter MJ. Unraveling the Etiology of Early-Onset Colorectal Cancer. JNCI J Natl Cancer Inst. 2021 May 4;113(5):505–6. Available from: https://dx.doi.org/10.1093/jnci/djaa165
25. Bajpai M, Seril DN, Van Gurp J, Geng X, Alvarez J, Minacapelli CD, et al. Effect of Long-Term Mesalamine Therapy on Cancer-Associated Gene Expression in Colonic Mucosa of Patients with Ulcerative Colitis. Dig Dis Sci. 2019 Mar 15;64(3):740–50. Available from: https://pubmed.ncbi.nlm.nih.gov/30478770/
26. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. Gut. 2001;48(4):526–35. Available from: https://pubmed.ncbi.nlm.nih.gov/11247898/
27. Jess T, Rungoe C, Peyrin-Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clin Gastroenterol Hepatol. 2012 Jun;10(6):639–45. Available from: https://pubmed.ncbi.nlm.nih.gov/22289873/
28. Magro F, Gionchetti P, Eliakim R, Ardizzone S, Armuzzi A, Barreiro-de Acosta M, et al. Third European Evidence-based Consensus on Diagnosis and Management of Ulcerative Colitis. Part 1: Definitions, Diagnosis, Extra-intestinal Manifestations, Pregnancy, Cancer Surveillance, Surgery, and Ileo-anal Pouch Disorders. J Crohns Colitis. 2017 Jun 1;11(6):649–70. Available from: https://pubmed.ncbi.nlm.nih.gov/28158501/
29. Maaser C, Sturm A, Vavricka SR, Kucharzik T, Fiorino G, Annese V, et al. ECCO-ESGAR Guideline for Diagnostic Assessment in IBD Part 1: Initial diagnosis, monitoring of known IBD, detection of complications. J Crohns Colitis. 2019 Feb 1;13(2):144–64. Available from: https://pubmed.ncbi.nlm.nih.gov/30137275/
30. Keller DS, Windsor A, Cohen R, Chand M. Colorectal cancer in inflammatory bowel disease: review of the evidence. Tech Coloproctol. 2019 Jan 28;23(1):3–13. Available from: https://pubmed.ncbi.nlm.nih.gov/30701345/
31. Gandomani H, Gandomani HS, yousefi SM, Aghajani M, Mohammadian-Hafshejani A, Tarazoj AA, et al. Colorectal cancer in the world: incidence, mortality and risk factors. Biomed Res Ther. 2017 Oct 14;4(10):1656–75. Available from: http://bmrat.org/index.php/BMRAT/article/view/372
32. Amersi F, Agustin M, Ko CY. Colorectal Cancer: Epidemiology, Risk Factors, and Health Services. Clin Colon Rectal Surg. 2005 Aug;18(3):133. Available from: /pmc/articles/PMC2780097/
33. Munkholm P. Review article: the incidence and prevalence of colorectal cancer in inflammatory bowel disease. Aliment Pharmacol Ther. 2003;18 Suppl 2(2):1–5. Available from: https://pubmed.ncbi.nlm.nih.gov/12950413/

34. Olén O, Erichsen R, Sachs MC, Pedersen L, Halfvarson J, Askling J, et al. Colorectal cancer in ulcerative colitis: a Scandinavian population-based cohort study. Lancet. 2020 Jan 11;395(10218):123–31.
35. Olén O, Erichsen R, Sachs MC, Pedersen L, Halfvarson J, Askling J, et al. Colorectal cancer in Crohn's disease: a Scandinavian population-based cohort study. lancet Gastroenterol Hepatol. 2020 May 1;5(5):475–84. Available from: https://pubmed.ncbi.nlm.nih.gov/32066530/
36. Kofla-Dłubacz A, Pytrus T, Akutko K, Sputa-Grzegrzółka P, Piotrowska A, Dzięgiel P. Etiology of IBD—Is It Still a Mystery? Int J Mol Sci. 2022 Oct 1;23(20). Available from: /pmc/articles/PMC9604112/
37. Lucafò M, Curci D, Franzin M, Decorti G, Stocco G. Inflammatory Bowel Disease and Risk of Colorectal Cancer: An Overview From Pathophysiology to Pharmacological Prevention. Front Pharmacol. 2021 Oct 20;12:772101.
38. Lucafò M, Curci D, Franzin M, Decorti G, Stocco G. Inflammatory Bowel Disease and Risk of Colorectal Cancer: An Overview From Pathophysiology to Pharmacological Prevention. Front Pharmacol. 2021 Oct 20;12. Available from: /pmc/articles/PMC8563785/
39. Kang M, Martin A. Microbiome and colorectal cancer: Unraveling host-microbiota interactions in colitis-associated colorectal cancer development. Semin Immunol. 2017 Aug 1;32:3–13. Available from: https://pubmed.ncbi.nlm.nih.gov/28465070/
40. Fan X, Jin Y, Chen G, Ma X, Zhang L. Gut Microbiota Dysbiosis Drives the Development of Colorectal Cancer. Digestion. 2021 Jun 1;102(4):508–15. Available from: https://pubmed.ncbi.nlm.nih.gov/32932258/
41. Ren J, Kirkness CS, Kim M, Asche C V., Puli S. Long-term risk of colorectal cancer by gender after positive colonoscopy: population-based cohort study. Curr Med Res Opin. 2016 Aug 2;32(8):1367–74. Available from: https://pubmed.ncbi.nlm.nih.gov/27050237/
42. Tuohy TMF, Rowe KG, Mineau GP, Pimentel R, Burt RW, Samadder NJ. Risk of colorectal cancer and adenomas in the families of patients with adenomas: a population-based study in Utah. Cancer. 2014 Jan 1;120(1):35–42. Available from: https://pubmed.ncbi.nlm.nih.gov/24150925/
43. Lowery JT, Ahnen DJ, Schroy PC, Hampel H, Baxter N, Boland CR, et al. Understanding the contribution of family history to colorectal cancer risk and its clinical implications: A state-of-the-science review. Cancer. 2016 Sep 1;122(17):2633–45. Available from: https://pubmed.ncbi.nlm.nih.gov/27258162/
44. Samadder NJ, Smith KR, Hanson H, Pimentel R, Wong J, Boucher K, et al. Increased Risk of Colorectal Cancer Among Family Members of All Ages, Regardless of Age of Index Case at Diagnosis. Clin Gastroenterol Hepatol. 2015 Dec 1;13(13):2305-2311.e2. Available from:

https://pubmed.ncbi.nlm.nih.gov/26188136/

45. Yurgelun MB, Kulke MH, Fuchs CS, Allen BA, Uno H, Hornick JL, et al. Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer. J Clin Oncol. 2017 Apr 1;35(10):1086–95. Available from: https://pubmed.ncbi.nlm.nih.gov/28135145/
46. Win AK, Lindor NM, Young JP, MacRae FA, Young GP, Williamson E, et al. Risks of Primary Extracolonic Cancers Following Colorectal Cancer in Lynch Syndrome. JNCI J Natl Cancer Inst. 2012 Sep 9;104(18):1363. Available from: /pmc/articles/PMC3529597/
47. Parry S, Win AK, Parry B, Macrae FA, Gurrin LC, Church JM, et al. Metachronous colorectal cancer risk for mismatch repair gene mutation carriers: the advantage of more extensive colon surgery. Gut. 2011 Jul;60(7):950–7. Available from: https://pubmed.ncbi.nlm.nih.gov/21193451/
48. Kohlmann W, Gruber SB. Lynch Syndrome. GeneReviews(®). 2021 Feb 4; Available from: https://www.ncbi.nlm.nih.gov/books/NBK1211/
49. Lazzeroni M, Bellerba F, Calvello M, Macrae F, Win AK, Jenkins M, et al. A meta-analysis of obesity and risk of colorectal cancer in patients with lynch syndrome: The impact of sex and genetics. Nutrients. 2021 May 1;13(5):1736. Available from: https://www.mdpi.com/2072-6643/13/5/1736/htm
50. Leoz ML, Carballal S, Moreira L, Ocaña T, Balaguer F. The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. Appl Clin Genet. 2015 Apr 16;8:95–107. Available from: https://pubmed.ncbi.nlm.nih.gov/25931827/
51. Yen T, Stanich PP, Axell L, Patel SG. APC-Associated Polyposis Conditions. Atlas Genet Cytogenet Oncol Haematol. 2022 May 12;24(12):477–82. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1345/
52. Vasen HFA, Tomlinson I, Castells A. Clinical management of hereditary colorectal cancer syndromes. Nat Rev Gastroenterol Hepatol. 2015 Jan 1;12(2):88–97. Available from: https://pubmed.ncbi.nlm.nih.gov/25582351/
53. Pati S, Irfan W, Jameel A, Ahmed S, Shahid RK. Obesity and Cancer: A Current Overview of Epidemiology, Pathogenesis, Outcomes, and Management. Cancers (Basel). 2023 Jan 1;15(2). Available from: /pmc/articles/PMC9857053/
54. Aleksandrova K, Nimptsch K, Pischon T. Obesity and colorectal cancer. Front Biosci (Elite Ed). 2013 Jan 1;5(1):61–77. Available from: https://pubmed.ncbi.nlm.nih.gov/23276970/
55. Sinicrope FA, Foster NR, Yothers G, Benson A, Seitz JF, Labianca R, et al. Body mass index at diagnosis and survival among colon cancer patients enrolled in clinical trials of adjuvant chemotherapy. Cancer. 2013 Apr 15;119(8):1528–36. Available from: https://pubmed.ncbi.nlm.nih.gov/23310947/

