HAL open science

Investigating the role of the intestinal microbiota in colorectal cancer development : epidemiological approaches

Semi Zouiouich

▶ To cite this version:

Semi Zouiouich. Investigating the role of the intestinal microbiota in colorectal cancer development : epidemiological approaches. Cancer. Université de Lyon, 2021. English. NNT : 2021LYSE1011 . tel-03662440

HAL Id: tel-03662440

https://theses.hal.science/tel-03662440

Submitted on 9 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N°d'ordre NNT : 2021LYSE1011

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de
l'Université Claude Bernard Lyon 1

Ecole Doctorale N°205
Ecole Doctorale Interdisciplinaire Science-Santé (EDISS)

Spécialité de doctorat : Épidémiologie, Santé Publique, Recherche sur les services de santé
Discipline : Épidémiologie

Soutenue publiquement le 20/01/2021, par :
Sémi Zouiouich

Étude du rôle du microbiote intestinal dans la carcinogenèse colorectale : approches épidémiologiques

---

Devant le jury composé de :

Fervers, Béatrice Professeure associée, Université Lyon 1	Présidente
Cross, Amanda Professeure, Imperial College	Rapporteure
Engstrand, Lars Professeur, Karolinska Institutet	Rapporteur
Touvier, Mathilde Directrice de Recherche INSERM, Université Paris 13	Rapporteure
Deschasaux, Mélanie Chargée de Recherche INSERM, Université Paris 13	Examinatrice
Gunter, Marc Responsable de laboratoire, CIRC	Directeur de thèse
Huybrechts, Inge Chercheuse, CIRC	Co-directrice de thèse
Smelov, Vitaly Technical Officer, WHO, Regional Office for Europe	Co-directeur de thèse

Université Claude Bernard – LYON 1

Administrateur provisoire de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Pierre ROLLAND

COMPOSANTES SANTE

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

Laboratory

Nutrition and Metabolism Section

International Agency for Research on Cancer (IARC) - World Health Organization (WHO)

150 cours Albert Thomas

69372 Lyon CEDEX 08, FRANCE

Abstract

The gut microbiome is the ensemble of microorganisms that inhabit the gastrointestinal tract. A dynamic host-microbiome symbiosis exists in the intestine and specific interactions between the microbiota and host metabolism and immune system are critical for shaping host physiology. Disturbance of this symbiosis has been hypothesized to play a role in the development of various chronic diseases, including colorectal cancer. Indeed, there is growing experimental evidence that the gut microbiome potentially influences tumor development through dysregulation of host metabolism and immune function. However, epidemiological data linking the gut microbiome with colorectal carcinogenesis remains limited as very few existing cohorts have collected fecal samples. Colorectal cancer screening programs, in which millions of stool samples are collected each year, might provide rich opportunities to establish population-based cohorts with repeated, prospectively collected samples.

The first part of the thesis systematically summarizes the current epidemiological literature that has been published in the past decade on the association of the human microbiome with cancer. Our findings emphasised that for most microbiome indicators, the evidence was still too weak to draw firm conclusions in relation to their role in cancer.

The second part of the thesis investigates the association between insulin resistance and inflammation - recognised colorectal cancer risk factors - and the gut microbiome in two population-based cohorts – the Northern Finland Birth Cohort and TwinsUK. Our study indicated that higher levels of insulin resistance and other markers of metabolic dysfunction were associated with lower microbiome diversity in both cohorts, even after control for obesity and other factors.

Finally, the third part of the thesis evaluates microbiome stability and accuracy in fecal samples collected using different methods employed in ongoing colorectal cancer screening programs. Our findings suggest that commonly used fecal sample collections such as fecal immunotests and paper-based collection cards are, in general, suitable media for microbiome measurements though storage factors such as ambient temperature can impact on stability for some methods. In addition, the

opportunistic collection of fecal samples in fecal immunochemical test tubes after colorectal cancer screening is likely a viable method for establishing cohorts with prediagnostic fecal specimens.

Overall, this thesis presents the state-of-the-art on epidemiological evidence for the role of the microbiome in tumorigenesis, provides novel insights on the association of metabolic risk factors for colorectal cancer with the gut microbiome in population-based studies, and finally has generated important methodological data on the impact of fecal sample collection tools on microbial measurements that is needed for future epidemiological research on the microbiome and colorectal cancer, as well as other chronic diseases.

Résumé

Le microbiome intestinal désigne l'ensemble des micro-organismes présents dans le tractus gastro-intestinal. Une symbiose dynamique entre l'hôte et le microbiome existe dans l'intestin et des interactions spécifiques entre le microbiome et le métabolisme de l'hôte, et son système immunitaire, sont essentielles pour façonner la physiologie de l'hôte. L'hypothèse que la perturbation de cette symbiose joue un rôle dans le développement de diverses maladies chroniques, y compris du cancer colorectal, a été émise. En effet, il existe de plus en plus de preuves expérimentales selon lesquelles le microbiome intestinal influence potentiellement le développement de tumeurs par le biais d'une dérégulation du métabolisme de l'hôte et de sa fonction immunitaire. Cependant, les données épidémiologiques reliant le microbiome intestinal à la carcinogenèse colorectale restent limitées car très peu de cohortes existantes ont collecté des échantillons fécaux. Les programmes de dépistage du cancer colorectal, dans lesquels des millions d'échantillons de selles sont prélevés chaque année, pourraient fournir de riches opportunités pour établir des cohortes basées sur la population avec des échantillons répétés et prospectivement collectés.