56. Lee J, Meyerhardt JA, Giovannucci E, Jeon JY. Association between body mass index and prognosis of colorectal cancer: a meta-analysis of prospective cohort studies. PLoS One. 2015 Mar 26;10(3). Available from: https://pubmed.ncbi.nlm.nih.gov/25811460/
57. Ortega LS, Bradbury KE, Cross AJ, Morris JS, Gunter MJ, Murphy N. A Prospective Investigation of Body Size, Body Fat Composition and Colorectal Cancer Risk in the UK Biobank. Sci Rep. 2017 Dec 1;7(1):17807. Available from: /pmc/articles/PMC5736687/
58. Larsson SC, Wolk A. Obesity and colon and rectal cancer risk: a meta-analysis of prospective studies. Am J Clin Nutr. 2007 Sep 1;86(3):556–65. Available from: https://pubmed.ncbi.nlm.nih.gov/17823417/
59. Aleksandrova K, Schlesinger S, Fedirko V, Jenab M, Bueno-De-Mesquita B, Freisling H, et al. Metabolic Mediators of the Association Between Adult Weight Gain and Colorectal Cancer: Data From the European Prospective Investigation into Cancer and Nutrition (EPIC) Cohort. Am J Epidemiol. 2017 May 5;185(9):751. Available from: /pmc/articles/PMC5860400/
60. Murphy N, Jenab M, Gunter MJ. Adiposity and gastrointestinal cancers: epidemiology, mechanisms and future directions. Nat Rev Gastroenterol Hepatol. 2018 Nov 1;15(11):659–70. Available from: https://pubmed.ncbi.nlm.nih.gov/29970888/
61. Gribovskaja-Rupp I, Kosinski L, Ludwig KA. Obesity and Colorectal Cancer. Clin Colon Rectal Surg. 2011;24(4):229. Available from: /pmc/articles/PMC3311490/
62. Iyengar NM, Gucalp A, Dannenberg AJ, Hudis CA. Obesity and Cancer Mechanisms: Tumor Microenvironment and Inflammation. J Clin Oncol. 2016 Dec 10;34(35):4270–6. Available from: https://pubmed.ncbi.nlm.nih.gov/27903155/
63. Kim H, Giovannucci EL. Sex differences in the association of obesity and colorectal cancer risk. Cancer Causes Control. 2017 Jan 1;28(1):1–4. Available from: https://pubmed.ncbi.nlm.nih.gov/27878394/
64. Renehan AG, Zwahlen M, Egger M. Adiposity and cancer risk: new mechanistic insights from epidemiology. Nat Rev Cancer. 2015 Aug 27;15(8):484–98. Available from: https://pubmed.ncbi.nlm.nih.gov/26205341/
65. Liu XZ, Pedersen L, Halberg N. Cellular mechanisms linking cancers to obesity. Cell Stress. 2021;5(5):55. Available from: /pmc/articles/PMC8090860/
66. Schmidt S, Monk JM, Robinson LE, Mourtzakis M. The integrative role of leptin, oestrogen and the insulin family in obesity-associated breast cancer: potential effects of exercise. Obes Rev. 2015 Jun 1;16(6):473–87. Available from: https://pubmed.ncbi.nlm.nih.gov/25875578/
67. Pham DV, Park PH. Tumor Metabolic Reprogramming by Adipokines as a Critical Driver of Obesity-Associated Cancer Progression. Int J Mol Sci 2021, Vol 22, Page 1444. 2021 Feb 1;22(3):1444. Available from: https://www.mdpi.com/1422-0067/22/3/1444/htm

68. Kelesidis I, Kelesidis T, Mantzoros CS. Adiponectin and cancer: a systematic review. Br J Cancer. 2006 May 5;94(9):1221. Available from: /pmc/articles/PMC2361397/
69. Ellulu MS, Patimah I, Khaza'ai H, Rahmat A, Abed Y. Obesity and inflammation: the linking mechanism and the complications. Arch Med Sci. 2017;13(4):851. Available from: /pmc/articles/PMC5507106/
70. Hopkins BD, Goncalves MD, Cantley LC. Obesity and Cancer Mechanisms: Cancer Metabolism. J Clin Oncol. 2016 Dec 10;34(35):4277–83. Available from: https://pubmed.ncbi.nlm.nih.gov/27903152/
71. Mctiernan A, Friedenreich CM, Katzmarzyk PT, Powell KE, Macko R, Buchner D, et al. Physical Activity in Cancer Prevention and Survival: A Systematic Review. Med Sci Sports Exerc. 2019 Jun 1;51(6):1252–61. Available from: https://pubmed.ncbi.nlm.nih.gov/31095082/
72. Matthews CE, Moore SC, Arem H, Cook MB, Trabert B, Hakansson N, et al. Amount and Intensity of Leisure-Time Physical Activity and Lower Cancer Risk. J Clin Oncol. 2020 Mar 1;38(7):686–98. Available from: https://pubmed.ncbi.nlm.nih.gov/31877085/
73. Mazzilli KM, Matthews CE, Salerno EA, Moore SC. Weight Training and Risk of 10 Common Types of Cancer. Med Sci Sports Exerc. 2019 Sep 1;51(9):1845–51. Available from: https://pubmed.ncbi.nlm.nih.gov/30920488/
74. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. BMJ. 2016;354. Available from: https://pubmed.ncbi.nlm.nih.gov/27510511/
75. Boyle T, Keegel T, Bull F, Heyworth J, Fritschi L. Physical activity and risks of proximal and distal colon cancers: a systematic review and meta-analysis. J Natl Cancer Inst. 2012 Oct 17;104(20):1548–61. Available from: https://pubmed.ncbi.nlm.nih.gov/22914790/
76. Wolin KY, Yan Y, Colditz GA, Lee IM. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer. 2009 Feb 2;100(4):611. Available from: /pmc/articles/PMC2653744/
77. Mctiernan A, Friedenreich CM, Katzmarzyk PT, Powell KE, Macko R, Buchner D, et al. Physical Activity in Cancer Prevention and Survival: A Systematic Review. Med Sci Sports Exerc. 2019 Jun 1;51(6):1252–61. Available from: https://journals.lww.com/acsm-msse/fulltext/2019/06000/physical_activity_in_cancer_prevention_and.20.aspx
78. Larsson SC, Rutegård J, Bergkvist L, Wolk A. Physical activity, obesity, and risk of colon and rectal cancer in a cohort of Swedish men. Eur J Cancer. 2006 Oct;42(15):2590–7.
79. Slattery ML, Boucher KM, Caan BJ, Potter JD, Ma KN. Eating patterns and risk of colon cancer. Am J Epidemiol. 1998 Jul 1;148(1):4–16. Available from:

https://pubmed.ncbi.nlm.nih.gov/9663397/

80. Bradbury KE, Appleby PN, Key TJ. Fruit, vegetable, and fiber intake in relation to cancer risk: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC). Am J Clin Nutr. 2014 Jul 1;100(SUPPL. 1):394S-398S.
81. Kim YI, Mason JB. Nutrition chemoprevention of gastrointestinal cancers: a critical review. Nutr Rev. 1996;54(9):259–79. Available from: https://pubmed.ncbi.nlm.nih.gov/9009668/
82. Wong SH, Yu J. Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev Gastroenterol Hepatol. 2019 Nov 1;16(11):690–704. Available from: https://pubmed.ncbi.nlm.nih.gov/31554963/
83. Kwong TNY, Wang X, Nakatsu G, Chow TC, Tipoe T, Dai RZW, et al. Association Between Bacteremia From Specific Microbes and Subsequent Diagnosis of Colorectal Cancer. Gastroenterology. 2018 Aug 1;155(2):383-390.e8. Available from: https://pubmed.ncbi.nlm.nih.gov/29729257/
84. Tabung FK, Liu L, Wang W, Fung TT, Wu K, Smith-Warner SA, et al. Association of Dietary Inflammatory Potential With Colorectal Cancer Risk in Men and Women. JAMA Oncol. 2018 Mar 1;4(3):366–73. Available from: https://pubmed.ncbi.nlm.nih.gov/29346484/
85. Bouvard V, Loomis D, Guyton KZ, Grosse Y, Ghissassi F El, Benbrahim-Tallaa L, et al. Carcinogenicity of consumption of red and processed meat. Lancet Oncol. 2015 Dec 1;16(16):1599–600. Available from: https://pubmed.ncbi.nlm.nih.gov/26514947/
86. Vieira AR, Abar L, Chan DSM, Vingeliene S, Polemiti E, Stevens C, et al. Foods and beverages and colorectal cancer risk: a systematic review and meta-analysis of cohort studies, an update of the evidence of the WCRF-AICR Continuous Update Project. Ann Oncol Off J Eur Soc Med Oncol. 2017 Aug 1;28(8):1788–802. Available from: https://pubmed.ncbi.nlm.nih.gov/28407090/
87. Kim E, Coelho D, Blachier F. Review of the association between meat consumption and risk of colorectal cancer. Nutr Res. 2013 Dec;33(12):983–94. Available from: https://pubmed.ncbi.nlm.nih.gov/24267037/
88. Michels KB, Giovannucci E, Joshipura KJ, Rosner BA, Stampfer MJ, Fuchs CS, et al. Prospective study of fruit and vegetable consumption and incidence of colon and rectal cancers. J Natl Cancer Inst. 2000 Nov 1;92(21):1740–52. Available from: https://pubmed.ncbi.nlm.nih.gov/11058617/
89. Van Duijnhoven FJB, Bueno-De-Mesquita HB, Ferrari P, Jenab M, Boshuizen HC, Ros MM, et al. Fruit, vegetables, and colorectal cancer risk: the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr. 2009 May 1;89(5):1441–52.
90. Lee JE, Chan AT. Fruit, Vegetables, and Folate: Cultivating the Evidence for Cancer Prevention. Gastroenterology. 2011;141(1):16. Available from:

/pmc/articles/PMC3391696/

91. Godos J, Bella F, Sciacca S, Galvano F, Grosso G. Vegetarianism and breast, colorectal and prostate cancer risk: an overview and meta-analysis of cohort studies. J Hum Nutr Diet. 2017 Jun 1;30(3):349–59. Available from: https://pubmed.ncbi.nlm.nih.gov/27709695/

92. Orlich MJ, Singh PN, Sabaté J, Fan J, Sveen L, Bennett H, et al. Vegetarian dietary patterns and the risk of colorectal cancers. JAMA Intern Med. 2015 May 1;175(5):767–76. Available from: https://pubmed.ncbi.nlm.nih.gov/25751512/

93. Schwingshackl L, Schwedhelm C, Hoffmann G, Knüppel S, Laure Preterre A, Iqbal K, et al. Food groups and risk of colorectal cancer. Int J cancer. 2018 May 1;142(9):1748–58. Available from: https://pubmed.ncbi.nlm.nih.gov/29210053/

94. Aune D, Chan DSM, Lau R, Vieira R, Greenwood DC, Kampman E, et al. Dietary fibre, whole grains, and risk of colorectal cancer: systematic review and dose-response meta-analysis of prospective studies. BMJ. 2011 Nov 26;343(7833):1082. Available from: https://pubmed.ncbi.nlm.nih.gov/22074852/

95. Kyrø C, Olsen A, Landberg R, Skeie G, Loft S, Åman P, et al. Plasma alkylresorcinols, biomarkers of whole-grain wheat and rye intake, and incidence of colorectal cancer. J Natl Cancer Inst. 2014 Jan 1;106(1). Available from: https://pubmed.ncbi.nlm.nih.gov/24317181/

96. Kunzmann AT, Coleman HG, Huang WY, Kitahara CM, Cantwell MM, Berndt SI. Dietary fiber intake and risk of colorectal cancer and incident and recurrent adenoma in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. Am J Clin Nutr. 2015 Oct 1;102(4):881–90. Available from: https://pubmed.ncbi.nlm.nih.gov/26269366/

97. Dahm CC, Keogh RH, Spencer EA, Greenwood DC, Key TJ, Fentiman IS, et al. Dietary fiber and colorectal cancer risk: a nested case-control study using food diaries. J Natl Cancer Inst. 2010 May;102(9):614–26. Available from: https://pubmed.ncbi.nlm.nih.gov/20407088/

98. Larsson SC, Giovannucci E, Bergkvist L, Wolk A. Whole grain consumption and risk of colorectal cancer: a population-based cohort of 60 000 women. Br J Cancer. 2005 May 5;92(9):1803. Available from: /pmc/articles/PMC2362029/

99. Bingham SA, Day NE, Luben R, Ferrari P, Slimani N, Norat T, et al. Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): An observational study. Lancet. 2003 May 3;361(9368):1496–501. Available from: https://pubmed.ncbi.nlm.nih.gov/12737858/

100. Peters U, Sinha R, Chatterjee N, Subar AF, Ziegler RG, Kulldorff M, et al. Dietary fibre and colorectal adenoma in a colorectal cancer early detection programme. Lancet (London, England). 2003 May 3;361(9368):1491–5. Available from:

https://pubmed.ncbi.nlm.nih.gov/12737857/

101. Asano TK, McLeod RS. Dietary fibre for the prevention of colorectal adenomas and carcinomas. Cochrane database Syst Rev. 2002 Jan 21;(2). Available from: https://pubmed.ncbi.nlm.nih.gov/12076480/

102. Rthur A, Chatzkin S, Laine E, Anza L, Onald D, Orle C, et al. Lack of Effect of a Low-Fat, High-Fiber Diet on the Recurrence of Colorectal Adenomas. https://doi.org/101056/NEJM200004203421601. 2000 Apr 20;342(16):1149–55. Available from: https://www.nejm.org/doi/full/10.1056/nejm200004203421601

103. Maclennan R, Macrae F, Bain C, Battistutta D, Chapuis P, Gratten H, et al. Randomized trial of intake of fat, fiber, and beta carotene to prevent colorectal adenomas. J Natl Cancer Inst. 1995 Dec 6;87(23):1760–6. Available from: https://pubmed.ncbi.nlm.nih.gov/7473832/

104. Park Y, Hunter DJ, Spiegelman D, Bergkvist L, Berrino F, Van Den Brandt PA, et al. Dietary fiber intake and risk of colorectal cancer: a pooled analysis of prospective cohort studies. JAMA. 2005 Dec 14;294(22):2849–57. Available from: https://pubmed.ncbi.nlm.nih.gov/16352792/

105. He X, Wu K, Zhang X, Nishihara R, Cao Y, Fuchs CS, et al. Dietary intake of fiber, whole grains and risk of colorectal cancer: an updated analysis according to food sources, tumor location and molecular subtypes in two large US cohorts. Int J cancer. 2019 Dec 12;145(11):3040. Available from: /pmc/articles/PMC7274214/

106. Ferguson LR, Harris PJ. The dietary fibre debate: More food for thought. Lancet. 2003 May 3;361(9368):1487–8. Available from: http://www.thelancet.com/article/S0140673603132199/fulltext

107. Dahm CC, Keogh RH, Spencer EA, Greenwood DC, Key TJ, Fentiman IS, et al. Dietary Fiber and Colorectal Cancer Risk: A Nested Case–Control Study Using Food Diaries. JNCI J Natl Cancer Inst. 2010 May 5;102(9):614–26. Available from: https://dx.doi.org/10.1093/jnci/djq092

108. Jones G. Pharmacokinetics of vitamin D toxicity. Am J Clin Nutr. 2008 Aug 1;88(2). Available from: https://pubmed.ncbi.nlm.nih.gov/18689406/

109. Deeb KK, Trump DL, Johnson CS. Vitamin D signalling pathways in cancer: potential for anticancer therapeutics. Nat Rev Cancer. 2007 Sep;7(9):684–700. Available from: https://pubmed.ncbi.nlm.nih.gov/17721433/

110. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nat 2008 4547203. 2008 Jul 1;454(7203):436–44. Available from: https://www.nature.com/articles/nature07205

111. Adams JS, Hewison M. Update in Vitamin D. J Clin Endocrinol Metab. 2010 Feb

1;95(2):471–8. Available from: https://dx.doi.org/10.1210/jc.2009-1773

112. International Agency for Research on Cancer. Vitamin D and Cancer. International Agency for Research on Cancer; 2008. 449 p.

113. Wong G, Lim WH, Lewis J, Craig JC, Turner R, Zhu K, et al. Vitamin D and cancer mortality in elderly women. BMC Cancer. 2015 Mar 8;15(1):1–9. Available from: https://bmccancer.biomedcentral.com/articles/10.1186/s12885-015-1112-5

114. Sha S, Nguyen TMN, Kuznia S, Niedermaier T, Zhu A, Brenner H, et al. Real-world evidence for the effectiveness of vitamin D supplementation in reduction of total and cause-specific mortality. J Intern Med. 2023 Mar 1;293(3):384–97. Available from: https://pubmed.ncbi.nlm.nih.gov/36208176/

115. Schöttker B, Haug U, Schomburg L, Köhrle J, Perna L, Müller H, et al. Strong associations of 25-hydroxyvitamin D concentrations with all-cause, cardiovascular, cancer, and respiratory disease mortality in a large cohort study. Am J Clin Nutr. 2013 Apr 1;97(4):782–93. Available from: https://pubmed.ncbi.nlm.nih.gov/23446902/

116. Lee DM, Vanderschueren D, Boonen S, O'Neill TW, Pendleton N, Pye SR, et al. Association of 25-hydroxyvitamin D, 1,25-dihydroxyvitamin D and parathyroid hormone with mortality among middle-aged and older European men. Age Ageing. 2014 Jul 1;43(4):528–35. Available from: https://pubmed.ncbi.nlm.nih.gov/24375224/

117. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, et al. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. Int J cancer. 2011 Mar 15;128(6):1414–24. Available from: https://pubmed.ncbi.nlm.nih.gov/20473927/

118. Autier P, Gandini S. Vitamin D supplementation and total mortality: a meta-analysis of randomized controlled trials. Arch Intern Med. 2007 Oct 9;167(16):1730–7. Available from: https://pubmed.ncbi.nlm.nih.gov/17846391/

119. Kuznia S, Zhu A, Akutsu T, Buring JE, Camargo CA, Cook NR, et al. Efficacy of vitamin D3 supplementation on cancer mortality: Systematic review and individual patient data meta-analysis of randomised controlled trials. Ageing Res Rev. 2023 Jun 1;87. Available from: https://pubmed.ncbi.nlm.nih.gov/37004841/