La première partie de la thèse résume de manière systématique la littérature épidémiologique actuelle qui a été publiée au cours de la dernière décennie sur l'association entre le microbiome et le cancer. Nos résultats ont souligné que pour la plupart des indicateurs du microbiome, les preuves étaient encore trop faibles pour tirer des conclusions définitives concernant leur rôle dans le cancer.

La deuxième partie de la thèse étudie l'association entre la résistance à l'insuline et l'inflammation - des facteurs de risque reconnus du cancer colorectal - et le microbiome intestinal, dans deux cohortes populationnelles : The Northern Finland Birth Cohort et TwinsUK. Notre étude a indiqué que des niveaux plus élevés de résistance à l'insuline et d'autres marqueurs de dysfonctionnement métabolique étaient associés à une diversité amoindrie du microbiome dans les deux cohortes, même après contrôle de l'obésité et d'autres facteurs.

Enfin, la troisième partie de la thèse évalue la stabilité et la concordance du microbiome dans les échantillons fécaux prélevés à l'aide de différentes méthodes

utilisées dans les programmes de dépistage du cancer colorectal en cours. Nos résultats suggèrent que les collections d'échantillons fécaux couramment utilisées, telles que les tests immunochimiques fécaux et les cartes de prélèvement sur papier, sont en général des supports appropriés pour mesurer la diversité microbiome, bien que des facteurs de stockage, tels que la température ambiante, puissent avoir un impact sur la stabilité de certaines méthodes. En outre, la collecte opportuniste d'échantillons fécaux dans des tubes à tests immunochimiques fécaux après le dépistage du cancer colorectal est probablement une méthode viable pour établir des cohortes avec des échantillons fécaux pré-diagnostics.

Dans l'ensemble, cette thèse présente l'état de l'art sur les preuves épidémiologiques du rôle du microbiome dans la tumorigenèse. Elle fournit de nouvelles informations sur l'association des facteurs de risque métaboliques du cancer colorectal avec le microbiome intestinal dans les études populationnelles. Enfin, elle a généré d'importantes données méthodologiques sur l'impact des outils de collecte d'échantillons fécaux sur les mesures microbiennes nécessaires pour les futures recherches épidémiologiques sur le microbiome et le cancer colorectal, ainsi que sur d'autres maladies chroniques.

Résumé substantiel

Le microbiome intestinal désigne l'ensemble des micro-organismes, incluant les bactéries commensales et pathogènes, les virus, les archées et les mycètes, ainsi que leur matériel génétique présents dans le tractus gastro-intestinal. Le microbiome intestinal possède de nombreuses fonctions métaboliques, telles que la production d'acides gras à chaîne courte via la fermentation de polysaccharides complexes, la déconjugaison d'acides biliaires, la synthèse de lipopolysaccharides spécifiques et la biosynthèse de certains acides aminés essentiels et vitamines. Une symbiose dynamique entre l'hôte et le microbiome existe dans l'intestin et des interactions spécifiques, telles que la digestion d'aliments et l'extraction de nutriments ou le métabolisme de certains médicaments, entre le microbiome et le métabolisme de l'hôte, et son système immunitaire, sont essentielles pour façonner la physiologie de l'hôte. Cependant, cette symbiose repose sur un équilibre fragile, et des perturbations des interactions entre l'hôte et le microbiome intestinal ont été associées à plusieurs maladies chroniques telles que l'obésité, le diabète et potentiellement différents types de cancer.

Le cancer colorectal est le troisième cancer le plus fréquent chez les hommes et le deuxième chez les femmes, et est également la deuxième cause de décès par cancer dans le monde chez les deux sexes combinés. Bien que les facteurs génétiques, tel que les antécédents familiaux de cancer colorectal, ainsi que l'inflammation chronique induite par les maladies inflammatoires de l'intestin, contribuent au développement du cancer colorectal, il existe des preuves indiquant que de nombreux facteurs modifiables liés à l'alimentation et au mode de vie affectent le risque de cancer colorectal. De plus en plus d'indices suggèrent que le microbiome intestinal influence potentiellement l'initiation et la progression du développement tumoral par le biais d'une dérégulation du métabolisme de l'hôte et de sa fonction immunitaire, ainsi que par des perturbations des processus homéostatiques favorisant l'inflammation chronique et l'instabilité du génome induisant des mutations.