120. Keum N, Lee DH, Greenwood DC, Manson JE, Giovannucci E. Vitamin D supplementation and total cancer incidence and mortality: a meta-analysis of randomized controlled trials. Ann Oncol Off J Eur Soc Med Oncol. 2019 May 1;30(5):733–43. Available from: https://pubmed.ncbi.nlm.nih.gov/30796437/

121. Bjelakovic G, Gluud LL, Nikolova D, Whitfield K, Wetterslev J, Simonetti RG, et al. Vitamin D supplementation for prevention of mortality in adults. Cochrane database Syst Rev. 2014 Jan 10;2014(1). Available from: https://pubmed.ncbi.nlm.nih.gov/24414552/

122. Autier P, Mullie P, Macacu A, Dragomir M, Boniol M, Coppens K, et al. Effect of vitamin D supplementation on non-skeletal disorders: a systematic review of meta-analyses and randomised trials. lancet Diabetes Endocrinol. 2017 Dec 1;5(12):986–1004. Available from: https://pubmed.ncbi.nlm.nih.gov/29102433/
123. Sha S, Chen LJ, Brenner H, Schöttker B. Associations of 25-hydroxyvitamin D status and vitamin D supplementation use with mortality due to 18 frequent cancer types in the UK Biobank cohort. Eur J Cancer. 2023 Sep 1;191:113241.
124. Manson JE, Cook NR, Lee I-M, Christen W, Bassuk SS, Mora S, et al. Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease. N Engl J Med. 2019 Jan 3;380(1):33–44. Available from: https://pubmed.ncbi.nlm.nih.gov/30415629/
125. Wactawski-Wende J, Kotchen JM, Anderson GL, Assaf AR, Brunner RL, O'Sullivan MJ, et al. Calcium plus vitamin D supplementation and the risk of colorectal cancer. N Engl J Med. 2006 Feb 16;354(7):684–96. Available from: https://pubmed.ncbi.nlm.nih.gov/16481636/
126. Baron JA, Barry EL, Mott LA, Rees JR, Sandler RS, Snover DC, et al. A Trial of Calcium and Vitamin D for the Prevention of Colorectal Adenomas. N Engl J Med. 2015 Oct 10;373(16):1519. Available from: /pmc/articles/PMC4643064/
127. Cao Y, Chan AT. Vitamin D and Early-Onset Colorectal Cancer—Rays of Hope? Gastroenterology. 2023 Oct 1;165(4):831–3. Available from: http://www.gastrojournal.org/article/S0016508523048424/fulltext
128. Cui A, Xiao P, Ma Y, Fan Z, Zhou F, Zheng J, et al. Prevalence, trend, and predictor analyses of vitamin D deficiency in the US population, 2001-2018. Front Nutr. 2022 Oct 3;9. Available from: https://pubmed.ncbi.nlm.nih.gov/36263304/
129. Kim H, Lipsyc-Sharf M, Zong X, Wang X, Hur J, Song M, et al. Total Vitamin D Intake and Risks of Early-Onset Colorectal Cancer and Precursors. Gastroenterology. 2021 Oct 1;161(4):1208-1217.e9. Available from: https://pubmed.ncbi.nlm.nih.gov/34245763/
130. Kim Y, Chang Y, Cho Y, Chang J, Kim K, Park D Il, et al. Serum 25-Hydroxyvitamin D Levels and Risk of Colorectal Cancer: An Age-Stratified Analysis. Gastroenterology. 2023 Oct 1;165(4):920–31. Available from: https://pubmed.ncbi.nlm.nih.gov/37429364/
131. Tobias DK, Luttmann-Gibson H, Mora S, Danik J, Bubes V, Copeland T, et al. Association of Body Weight With Response to Vitamin D Supplementation and Metabolism. JAMA Netw Open. 2023 Jan 3;6(1):e2250681. Available from: /pmc/articles/PMC9856931/
132. Manson JE, Cook NR, Lee I-M, Christen W, Bassuk SS, Mora S, et al. Vitamin D Supplements and Prevention of Cancer and Cardiovascular Disease. N Engl J Med. 2019 Jan 3;380(1):33–44. Available from: https://www.nejm.org/doi/full/10.1056/nejmoa1809944
133. Chandler PD, Chen WY, Ajala ON, Hazra A, Cook N, Bubes V, et al. Effect of Vitamin D3 Supplements on Development of Advanced Cancer: A Secondary Analysis of the VITAL

Randomized Clinical Trial. JAMA Netw open. 2020 Nov 18;3(11). Available from: https://pubmed.ncbi.nlm.nih.gov/33206192/

134. Secretan B, Straif K, Baan R, Grosse Y, Ghissassi F El, Bouvard V, et al. A review of human carcinogens--Part E: tobacco, areca nut, alcohol, coal smoke, and salted fish. Lancet Oncol. 2009 Nov 1;10(11):1033–4. Available from: https://europepmc.org/article/med/19891056

135. Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA. 2008 Dec 17;300(23):2765–78. Available from: https://pubmed.ncbi.nlm.nih.gov/19088354/

136. Botteri E, Iodice S, Raimondi S, Maisonneuve P, Lowenfels AB. Cigarette smoking and adenomatous polyps: a meta-analysis. Gastroenterology. 2008;134(2). Available from: https://pubmed.ncbi.nlm.nih.gov/18242207/

137. Buchanan DD, Sweet K, Drini M, Jenkins MA, Win AK, English DR, et al. Risk Factors for Colorectal Cancer in Patients with Multiple Serrated Polyps: A Cross-Sectional Case Series from Genetics Clinics. PLoS One. 2010 Jul 16;5(7):e11636. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0011636

138. Wallace K, Grau M V., Ahnen D, Snover DC, Robertson DJ, Mahnke D, et al. The association of lifestyle and dietary factors with the risk of serrated polyps of the colorectum. Cancer Epidemiol Biomarkers Prev. 2009 Aug;18(8):2310. Available from: /pmc/articles/PMC3669681/

139. Ordóñez-Mena JM, Walter V, Schöttker B, Jenab M, O'Doherty MG, Kee F, et al. Impact of prediagnostic smoking and smoking cessation on colorectal cancer prognosis: a meta-analysis of individual patient data from cohorts within the CHANCES consortium. Ann Oncol Off J Eur Soc Med Oncol. 2018 Feb 1;29(2):472–83. Available from: https://pubmed.ncbi.nlm.nih.gov/29244072/

140. Yang B, Jacobs EJ, Gapstur SM, Stevens V, Campbell PT. Active smoking and mortality among colorectal cancer survivors: the Cancer Prevention Study II nutrition cohort. J Clin Oncol. 2015 Mar 10;33(8):885–93. Available from: https://pubmed.ncbi.nlm.nih.gov/25646196/

141. Caini S, Del Riccio M, Vettori V, Raimondi S, Assedi M, Vignati S, et al. The Prognostic Impact of Quitting Smoking at or around Diagnosis on the Survival of Patients with Gastrointestinal Cancers: A Systematic Literature Review. Cancers (Basel). 2022 Aug 1;14(16). Available from: https://pubmed.ncbi.nlm.nih.gov/36010851/

142. Mizoue T, Inoue M, Wakai K, Nagata C, Shimazu T, Tsuji I, et al. Alcohol drinking and colorectal cancer in Japanese: a pooled analysis of results from five cohort studies. Am J Epidemiol. 2008;167(12):1397–406. Available from: https://pubmed.ncbi.nlm.nih.gov/18420544/

143. Cho E, Smith-Warner SA, Ritz J, Van Den Brandt PA, Colditz GA, Folsom AR, et al. Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies. Ann Intern Med. 2004 Apr 20;140(8). Available from: https://pubmed.ncbi.nlm.nih.gov/15096331/
144. Wang Y, Duan H, Yang H, Lin J. A pooled analysis of alcohol intake and colorectal cancer. Int J Clin Exp Med. 2015 May 30;8(5):6878. Available from: /pmc/articles/PMC4509170/
145. Carding S, Verbeke K, Vipond DT, Corfe BM, Owen LJ. Dysbiosis of the gut microbiota in disease. Microb Ecol Health Dis. 2015;26(1):26191.
146. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, Zolfo M, et al. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. 2019 Apr 1;25(4):667–78. Available from: https://pubmed.ncbi.nlm.nih.gov/30936548/
147. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, Milanese A, et al. Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med. 2019 Apr 1;25(4):679–89. Available from: https://www.nature.com/articles/s41591-019-0406-6
148. Wong CC, Yu J. Gut microbiota in colorectal cancer development and therapy. Nat Rev Clin Oncol. 2023;1–24.
149. Yamamoto S, Kinugasa H, Hirai M, Terasawa H, Yasutomi E, Oka S, et al. Heterogeneous distribution of Fusobacterium nucleatum in the progression of colorectal cancer. J Gastroenterol Hepatol. 2021;36(7):1869–76.
150. Okuda S, Shimada Y, Tajima Y, Yuza K, Hirose Y, Ichikawa H, et al. Profiling of host genetic alterations and intra-tumor microbiomes in colorectal cancer. Comput Struct Biotechnol J. 2021;19:3330–8.
151. Liu W, Zhang X, Xu H, Li S, Lau HC-H, Chen Q, et al. Microbial community heterogeneity within colorectal neoplasia and its correlation with colorectal carcinogenesis. Gastroenterology. 2021;160(7):2395–408.
152. Bullman S, Pedamallu CS, Sicinska E, Clancy TE, Zhang X, Cai D, et al. Analysis of Fusobacterium persistence and antibiotic response in colorectal cancer. Science (80-). 2017;358(6369):1443–8.
153. Li L, Li X, Zhong W, Yang M, Xu M, Sun Y, et al. Gut microbiota from colorectal cancer patients enhances the progression of intestinal adenoma in Apc (min/+) mice. EBioMedicine 48: 301–315. 2019.
154. Wong SH, Zhao L, Zhang X, Nakatsu G, Han J, Xu W, et al. Gavage of fecal samples from patients with colorectal cancer promotes intestinal carcinogenesis in germ-free and conventional mice. Gastroenterology. 2017;153(6):1621–33.
155. Chung L, Orberg ET, Geis AL, Chan JL, Fu K, Shields CED, et al. Bacteroides fragilis toxin

coordinates a pro-carcinogenic inflammatory cascade via targeting of colonic epithelial cells. Cell Host Microbe. 2018;23(2):203–14.

156. Akyol A, Hinoi T, Feng Y, Bommer GT, Glaser TM, Fearon ER. Generating somatic mosaicism with a Cre recombinase–microsatellite sequence transgene. Nat Methods. 2008;5(3):231–3.

157. Yang J, Wei H, Zhou Y, Szeto C-H, Li C, Lin Y, et al. High-fat diet promotes colorectal tumorigenesis through modulating gut microbiota and metabolites. Gastroenterology. 2022;162(1):135–49.

158. Coker OO, Liu C, Wu WKK, Wong SH, Jia W, Sung JJY, et al. Altered gut metabolites and microbiota interactions are implicated in colorectal carcinogenesis and can be non-invasive diagnostic biomarkers. Microbiome. 2022;10(1):35.

159. Gao R, Wu C, Zhu Y, Kong C, Zhu Y, Gao Y, et al. Integrated analysis of colorectal cancer reveals cross-cohort gut microbial signatures and associated serum metabolites. Gastroenterology. 2022;163(4):1024–37.

160. Kong C, Liang L, Liu G, Du L, Yang Y, Liu J, et al. Integrated metagenomic and metabolomic analysis reveals distinct gut-microbiome-derived phenotypes in early-onset colorectal cancer. Gut. 2023;72(6):1129–42.

161. Chen F, Dai X, Zhou C-C, Li K, Zhang Y, Lou X-Y, et al. Integrated analysis of the faecal metagenome and serum metabolome reveals the role of gut microbiome-associated metabolites in the detection of colorectal cancer and adenoma. Gut. 2022;71(7):1315–25.

162. Tsilidis KK, Branchini C, Guallar E, Helzlsouer KJ, Erlinger TP, Platz EA. C-reactive protein and colorectal cancer risk: A systematic review of prospective studies. Vol. 123, International Journal of Cancer. Int J Cancer; 2008. p. 1133–40. Available from: https://pubmed.ncbi.nlm.nih.gov/18528865/

163. Pohl C, Hombach A, Kruis W. Chronic inflammatory bowel disease and cancer. Vol. 47, Hepato-Gastroenterology. 2000. p. 57–70. Available from: https://europepmc.org/article/med/10690586

164. Grivennikov SI, Karin M. Inflammatory cytokines in cancer: Tumour necrosis factor and interleukin 6 take the stage. In: Annals of the Rheumatic Diseases. Ann Rheum Dis; 2011. Available from: https://pubmed.ncbi.nlm.nih.gov/21339211/

165. Doyle SL, Donohoe CL, Lysaght J, Reynolds J V. Visceral obesity, metabolic syndrome, insulin resistance and cancer. In: Proceedings of the Nutrition Society. Proc Nutr Soc; 2012. p. 181–9. Available from: https://pubmed.ncbi.nlm.nih.gov/22051112/

166. Holick MF, Chen TC. Vitamin D deficiency: A worldwide problem with health consequences. Am J Clin Nutr. 2008 Apr 1;87(4). Available from: https://pubmed.ncbi.nlm.nih.gov/18400738/

167. Gandini S, Boniol M, Haukka J, Byrnes G, Cox B, Sneyd MJ, et al. Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. Int J Cancer. 2011 Mar 15;128(6):1414–24. Available from: http://doi.wiley.com/10.1002/ijc.25439

168. Keum N, Chen QY, Lee DH, Manson JE, Giovannucci E. Vitamin D supplementation and total cancer incidence and mortality by daily vs. infrequent large-bolus dosing strategies: a meta-analysis of randomised controlled trials. Br J Cancer. 2022;127(5). Available from: https://pubmed.ncbi.nlm.nih.gov/35676320/

169. Liu PT, Stenger S, Li H, Wenzel L, Tan BH, Krutzik SR, et al. Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response. Science (80-). 2006 Mar 24;311(5768):1770–3. Available from: https://pubmed.ncbi.nlm.nih.gov/16497887/

170. Adorini L, Penna G, Giarratana N, Roncari A, Amuchastegui S, Daniel KC, et al. Dendritic cells as key targets for immunomodulation by Vitamin D receptor ligands. In: Journal of Steroid Biochemistry and Molecular Biology. J Steroid Biochem Mol Biol; 2004. p. 437–41. Available from: https://pubmed.ncbi.nlm.nih.gov/15225816/

171. Bashir M, Prietl B, Tauschmann M, Mautner SI, Kump PK, Treiber G, et al. Effects of high doses of vitamin D3 on mucosa-associated gut microbiome vary between regions of the human gastrointestinal tract. Eur J Nutr. 2016 Jun 1;55(4):1479–89. Available from: http://www.mothur.org/

172. Serrano D, Pozzi C, Guglietta S, Fosso B, Suppa M, Gnagnarella P, et al. Microbiome as Mediator of Diet on Colorectal Cancer Risk: The Role of Vitamin D, Markers of Inflammation and Adipokines. Nutrients. 2021 Feb 1;13(2):1–19. Available from: https://pubmed.ncbi.nlm.nih.gov/33504116/

173. Manzari C, Fosso B, Marzano M, Annese A, Caprioli R, D'Erchia AM, et al. The influence of invasive jellyfish blooms on the aquatic microbiome in a coastal lagoon (Varano, SE Italy) detected by an Illumina-based deep sequencing strategy. Biol Invasions. 2015 Mar 1;17(3):923–40. Available from: https://link.springer.com/article/10.1007/s10530-014-0810-2

174. Manzari C, Chiara M, Costanza A, Leoni C, Volpicella M, Picardi E, et al. Draft genome sequence of Sphingobium sp. strain ba1, resistant to kanamycin and nickel ions. FEMS Microbiol Lett. 2014 Dec 1;361(1):8–9. Available from: https://academic.oup.com/femsle/article/361/1/8/2908306

175. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017 Sep 12;35(9):833–44. Available from: https://pubmed.ncbi.nlm.nih.gov/28898207/

176. International WCRF. Meat, fish and dairy products and the risk of cancer. Contin Updat

Proj Expert Rep. 2018;

177. Fund WCR. Diet, Nutrition, Physical Activity and Cancer: a Global Perspective. Contin Updat Proj Expert Report. 2018;2018.

178. Fosso B, Santamaria M, Marzano M, Alonso-Alemany D, Valiente G, Donvito G, et al. BioMaS: A modular pipeline for Bioinformatic analysis of Metagenomic AmpliconS. BMC Bioinformatics. 2015 Dec 12;16(1):1–11. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-015-0595-z

179. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014 Mar;30(5):614–20. Available from: https://pubmed.ncbi.nlm.nih.gov/24142950/

180. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, et al. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 2009;37:D141. Available from: /pmc/articles/PMC2686447/

181. Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Res. 2014 Jan 1;42(Database issue). Available from: https://pubmed.ncbi.nlm.nih.gov/24288368/

182. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Apr;9(4):357–9. Available from: https://pubmed.ncbi.nlm.nih.gov/22388286/

183. Mende DR, Sunagawa S, Zeller G, Bork P. Accurate and universal delineation of prokaryotic species. Nat Methods. 2013 Sep;10(9):881–4. Available from: https://pubmed.ncbi.nlm.nih.gov/23892899/

184. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):1–21. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8

185. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015 Sep 29;12(10):902–3. Available from: https://pubmed.ncbi.nlm.nih.gov/26418763/

186. Franzosa EA, McIver LJ, Rahnavard G, Thompson LR, Schirmer M, Weingart G, et al. Species-level functional profiling of metagenomes and metatranscriptomes. Nat Methods. 2018 Nov 1;15(11):962–8. Available from: https://pubmed.ncbi.nlm.nih.gov/30377376/

187. Chatfield SM, Brand C, Ebeling PR, Russell DM. Vitamin D deficiency in general medical inpatients in summer and winter. Intern Med J. 2007 Jun;37(6):377–82. Available from: https://pubmed.ncbi.nlm.nih.gov/17535381/

188. Docio S, Riancho JA, Pérez A, Olmos JM, Amado JA, González-Macías J. Seasonal deficiency of vitamin D in children: a potential target for osteoporosis-preventing strategies? J Bone Miner Res. 1998 Apr;13(4):544–8. Available from:

https://pubmed.ncbi.nlm.nih.gov/9556054/

189. Gallagher JC, Sai AJ. Vitamin D insufficiency, deficiency, and bone health. J Clin Endocrinol Metab. 2010;95(6):2630–3. Available from: https://pubmed.ncbi.nlm.nih.gov/20525913/