Le lien potentiel entre le microbiome intestinal et la néoplasie colorectale représente une hypothèse prometteuse pour étudier les caractéristiques inexpliquées du cancer colorectal. Cependant, les données épidémiologiques reliant le microbiome intestinal

à la carcinogenèse colorectale restent limitées car très peu de cohortes existantes ont collecté des échantillons fécaux. Les données basées sur des études cas-témoins ne permettent pas de déterminer si le processus de cancérogenèse modifie l'environnement local et crée une nouvelle niche pour les micro-organismes, ou si les altérations de la population microbienne contribuent à la cancérogenèse. Des échantillons répétés, collectés prospectivement à partir d'études de cohortes populationnelles, sont donc nécessaires pour une meilleure compréhension de la temporalité des associations microbiennes avec le cancer colorectal. Les programmes de dépistage du cancer colorectal, dans lesquels des millions d'échantillons de selles sont prélevés chaque année, pourraient fournir de riches opportunités pour établir des cohortes basées sur la population avec des échantillons répétés et collectés prospectivement.

L'objectif de cette thèse est de contribuer à déchiffrer les associations supposées entre le microbiome intestinal et le développement du cancer colorectal en renforçant les preuves épidémiologiques sur la relation entre le microbiome intestinal et des facteurs de risque établis de cancer colorectal, notamment la résistance à l'insuline et l'inflammation chronique, ainsi qu'en évaluant des méthodes de collection d'échantillons de matières fécales utilisées dans les programmes de dépistage.

La première partie de la thèse résume de manière systématique la littérature épidémiologique actuelle qui a été publiée au cours de la dernière décennie sur l'association entre le microbiome et le cancer. En effet, si certaines revues et méta-analyses portant sur les profils de microbiome en lien avec des cancers spécifiques ont déjà été publiées, aucun aperçu systématique des associations les plus significatives entre la composition du microbiome (par exemple les taxons et la diversité) et les différents sites de cancer n'a été publié à ce jour. En raison de la diversité des données publiées, un aperçu des preuves épidémiologiques disponibles liant le microbiome au cancer est nécessaire.

Nos résultats indiquent que bien que des preuves soient disponibles pour le lien entre certains taxons du microbiome intestinal et le risque de cancer colorectal, et pour le microbiome buccal et le risque de cancer buccal, pour la plupart des indicateurs du microbiome, les preuves sont encore trop faibles pour tirer des conclusions définitives

concernant leur rôle dans le cancer. La plupart des preuves nécessitent une validation dans des études prospectives basées sur la population.

La deuxième partie de la thèse étudie l'association entre la résistance à l'insuline et l'inflammation - des facteurs de risque reconnus du cancer colorectal - et le microbiome intestinal, dans deux cohortes populationnelles : The Northern Finland Birth Cohort et TwinsUK. Alors que l'hypothèse selon laquelle le microbiome intestinal influence le développement de la résistance à l'insuline a été émise, les données humaines sur la résistance à l'insuline et d'autres paramètres métaboliques et le microbiome intestinal sont encore limitées, car très peu de cohortes ont collecté des échantillons sanguins et fécaux. La résistance à l'insuline est une cause potentielle du diabète et d'autres maladies métaboliques, et constitue un facteur de risque établi de cancer colorectal. Par conséquent, déchiffrer la physiopathologie de la résistance à l'insuline à l'aide de données populationnelles pourrait aider à comprendre les mécanismes de développement du cancer colorectal et à renforcer les preuves de stratégies préventives potentielles.

Notre étude a indiqué que des niveaux plus élevés de résistance à l'insuline et d'autres marqueurs de dysfonctionnement métabolique étaient associés à une diversité amoindrie du microbiome dans les deux cohortes, même après contrôle de l'obésité et d'autres facteurs. Cependant, en raison de la nature transversale de cette étude, il est impossible de conclure si ces différences dans la composition du microbiome et les associations de taxons sont une cause ou une conséquence d'un dysfonctionnement métabolique.

Enfin, la troisième partie de la thèse évalue la stabilité et la concordance du microbiome dans les échantillons fécaux prélevés à l'aide de différentes méthodes utilisées dans les programmes de dépistage du cancer colorectal en cours dans différents pays. En effet, pour potentiellement conseiller l'établissement d'études internationales ayant des échantillons fécaux répétés et collectés prospectivement basés sur des programmes de dépistage du cancer colorectal, un travail méthodologique est nécessaire pour tester les tubes de test immunochimique fécal et les cartes de prélèvement d'échantillons utilisées dans les pays où différents dispositifs de dépistage sont utilisés.

Nos résultats suggèrent que les collections d'échantillons fécaux couramment utilisées, telles que les tests immunochimiques fécaux et les cartes de prélèvement sur papier, sont en général des supports appropriés pour mesurer la diversité microbiome, bien que des facteurs de stockage, tels que la température ambiante, puissent avoir un impact sur la stabilité de certaines méthodes. En outre, la collecte opportuniste d'échantillons fécaux dans des tubes à tests immunochimiques fécaux après le dépistage du cancer colorectal est probablement une méthode viable pour établir des cohortes avec des échantillons fécaux pré-diagnostics.