190. Ding C, Parameswaran V, Udayan R, Burgess J, Jones G. Circulating levels of inflammatory markers predict change in bone mineral density and resorption in older adults: a longitudinal study. J Clin Endocrinol Metab. 2008;93(5):1952–8. Available from: https://pubmed.ncbi.nlm.nih.gov/18285417/

191. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, et al. Metagenomic biomarker discovery and explanation. Genome Biol. 2011 Jun 24;12(6). Available from: https://pubmed.ncbi.nlm.nih.gov/21702898/

192. Ter Braak CJF. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology. 1986;67(5):1167–79.

193. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 2019 Sep 1;35(17):3055–62. Available from: https://pubmed.ncbi.nlm.nih.gov/30657866/

194. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Comput Biol. 2017 Nov 1;13(11):e1005752. Available from: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005752

195. VanderWeele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol. 2010 Dec 15;172(12):1339–48. Available from: https://pubmed.ncbi.nlm.nih.gov/21036955/

196. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013 Jun;18(2):137–50. Available from: https://pubmed.ncbi.nlm.nih.gov/23379553/

197. Liu D, Jiang XY, Zhou LS, Song JH, Zhang X. Effects of Probiotics on Intestinal Mucosa Barrier in Patients With Colorectal Cancer after Operation: Meta-Analysis of Randomized Controlled Trials. Medicine (Baltimore). 2016 Apr 1;95(15). Available from: https://pubmed.ncbi.nlm.nih.gov/27082589/

198. Zou S, Fang L, Lee MH. Dysbiosis of gut microbiota in promoting the development of colorectal cancer. Gastroenterol Rep. 2018 Feb 1;6(1):1–12. Available from: https://pubmed.ncbi.nlm.nih.gov/29479437/

199. Ley RE, Turnbaugh PJ, Klein S, Gordon JI. Microbial ecology: human gut microbes associated with obesity. Nature. 2006 Dec 21;444(7122):1022–3. Available from:

https://pubmed.ncbi.nlm.nih.gov/17183309/

200. Ferrocino I, Di Cagno R, De Angelis M, Turroni S, Vannini L, Bancalari E, et al. Fecal Microbiota in Healthy Subjects Following Omnivore, Vegetarian and Vegan Diets: Culturable Populations and rRNA DGGE Profiling. PLoS One. 2015 Jun 2;10(6). Available from: /pmc/articles/PMC4452701/

201. Watanabe Y, Nagai F, Morotomi M. Characterization of Phascolarctobacterium succinatutens sp. nov., an Asaccharolytic, Succinate-Utilizing Bacterium Isolated from Human Feces. Appl Environ Microbiol. 2012 Jan;78(2):511. Available from: /pmc/articles/PMC3255759/

202. Stojanov S, Berlec A, Štrukelj B. The Influence of Probiotics on the Firmicutes/Bacteroidetes Ratio in the Treatment of Obesity and Inflammatory Bowel disease. Microorg 2020, Vol 8, Page 1715. 2020 Nov 1;8(11):1715. Available from: https://www.mdpi.com/2076-2607/8/11/1715/htm

203. Bikle DD. Vitamin D metabolism, mechanism of action, and clinical applications. Chem Biol. 2014 Mar 20;21(3):319–29. Available from: https://pubmed.ncbi.nlm.nih.gov/24529992/

204. Christakos S, Dhawan P, Verstuyf A, Verlinden L, Carmeliet G. Vitamin D: Metabolism, Molecular Mechanism of Action, and Pleiotropic Effects. Physiol Rev. 2016 Dec 16;96(1):365–408. Available from: https://pubmed.ncbi.nlm.nih.gov/26681795/

205. Heath AK, Kim IY, Hodge AM, English DR, Muller DC. Vitamin D Status and Mortality: A Systematic Review of Observational Studies. Int J Environ Res Public Health. 2019 Feb 1;16(3). Available from: https://pubmed.ncbi.nlm.nih.gov/30700025/

206. Gandini S, Raimondi S, Gnagnarella P, Doré JF, Maisonneuve P, Testori A. Vitamin D and skin cancer: a meta-analysis. Eur J Cancer. 2009 Mar;45(4):634–41. Available from: https://pubmed.ncbi.nlm.nih.gov/19008093/

207. Parker J, Hashmi O, Dutton D, Mavrodaris A, Stranges S, Kandala NB, et al. Levels of vitamin D and cardiometabolic disorders: systematic review and meta-analysis. Maturitas. 2010 Mar;65(3):225–36. Available from: https://pubmed.ncbi.nlm.nih.gov/20031348/

208. Charoenngam N, Holick MF. Immunologic Effects of Vitamin D on Human Health and Disease. Nutrients. 2020 Jul 1;12(7):1–28. Available from: /pmc/articles/PMC7400911/

209. Greenstein RJ, Su L, Brown ST. Vitamins A & D Inhibit the Growth of Mycobacteria in Radiometric Culture. PLoS One. 2012 Jan 3;7(1):e29631. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0029631

210. Fakhoury HMA, Kvietys PR, AlKattan W, Anouti F Al, Elahi MA, Karras SN, et al. Vitamin D and intestinal homeostasis: Barrier, microbiota, and immune modulation. J Steroid Biochem Mol Biol. 2020 Jun 1;200. Available from: https://pubmed.ncbi.nlm.nih.gov/32194242/

211. Rinninella E, Raoul P, Cintoni M, Franceschi F, Miggiano GAD, Gasbarrini A, et al. What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. Microorganisms. 2019 Jan 1;7(1):14. Available from: /pmc/articles/PMC6351938/

212. Bellerba F, Muzio V, Gnagnarella P, Facciotti F, Chiocca S, Bossi P, et al. The Association between Vitamin D and Gut Microbiota: A Systematic Review of Human Studies. Nutr 2021, Vol 13, Page 3378. 2021 Sep 26;13(10):3378. Available from: https://www.mdpi.com/2072-6643/13/10/3378/htm

213. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston R V., Thomas J. Defining the criteria for including studies and how they will be grouped for the synthesis. Cochrane Handb Syst Rev Interv. 2019 Jan 1;33–65. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/9781119536604.ch3

214. Sordillo JE, Zhou Y, McGeachie MJ, Ziniti J, Lange N, Laranjo N, et al. Factors influencing the infant gut microbiome at age 3-6 months: Findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). J Allergy Clin Immunol. 2017 Feb 1;139(2):482-491.e14. Available from: https://pubmed.ncbi.nlm.nih.gov/27746239/

215. Naderpoor N, Mousa A, Arango LFG, Barrett HL, Nitert MD, de Courten B. Effect of Vitamin D Supplementation on Faecal Microbiota: A Randomised Clinical Trial. Nutrients. 2019 Dec 1;11(12). Available from: https://pubmed.ncbi.nlm.nih.gov/31783602/

216. Missailidis C, Sørensen N, Ashenafi S, Amogne W, Kassa E, Bekele A, et al. Vitamin D and Phenylbutyrate Supplementation Does Not Modulate Gut Derived Immune Activation in HIV-1. Nutrients. 2019 Jul 1;11(7). Available from: https://pubmed.ncbi.nlm.nih.gov/31330899/

217. Kanhere M, He J, Chassaing B, Ziegler TR, Alvarez JA, Ivie EA, et al. Bolus Weekly Vitamin D3 Supplementation Impacts Gut and Airway Microbiota in Adults With Cystic Fibrosis: A Double-Blind, Randomized, Placebo-Controlled Clinical Trial. J Clin Endocrinol Metab. 2018 Feb 1;103(2):564–74. Available from: https://pubmed.ncbi.nlm.nih.gov/29161417/

218. Hjelmsø MH, Shah SA, Thorsen J, Rasmussen M, Vestergaard G, Mortensen MS, et al. Prenatal dietary supplements influence the infant airway microbiota in a randomized factorial clinical trial. Nat Commun. 2020 Dec 1;11(1). Available from: https://pubmed.ncbi.nlm.nih.gov/31969566/

219. Charoenngam N, Shirvani A, Kalajian TA, Song A, Holick MF. The Effect of Various Doses of Oral Vitamin D3 Supplementation on Gut Microbiota in Healthy Adults: A Randomized, Double-blinded, Dose-response Study. Anticancer Res. 2020;40(1):551–6. Available from: https://pubmed.ncbi.nlm.nih.gov/31892611/

220. Ciubotaru I, Green SJ, Kukreja S, Barengolts E. Significant differences in fecal microbiota

are associated with various stages of glucose tolerance in African American male veterans. Transl Res. 2015 Nov 1;166(5):401–11. Available from: https://pubmed.ncbi.nlm.nih.gov/26209747/