Dans l'ensemble, cette thèse présente l'état de l'art sur les preuves épidémiologiques du rôle du microbiome dans la tumorigenèse. Elle fournit de nouvelles informations sur l'association des facteurs de risque métaboliques du cancer colorectal avec le microbiome intestinal dans les études populationnelles. Enfin, elle a généré d'importantes données méthodologiques sur l'impact des outils de collecte d'échantillons fécaux sur les mesures microbiennes nécessaires pour les futures recherches épidémiologiques sur le microbiome et le cancer colorectal, ainsi que sur d'autres maladies chroniques.

List of publications

Published

Huybrechts I*, **Zouiouich S***, Loobuyck A, Vandenbulcke Z, Vogtmann E, Pisanu S, Iguacel I, Scalbert A, Indave I, Smelov V, Gunter MJ, Michels N (**2020**) The human microbiome in relation to cancer risk: a systematic review of epidemiological studies. *Cancer Epidemiology, Biomarkers & Prevention*. doi: 10.1158/1055-9965.EPI-20-0288.*co-first authors

Butt J, Jenab M, Pawlita M, Tjønneland A, Kyrø C, Boutron-Ruault MC, Carbonnel F, Dong C, Kaaks R, Kühn T, Boeing H, Schulze MB, Trichopoulou A, Karakatsani A, La Vecchia C, Palli D, Agnoli C, Tumino R, Sacerdote C, Panico S, Bueno-de-Mesquita B, Vermeulen R, Gram IT, Weiderpass E, Borch KB, Quirós JR, Agudo A, Rodríguez-Barranco M, Santiuste C, Ardanaz E, Van Guelpen B, Harlid S, Imaz L, Perez-Cornago A, Gunter MJ, **Zouiouich S**, Park JY, Riboli E, Cross AJ, Heath AK, Waterboer T, Hughes DJ (**2020**) Antibody Responses to Helicobacter pylori and Risk of Developing Colorectal Cancer in a European Cohort. *Cancer Epidemiology, Biomarkers & Prevention*. doi: 10.1158/1055-9965.EPI-19-1545.

Baena A, De Vuyst H, Mesher D, Kasubi M, Yuma S, Mwaiselage J, **Zouiouich S**, Mlay P, Kahesa C, Landoulsi S, de la Luz Hernandez M, Lucas E, Herrero R, Almonte M, Broutet N (**2020**) Reproducibility of a rapid human papillomavirus test at different levels of the healthcare system in Tanzania: the AISHA study. *Cancer Epidemiology, Biomarkers & Prevention*. https://doi.org/10.1158/1055-9965.EPI-20-0422.

Ongoing submission / Article in preparation

Zouiouich S, Loftfield E, Huybrechts I, Viallon V, Louca P, Vogtmann E, Wells P, Steves C, Herzig KH, Menni C, Jarvelin MR, Sinha R, Gunter MJ (**2020**) Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies. *Diabetologia*. (**Provisionally accepted - Under revision**)

Zouiouich S, Mariadassou M, Rué O, Vogtmann E, Huybrechts I, Severi G, Boutron-Ruault MC, Senore C, Naccarati A, Mengozzi G, Kozlakidis Z, Jenab M, Sinha R,

Gunter MJ, Leclerc M (**2020**) Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs. *mSphere*. (**Submitted**)

Noh H, Jang HH, Kim G, **Zouiouich S,** Cho SY, Kim HJ, Kim J, Choe JS, Gunter MJ, Ferrari P, Scalbert A, Freisling H (**2020**) Taxonomic composition and diversity of the gut microbiota in relation to habitual diet in Korean adults. Preprint. doi: 10.21203/rs.2.22869/v1. (**Submitted**)

Table of contents

Chapter I: Introduction 13
1. The human gut microbiome 13
1.1. General information on the human gut microbiome 13
1.2. Gut microbiome metabolism 19
1.3. Gut microbiome in insulin resistance, diabetes and cancer 22
2. Colorectal cancer 26
2.1. Colorectal cancer: pathophysiology, epidemiology and etiology 26
2.2. Gut microbiome and colorectal cancer 29
2.3. Colorectal cancer screening and gut microbiome 31
3. Thesis aim and objectives 33
3.1. The human microbiome in relation to cancer risk 33
3.2. Markers of metabolic health and gut microbiome diversity 33
3.3. Comparison of fecal sample collection methods for microbial analysis 34
Chapter II: The human microbiome in relation to cancer risk: a systematic review of epidemiological studies 35
1. Background 35
2. Objective 35
3. Methods 35
4. Results 36
5. Conclusion 37
6. Scientific article 37
Chapter III: Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies 51
1. Background 51
2. Objective 51
3. Methods 51
4. Results 52
5. Conclusion 52
6. Scientific article 53
Chapter IV: Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs 77
1. Background 77
2. Objective 77

3. Methods 77
4. Results 78
5. Conclusion 79
6. Scientific article 79
Chapter V: Discussion and perspectives 114
1. Discussion 114
2. Research perspectives 118
3. Conclusion 124
References 126
Appendices 146
Supplementary material: The human microbiome in relation to cancer risk: a systematic review of epidemiological studies 146
Supplementary Table S1 146
Supplementary Table S2 147
Supplementary Table S3 148
Supplementary Table S4 179
Supplementary Table S5 224
Supplementary material: Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs 246
Table S2a 246
Table S2b 247
Table S2c 249

Chapter I: Introduction

1. The human gut microbiome

1.1. General information on the human gut microbiome

1.1.1. Definitions

The rapid evolution of the microbiomics field has been accompanied by the introduction of a wide vocabulary used to describe different aspects of these communities and their environments. In this chapter of the thesis, key terms, such as microbiota, microbiome, taxonomy, 16S ribosomal RNA and diversity will be defined.

Human microbiota

The human body, including mucosal and skin environments, is not only composed of somatic cells but also of a multitude of microbial cells derived from commensal, symbiotic and pathogenic bacteria, viruses, archaea and fungi (1). Together, these microorganisms form a complex bacterial ecosystem termed the human microbiota.

Gut microbiota

The human gut microbiota is the microbial ecosystem of the gastrointestinal digestive tract and is among the most complex of the body sites, including 500 to 1,000 different species (2,3). In most scientific publications, the gut microbiota refers to the bacterial population present in the digestive tract, without taking into account other microorganisms (i.e. viruses, archaea and fungi).

Gut microbiome

These trillions of microorganisms interact with their host through the expression of several millions of genes, which together comprise the human microbiome (4). Each bacterial strain has a genome containing thousands of genes, offering more genetic diversity than the human genome (2). A rough approximation of 1,000 bacterial species in the gut with 2,000 genes per species yields an estimate of 2,000,000 genes, 100 times the number of approximately 20,000 genes composing the human genome (5,6).

In many studies, the terms “microbiome” and “microbiota” are used interchangeably and are often synonymous. Even if these definitions are overlapping, small but relevant differences exist (7) (Table 1).

Table 1. Differences between three major terms of the vocabulary of microbiome research

Microbiome	Microbiota	Metagenome
The microorganisms, their genomes and the surrounding environmental conditions	The microorganisms present in a defined environment	The collection of genomes and genes from the members of a microbiota
Refers to the bacteria (primary) and their genes (secondary)	Refers to the taxonomy of microorganisms	Refers to the collective functions of microbial genes

Bacterial taxonomy

Bacterial taxonomy is the rank-based classification of bacteria. A taxon groups together different organisms to form a unit. The current bacterial classification includes domain, phyla, class, order, family, genus and species. The species constitutes the basic unit in the classification of living organisms (8). Nowadays, the most generally applied method for determining phylogenetic relationships between microorganisms is based on metagenomics, as high-throughput sequencing technologies have seen a rapid development in recent years. The purpose of metagenomics is to quantify the DNA of a multitude of species in a particular ecosystem at once, allowing the identification of non-cultivable bacterial species by standard methods in laboratories (9,10). Two sequencing approaches are used today: targeted metagenomics, consisting in amplifying and sequencing a single gene present in several species (11), and shotgun metagenomics, consisting in sequencing all the DNA from all organisms in a sample (12). Comparative analysis of the 16S ribosomal RNA (or 16S rRNA) gene sequences is particularly common for the study of the gut microbiome (13).

16S ribosomal RNA

16S rRNA is part of the 30S small subunit of the ribosome present in every prokaryotic cell. The gene encoding for the subunit,16S rRNA gene, and its sequence is approximately 1,550 base pairs long (14). The 16S rRNA gene is used as a

phylogenetic marker as it is highly conserved between different species of bacteria and archaea (15). In addition to highly conserved primer binding sites, 16S rRNA gene sequences contain hypervariable regions that can provide species-specific signature sequences useful for identification of bacteria (16) (Figure 1).

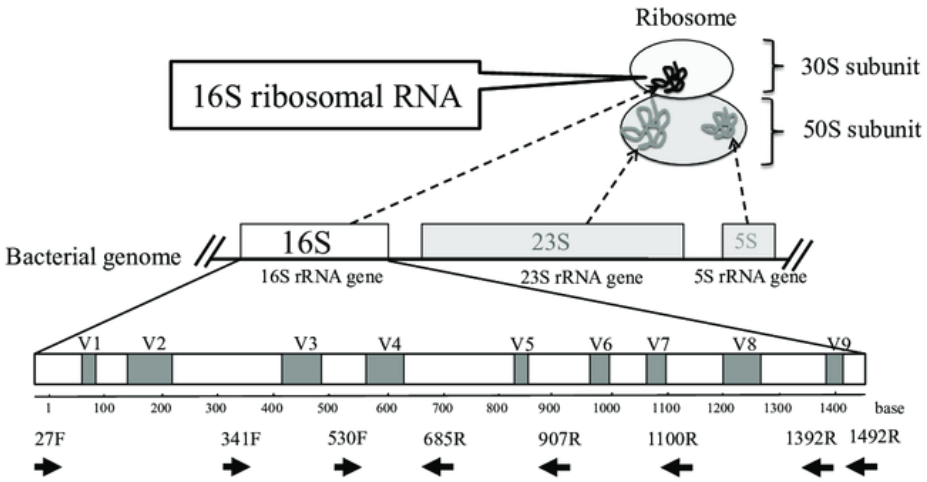


Figure 1. Schematic representation of ribosome complex and 16S rRNA gene
From Fukuda et al, 2016 (17)