221. Talsness CE, Penders J, Jansen EHJM, Damoiseaux J, Thijs C, Mommers M. Influence of vitamin D on key bacterial taxa in infant microbiota in the KOALA Birth Cohort Study. PLoS One. 2017 Nov 1;12(11):e0188011. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188011
222. Mandal S, Godfrey KM, McDonald D, Treuren W V., Bjørnholt J V., Midtvedt T, et al. Fat and vitamin intakes during pregnancy have stronger relations with a pro-inflammatory maternal microbiota than does carbohydrate intake. Microbiome. 2016;4. Available from: /pmc/articles/PMC5070355/
223. Kassem Z, Sitarik A, Levin AM, Lynch S V., Havstad S, Fujimura K, et al. Maternal and cord blood vitamin D level and the infant gut microbiota in a birth cohort study. Matern Heal Neonatol Perinatol. 2020 Dec;6(1). Available from: https://pubmed.ncbi.nlm.nih.gov/33101701/
224. Drall KM, Field CJ, Haqq AM, de Souza RJ, Tun HM, Morales-Lizcano NP, et al. Vitamin D supplementation in pregnancy and early infancy in relation to gut microbiota composition and C. difficile colonization: implications for viral respiratory infections. Gut Microbes. 2020 Nov 9;12(1). Available from: https://pubmed.ncbi.nlm.nih.gov/32779963/
225. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, et al. Linking long-term dietary patterns with gut microbial enterotypes. Science. 2011 Oct 7;334(6052):105–8. Available from: https://pubmed.ncbi.nlm.nih.gov/21885731/
226. Thomas RL, Jiang L, Adams JS, Xu ZZ, Shen J, Janssen S, et al. Vitamin D metabolites and the gut microbiome in older men. Nat Commun 2020 111. 2020 Nov 26;11(1):1–10. Available from: https://www.nature.com/articles/s41467-020-19793-8
227. Seura T, Yoshino Y, Fukuwatari T. The Relationship between Habitual Dietary Intake and Gut Microbiota in Young Japanese Women. J Nutr Sci Vitaminol (Tokyo). 2017;63(6):396–404. Available from: https://pubmed.ncbi.nlm.nih.gov/29332901/
228. Luthold R V., Fernandes GR, Franco-de-Moraes AC, Folchetti LGD, Ferreira SRG. Gut microbiota interactions with the immunomodulatory role of vitamin D in normal individuals. Metabolism. 2017 Apr 1;69:76–86. Available from: https://pubmed.ncbi.nlm.nih.gov/28285654/
229. Jackson MA, Verdi S, Maxan ME, Shin CM, Zierer J, Bowyer RCE, et al. Gut microbiota associations with common diseases and prescription medications in a population-based cohort. Nat Commun. 2018 Dec 1;9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29985401/

230. Soltys K, Stuchlikova M, Hlavaty T, Gaalova B, Budis J, Gazdarica J, et al. Seasonal changes of circulating 25-hydroxyvitamin D correlate with the lower gut microbiome composition in inflammatory bowel disease patients. Sci Rep. 2020 Dec 1;10(1). Available from: https://pubmed.ncbi.nlm.nih.gov/32265456/

231. Weng YJ, Gan HY, Li X, Huang Y, Li ZC, Deng HM, et al. Correlation of diet, microbiota and metabolite networks in inflammatory bowel disease. J Dig Dis. 2019 Sep 1;20(9):447–59. Available from: https://pubmed.ncbi.nlm.nih.gov/31240835/

232. Mandal S, Godfrey KM, McDonald D, Treuren W V., Bjørnholt J V., Midtvedt T, et al. Fat and vitamin intakes during pregnancy have stronger relations with a proinflammatory maternal microbiota than does carbohydrate intake. Microbiome. 2016 Oct 19;4(1):1–11. Available from: https://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-016-0200-3

233. Kassem Z, Sitarik A, Levin AM, Lynch S V., Havstad S, Fujimura K, et al. Maternal and cord blood vitamin D level and the infant gut microbiota in a birth cohort study. Matern Heal Neonatol Perinatol 2020 61. 2020 Oct 20;6(1):1–10. Available from: https://mhnpjournal.biomedcentral.com/articles/10.1186/s40748-020-00119-x

234. Drall KM, Field CJ, Haqq AM, de Souza RJ, Tun HM, Morales-Lizcano NP, et al. Vitamin D supplementation in pregnancy and early infancy in relation to gut microbiota composition and C. difficile colonization: implications for viral respiratory infections. Gut Microbes. 2020 Nov 9;12(1). Available from: /pmc/articles/PMC7524344/

235. Garg M, Hendy P, Ding JN, Shaw S, Hold G, Hart A. The Effect of Vitamin D on Intestinal Inflammation and Faecal Microbiota in Patients with Ulcerative Colitis. J Crohns Colitis. 2018 Jul 30;12(8):963–72. Available from: https://pubmed.ncbi.nlm.nih.gov/29726893/

236. Schäffler H, Herlemann DPR, Klinitzke P, Berlin P, Kreikemeyer B, Jaster R, et al. Vitamin D administration leads to a shift of the intestinal bacterial composition in Crohn's disease patients, but not in healthy controls. J Dig Dis. 2018 Apr 1;19(4):225–34. Available from: https://pubmed.ncbi.nlm.nih.gov/29573237/

237. Cantarel BL, Waubant E, Chehoud C, Kuczynski J, Desantis TZ, Warrington J, et al. Gut microbiota in multiple sclerosis: possible influence of immunomodulators. J Investig Med. 2015 Jun 3;63(5):729–34. Available from: https://pubmed.ncbi.nlm.nih.gov/25775034/

238. Singh P, Rawat A, Alwakeel M, Sharif E, Al Khodor S. The potential role of vitamin D supplementation as a gut microbiota modifier in healthy individuals. Sci Reports 2020 101. 2020 Dec 10;10(1):1–14. Available from: https://www.nature.com/articles/s41598-020-77806-4

239. Bashir M, Prietl B, Tauschmann M, Mautner SI, Kump PK, Treiber G, et al. Effects of high doses of vitamin D3 on mucosa-associated gut microbiome vary between regions of the

human gastrointestinal tract. Eur J Nutr. 2016 Jun 1;55(4):1479–89. Available from: https://pubmed.ncbi.nlm.nih.gov/26130323/

240. Bosman ES, Albert AY, Lui H, Dutz JP, Vallance BA. Skin exposure to narrow band ultraviolet (Uvb) light modulates the human intestinal microbiome. Front Microbiol. 2019;10(OCT). Available from: /pmc/articles/PMC6821880/

241. Tabatabaeizadeh SA, Fazeli M, Meshkat Z, Khodashenas E, Esmaeili H, Mazloum S, et al. The effects of high doses of vitamin D on the composition of the gut microbiome of adolescent girls. Clin Nutr ESPEN. 2020 Feb 1;35:103–8. Available from: https://pubmed.ncbi.nlm.nih.gov/31987101/

242. Furet JP, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S, et al. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol. 2009 Jun;68(3):351–62. Available from: https://pubmed.ncbi.nlm.nih.gov/19302550/

243. RefSeq: NCBI Reference Sequence Database.. Available from: https://www.ncbi.nlm.nih.gov/refseq/

244. Gabriel Al-Ghalith DK. BURST enables optimal exhaustive DNA alignment for big data.

245. McIver LJ, Abu-Ali G, Franzosa EA, Schwager R, Morgan XC, Waldron L, et al. bioBakery: a meta'omic analysis environment. Bioinformatics. 2018 Apr 1;34(7):1235–7. Available from: https://pubmed.ncbi.nlm.nih.gov/29194469/

246. Beghini F, McIver LJ, Blanco-Míguez A, Dubois L, Asnicar F, Maharjan S, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. Elife. 2021 May 1;10.

247. Quinn TP, Erb I, Richardson MF, Crowley TM. Understanding sequencing data as compositions: an outlook and review. Bioinformatics. 2018 Aug 8;34(16):2870. Available from: /pmc/articles/PMC6084572/

248. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome Datasets Are Compositional: And This Is Not Optional. Front Microbiol. 2017 Nov 15;8(NOV). Available from: https://pubmed.ncbi.nlm.nih.gov/29187837/

249. Nearing JT, Douglas GM, Hayes MG, MacDonald J, Desai DK, Allward N, et al. Microbiome differential abundance methods produce different results across 38 datasets. Nat Commun 2022 131. 2022 Jan 17;13(1):1–16. Available from: https://www.nature.com/articles/s41467-022-28034-z

250. Calgaro M, Romualdi C, Waldron L, Risso D, Vitulo N. Assessment of statistical methods from single cell, bulk RNA-seq, and metagenomics applied to microbiome data. Genome Biol. 2020 Aug 3;21(1):1–31. Available from: https://genomebiology.biomedcentral.com/articles/10.1186/s13059-020-02104-1

251. Pawlowsky-Glahn V, Egozcue JJ, Lovell D. Tools for compositional data with a total. http://dx.doi.org/101177/1471082X14535526. 2014 Nov 25;15(2):175–90. Available from: https://journals.sagepub.com/doi/abs/10.1177/1471082X14535526
252. Greenacre M. Compositional Data Analysis. https://doi.org/101146/annurev-statistics-042720-124436. 2021 Mar 8;8:271–99. Available from: https://www.annualreviews.org/doi/abs/10.1146/annurev-statistics-042720-124436
253. Aitchison J. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B. 1982;44(2):139–77.
254. Aitchison J. The Statistical Analysis of Compositional Data. J R Stat Soc Ser B. 1982 Jan 1;44(2):139–60. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.2517-6161.1982.tb01195.x
255. Pearson K. Mathematical Contributions to the Theory of Evolution.--On a Form of Spurious Correlation Which May Arise When Indices Are Used in the Measurement of Organs. RSPS. 1896;60:489–98. Available from: https://ui.adsabs.harvard.edu/abs/1896RSPS...60..489P/abstract
256. Greenacre M, Grunsky E, Bacon-Shone J. A comparison of isometric and amalgamation logratio balances in compositional data analysis. Comput Geosci. 2021 Mar 1;148:104621.
257. Aitchison J (John). The statistical analysis of compositional data. 1986;416. Available from: https://www.worldcat.org/title/13215689
258. Aitchison J, Barceló-Vidal C, Martín-Fernández JA, Pawlowsky-Glahn V. Logratio analysis and compositional distance. Math Geol. 2000;32(3):271–5.
259. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric Logratio Transformations for Compositional Data Analysis. Math Geol. 2003 Apr;35(3):279–300. Available from: https://link.springer.com/article/10.1023/A:1023818214614
260. Quinn TP, Erb I, Richardson MF, Crowley TM. Understanding sequencing data as compositions: an outlook and review. Bioinformatics. 2018 Aug 15;34(16):2870–8. Available from: https://dx.doi.org/10.1093/bioinformatics/bty175
261. Lubbe S, Filzmoser P, Templ M. Comparison of zero replacement strategies for compositional data with large numbers of zeros. Chemom Intell Lab Syst. 2021 Mar 15;210:104248.
262. Filzmoser P, Hron K, Templ M. Applied Compositional Data Analysis. 2018; Available from: http://link.springer.com/10.1007/978-3-319-96422-5
263. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. Dealing With Zeros and Missing Values in Compositional Data Sets Using Nonparametric Imputation 1. Math Geol. 2003;35(3).
264. Palarea-Albaladejo J, Martín-Fernández JA. zCompositions — R package for multivariate

imputation of left-censored data under a compositional approach. Chemom Intell Lab Syst. 2015 Apr 15;143:85–96.