Different DNA-based techniques, such as quantitative PCR (qPCR) or 16S rRNA sequencing, are widely used in microbiology studies to identify diversity in bacteria. When analysing 16S rRNA using qPCR, a universal primer is applied to the sample. The partial 16S gene amplicons that result from this primer contain the hypervariable regions, which is compared to other known sequences through sequence alignment. The result helps deduce phylogenetic relationships between bacteria. qPCR is a quantitative technique with the advantage of being fast, specific and precise for the detection of minor populations of bacteria (18). However, the main disadvantage of qPCR is that the primers are designed to only quantify a selection of specific bacterial strains.

16S rRNA sequencing has been widely used for identification and taxonomic classification of bacterial species. 16s rRNA sequencing refers to sequencing amplicons of the 16s rRNA gene. While sequencing the entire 16s rRNA gene is difficult due to the read length restrictions of many next generation sequencing platforms, sequencing one or more hypervariable regions is relatively rapid and cheap

(19). In contrast to qPCR, 16S rRNA sequencing does not only focus on specific bacterial strains but rather on the complex bacterial community. Using 16S rRNA gene sequences, numerous bacterial genera and species have been reclassified and renamed, classification of uncultivable bacteria has been made possible, phylogenetic relationships have been determined, and the discovery and classification of novel bacterial strains has been facilitated (20). However, although 16S rRNA gene sequencing is highly useful in regard to bacterial classification, it has low phylogenetic power at the species level and poor discriminatory power for some genera (21).

Operational taxonomic units (OTUs) are clusters of similar sequence variants of the 16S rRNA gene sequences (22). Each of these clusters are intended to represent a taxonomic unit of a bacteria species depending on the sequence similarity threshold. In past years, using a 97% identity cut-off was a standard approach to distinguish bacteria at the genus level. Recent approaches are now focused towards amplicon sequence variants (ASVs), focusing on minor differences (in many cases single nucleotide variations) to identify unique bacterial species (23,24).

Microbial diversity

Changes in the microbiome composition and population diversity, is typically measured with taxonomic diversity. The taxonomic diversity is defined by two components: the richness which describes how many different types of bacteria are observed in a sample, in terms of their taxonomic ranking and the evenness, which designates how evenly these bacteria are distributed (25).

To measure biodiversity, two concepts are generally used in microbiome studies: alpha and beta diversity (26). The alpha diversity refers to the diversity within a single sample and is usually expressed by the number of species in that ecosystem. The beta diversity shows the difference between microbial communities and generally focuses on taxonomic abundance profiles from different samples.

Even though there is no general agreement on which diversity index is the best to use, several diversity indices have been commonly used to study microbial diversity (Table 2). For alpha diversity, these metrics are the Shannon diversity index, inverse Simpson

index, and the number of observed OTUs/ASVs. For beta diversity, these metrics include Bray-Curtis dissimilarity, Jaccard index and UniFrac.

Table 2. Ecological diversity measures commonly used in microbiome studies

Diversity indices	Description	References
Shannon diversity index	Based on species richness and species evenness: more weight on species richness	(27)
Inverse Simpson index	Based on species richness and species evenness: more weight on species evenness	(27,28)
Observed OTUs/ASVs	Based on species richness: number of species present in a community	(29,30)
Bray-Curtis dissimilarity	Based on abundance: differences in microbial abundances between two samples	(31)
Jaccard index	Based on presence or absence of species: differences in microbial composition between two samples	(32)
UniFrac	Based on sequence distances: fraction of branch length shared between two samples or unique to one or the other sample	(33)

1.1.2. Composition of the human gut microbiota

Composition at high taxonomic levels shows the overall stability of the gut microbiota between individuals. In healthy individuals, intestinal microbiota are typically dominated, at the phyla level, by Gram-negative *Bacteroidetes* and Gram-positive *Firmicutes*, with a smaller abundance of *Actinobacteria* (34,35). However, the proportion of these phyla is not static, and different bacterial strains may compete to fulfil distinct ecological niches, inducing considerable inter-individual variation between phenotypically similar hosts. Therefore, each individual harbours radically different complexity and diversity of bacterial community in the digestive tract, making each gut microbiota unique (36,37).