265. Susin A, Wang Y, Cao KAL, Luz Calle M. Variable selection in microbiome compositional data analysis. NAR Genomics Bioinforma. 2020 Jun 1;2(2). Available from: https://dx.doi.org/10.1093/nargab/lqaa029

266. Luz Calle M. Statistical Analysis of Metagenomics Data. Genomics Inform. 2019;17(1). Available from: /pmc/articles/PMC6459172/

267. Cao KAL, Costello ME, Lakis VA, Bartolo F, Chua XY, Brazeilles R, et al. MixMC: A Multivariate Statistical Framework to Gain Insight into Microbial Communities. PLoS One. 2016 Aug 1;11(8):e0160169. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0160169

268. Valeri L, VanderWeele TJ. Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. Psychol Methods. 2013 Jun;18(2):137–50. Available from: https://pubmed.ncbi.nlm.nih.gov/23379553/

269. Vanderweele TJ, Vansteelandt S. Conceptual issues concerning mediation, interventions and composition. Stat Interface. 2009;2(4):457–68.

270. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010 Jun 15;26(12):1572–3. Available from: https://dx.doi.org/10.1093/bioinformatics/btq170

271. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995 Jan 1;57(1):289–300. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/j.2517-6161.1995.tb02031.x

272. Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008 Jul;9(3):432–41. Available from: https://pubmed.ncbi.nlm.nih.gov/18079126/

273. Bellerba F, Serrano D, Johansson H, Pozzi C, Segata N, NabiNejad A, et al. Colorectal cancer, Vitamin D and microbiota: A double-blind Phase II randomized trial (ColoViD) in colorectal cancer patients. Neoplasia. 2022 Oct;34:100842.

274. Wang H, Luo K, Guan Z, Li Z, Xiang J, Ou S, et al. Identification of the Crucial Role of CCL22 in F. nucleatum-Related Colorectal Tumorigenesis that Correlates With Tumor Microenvironment and Immune Checkpoint Therapy. Front Genet. 2022 Feb 28;13:1. Available from: /pmc/articles/PMC8918684/

275. Morishita A, Oura K, Tadokoro T, Shi T, Fujita K, Tani J, et al. Galectin-9 in Gastroenterological Cancer. Int J Mol Sci. 2023 Apr 1;24(7). Available from: /pmc/articles/PMC10094448/

276. Wang Y, Sun J, Ma C, Gao W, Song B, Xue H, et al. Reduced Expression of Galectin-9 Contributes to a Poor Outcome in Colon Cancer by Inhibiting NK Cell Chemotaxis Partially through the Rho/ROCK1 Signaling Pathway. PLoS One. 2016 Mar 1;11(3). Available from: /pmc/articles/PMC4814049/

277. Zhou X, Sun L, Jing D, Xu G, Zhang J, Lin L, et al. Galectin-9 Expression Predicts Favorable Clinical Outcome in Solid Tumors: A Systematic Review and Meta-Analysis. Front Physiol. 2018 Apr 26;9(APR). Available from: https://pubmed.ncbi.nlm.nih.gov/29765332/

278. Zafar H, Saier MH. Gut Bacteroides species in health and disease. Gut Microbes. 2021;13(1):1–20. Available from: https://pubmed.ncbi.nlm.nih.gov/33535896/

279. Zagato E, Pozzi C, Bertocchi A, Schioppa T, Saccheri F, Guglietta S, et al. Endogenous murine microbiota member Faecalibaculumrodentium and its human homolog protect from intestinal tumorgrowth. Nat Microbiol. 2020 Mar 1;5(3):511. Available from: /pmc/articles/PMC7048616/

280. Leylabadlo HE, Ghotaslou R, Feizabadi MM, Farajnia S, Moaddab SY, Ganbarov K, et al. The critical role of Faecalibacterium prausnitzii in human health: An overview. Microb Pathog. 2020 Dec 1;149. Available from: https://pubmed.ncbi.nlm.nih.gov/32534182/

281. Venegas DP, De La Fuente MK, Landskron G, González MJ, Quera R, Dijkstra G, et al. Short chain fatty acids (SCFAs)mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. Front Immunol. 2019 Mar 11;10(MAR):424615.

282. Dikeocha IJ, Al-Kabsi AM, Chiu HT, Alshawsh MA. Faecalibacterium prausnitzii Ameliorates Colorectal Tumorigenesis and Suppresses Proliferation of HCT116 Colorectal Cancer Cells. Biomed 2022, Vol 10, Page 1128. 2022 May 13;10(5):1128. Available from: https://www.mdpi.com/2227-9059/10/5/1128/htm

283. Martini LA, Wood RJ. Vitamin D status and the metabolic syndrome. Nutr Rev. 2006 Nov;64(11):479–86. Available from: https://pubmed.ncbi.nlm.nih.gov/17131943/

284. Wortsman J, Matsuoka LY, Chen TC, Lu Z, Holick MF. Decreased bioavailability of vitamin D in obesity. Am J Clin Nutr. 2000;72(3):690–3. Available from: https://pubmed.ncbi.nlm.nih.gov/10966885/

285. Valeri F, Endres K. How biological sex of the host shapes its gut microbiota. Front Neuroendocrinol. 2021 Apr 1;61:100912.

286. Yoon K, Kim N. Roles of Sex Hormones and Gender in the Gut Microbiota. J Neurogastroenterol Motil. 2021 Jul 1;27(3):314–25. Available from: https://pubmed.ncbi.nlm.nih.gov/33762473/

287. Kim YS, Unno T, Kim BY, Park MS. Sex Differences in Gut Microbiota. World J Mens Health. 2020 Jan 1;38(1):48–60. Available from: https://doi.org/10.5534/wjmh.190009

288. Morris A, Ta ME. Microbiota drives sex-specific differences. Nat Rev Endocrinol 2018 151.

2018 Nov 13;15(1):4–4. Available from: https://www.nature.com/articles/s41574-018-0127-9

289. Ma Z, Li Ma WZ, Li W, Ma Z. How and Why Men and Women Differ in Their Microbiomes: Medical Ecology and Network Analyses of the Microgenderome. Adv Sci. 2019 Dec 1;6(23):1902054. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/advs.201902054

290. Chary S, Amrein K, Mahmoud SH, Lasky-Su JA, Christopher KB. Sex-Specific Catabolic Metabolism Alterations in the Critically Ill following High Dose Vitamin D. 2022;12(3):207. Available from: https://www.mdpi.com/2218-1989/12/3/207/htm

291. Nirmagustina DE, Yang Y, Kumrungsee T, Yanaka N, Kato N. Gender Difference and Dietary Supplemental Vitamin B 6: Impact on Colon Luminal Environment. J Nutr Sci Vitaminol (Tokyo). 2018;64(2):116–28. Available from: https://pubmed.ncbi.nlm.nih.gov/29710029/

292. Bashir A, Miskeen AY, Hazari YM, Asrafuzzaman S, Fazili KM. Fusobacterium nucleatum, inflammation, and immunity: the fire within human gut. Tumour Biol. 2016 Mar 1;37(3):2805–10. Available from: https://pubmed.ncbi.nlm.nih.gov/26718210/

293. Sun C-H, Li B-B, Wang B, Zhao J, Zhang X-Y, Li T-T, et al. The role of Fusobacterium nucleatum in colorectal cancer: from carcinogenesis to clinical management. Chronic Dis Transl Med. 2019 Sep;5(3):178. Available from: /pmc/articles/PMC6926109/

294. Liu W, Zhang L, Xu H-J, Li Y, Hu C-M, Yang J-Y, et al. The Anti-Inflammatory Effects of Vitamin D in Tumorigenesis. Int J Mol Sci. 2018 Sep 13;19(9):2736. Available from: http://www.mdpi.com/1422-0067/19/9/2736

295. Morishita A, Nomura K, Tani J, Fujita K, Iwama H, Takuma K, et al. Galectin-9 suppresses the tumor growth of colon cancer in vitro and in vivo. Oncol Rep. 2021 Jun 1;45(6). Available from: /pmc/articles/PMC8072828/