Our understanding of what leads to inter-individual variation in diversity remains limited, even if several studies tend to show that a number of host (e.g. genetics, age) and lifestyle factors (e.g. diet, obesity, medications, mode of delivery at birth) could

influence the composition of the gut microbiota through the life-course, potentially to the benefit or detriment of the host (38–42) (Figure 2).

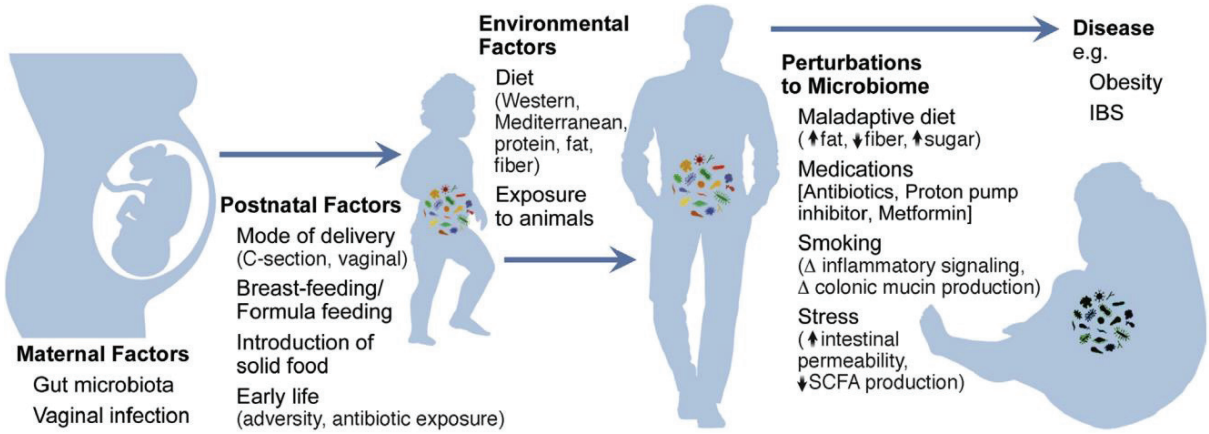


Figure 2. Evolution of the gut microbiome over the human life-course
Adapted from Dong & Gupta, 2019 (43), SCFA: short-chain fatty acid, IBS: inflammatory bowel syndrome

Typically-consumed diets, such as the Western diet and the Mediterranean diet, can have a significant impact on the gut microbiome (44). The Western diet, characterised by high-fat, high-sugar, high level of red and processed meat, high levels of refined grains and a lower level of fibre, was associated to a decrease in overall bacterial richness/diversity, and beneficial *Bifidobacteria* and *Eubacteria* (43,45). This alteration of the gut microbiome was associated with gut barrier degradation, reduced levels of short-chain fatty acids (SCFAs), higher levels of lipopolysaccharides (LPS) and inflammation (46,47). In comparison, the Mediterranean diet, characterised by a beneficial fatty acid profile, higher intake of fibre, vegetables, and fruits, and with lower intake of sugar and red meat, was associated to increased bacterial richness/diversity, higher levels of SCFAs and decreased inflammation (43,48). Consumption of dietary fibre increases bacterial richness and diversity with a relative abundance in such genera as *Prevotella* and *Treponema*, and enhance the production of SCFAs (39,49).

In addition to diet and other early life events (50), such as the mode of delivery at birth and breast-feeding, there are many other lifestyle exposures throughout adulthood that can lead to changes in microbial composition. Certain medications, such as antibiotics and proton pump inhibitors, can trigger gut microbiome imbalances which can persist

over time and increase the host's susceptibility to infection and metabolic syndrome (51,52). Smoking has also been associated to changes in gut microbiota composition, as shown by strong microbial alterations before and after smoking cessation (53). More recently, in addition to the previously mentioned factors that are relatively well described in the literature, it has been shown that alcohol consumption and bowel movement quality might have a very strong impact on the gut microbiome (54–56). Most importantly, in contrast to other dietary factors influencing the gut microbiome, these two factors have also been shown to differ when comparing diseased subjects to healthy individuals and to influence disease associations by confounding the effects of disease (56).

1.2. Gut microbiome metabolism

The gut microbiome is highly metabolically active and helps define essential physiological functions such as digestion of food and extraction of nutrients (57), drug metabolism (58), and modification of the host immune response and metabolism (59). Interestingly, microbial diversity between individuals does not appear to critically influence core functions in microbial metabolism, including the production of SCFAs via fermentation of complex polysaccharides, deconjugation of bile acids, synthesis of specific LPS and the biosynthesis of some essential amino acids and vitamins (5,60–62).

1.2.1. Short-chain fatty acids

Since the human body lacks the digestive enzymes required to digest many dietary plant polysaccharides and oligosaccharides, these dietary fibres are metabolically processed by the gut microbiota (63,64). Bacterial fermentation of dietary non-digestible polysaccharides in the human colon produces SCFAs, largely composed of acetate, propionate and butyrate (65,66). These end products of bacterial fermentation are produced by different bacteria in the colon. For instance, butyrate is produced by several bacteria from the phyla *Firmicutes* such as *Faecalibacterium prausnitzii*, or *Roseburia spp* (67).

SCFAs play a major role in the maintenance of normal physiology in the host including the stimulation of water absorption, the production of energy for intestinal epithelial cell

growth, and can either be absorbed from the colon or bind to free fatty acid receptors which are expressed in the intestinal epithelium (68–70). More specifically, butyrate is an important source of energy for colonic epithelial cells, propionate is utilized in the liver as a substrate for the production of glucose via gluconeogenesis, and acetate, once transformed into acetyl-CoA, is an important biochemical intermediate in human metabolism (68,71).

1.2.2. Bile acids

Bile acids represent the primary pathway for cholesterol catabolism in the liver and they primarily promote absorption of lipids, including fat-soluble vitamins in the intestine (72). Deconjugation of primary bile acids by the gut microbiota results in the formation of secondary bile acids, including deoxycholic and lithocholic acids, derived from cholic acid and chenodeoxycholic acid, respectively (73).

These bioconversions modulate the signalling properties of bile acids through the nuclear farnesoid X receptor and the G protein-coupled membrane receptor 5, which regulate diverse metabolic pathways in relation to glucose and lipid metabolism in the host (74). More precisely, secondary bile acids regulate the hepatic triglyceride pool via nuclear farnesoid X receptor-dependent inhibition of sterol regulatory element binding protein-1, therefore inhibiting lipogenesis (75). Secondary bile acids also stimulate the secretion of the incretin glucagon like peptide-1 through binding to the G protein-coupled membrane receptor 5 on enteroendocrine cells, thus improving insulin sensitivity, attenuating weight gain and increasing energy expenses (76).

1.2.3. Lipopolysaccharide

LPS is the major component of the outer layer of the membrane of Gram-negative bacteria and is, in non-capsulated strains, exposed on the cell surface (77). Distinct structures of LPS can be found on the surface of many pathogenic bacteria (e.g. *Escherichia coli*, *Salmonella*, *Yersinia pestis*) and therefore, have been termed endotoxin in reference to the cell-associated toxicity. LPS performs several functions in Gram-negative bacteria such as serving as a major structural component of the outer membrane, transforming the outer membrane into an effective permeability barrier against small, hydrophobic molecules and most importantly (78), playing a crucial role

in bacteria-host interactions by modulating responses by the host immune system. Gut microbial LPS is thought to be one of the most important mediators of the microbiome's influence on host physiology, notably because chronic immune cell activation is, in part, caused by the LPS-mediated stimulation of toll-like receptor 4 (79). However, our understanding of the mechanisms that support host-microbe interactions through LPS is still limited. While many assume that proinflammatory signalling by LPS is a central feature of the gut microbiome activity, others consider that proinflammatory pathways commonly triggered by bacterial pathogens upon interaction with the host are, in fact, actively repressed by beneficial bacteria of the gut microbiome to maintain homeostasis in the gut (80,81).

1.2.4. Amino acids

The gut microbiota is also involved in the digestion and metabolism of endogenous and exogenously-derived amino acids. Amino acids are among the major nutrients in the diet and regulate energy and protein homeostasis in organisms (82,83). Some functional amino acids, such as tryptophan, glutamine, methionine, and branched-chain amino acids, also have beneficial effects on the immune system through modulation of key metabolic signalling pathways (84). Along the gastrointestinal tract, alimentary and endogenous proteins are hydrolysed into peptides and amino acids by host- and bacteria-derived proteases and peptidases (85). A number of microorganisms, such as bacteria belonging to *Clostridium* and *Streptococcus*, are particularly involved in amino acid fermentation in the large intestine (86).

Amino acid-fermenting bacteria are not only important for protein digestion and subsequent amino acid absorption, but they also secrete various metabolites, characterised by nitrogen- and sulfur-containing material. The products are, sometimes, detrimental to the health of the host (87) but may also influence cellular signalling pathways and exert immune and barrier effects and thus, maintain homeostasis of the host (83). After being metabolised by microorganisms, amino acids can also be used to synthesise constituents, such as SCFAs or other bacterial proteins (88). Additionally, the gut bacteria can be involved in *de novo* biosynthesis of certain amino acids, suggesting that the exchange of amino acids between the host and the gut microbiota is bidirectional (89).

1.3. Gut microbiome in insulin resistance, diabetes and cancer

Elaborate mechanisms relying on host-driven tolerance are required to maintain the host-microbe symbiosis in the intestine, including physical barriers of the oriented gut tissues or secretion of antimicrobial peptides and secreted antibodies (90). However, this symbiosis is based on a dynamic and fragile equilibrium, and disturbances of the interactions between the host and the gut microbiome have been associated with several chronic diseases such as obesity (91,92), diabetes (93,94) and potentially different types of cancer (95,96).

1.3.1. Gut microbiome in insulin resistance and metabolic syndrome

Type 2 diabetes (T2D) has reached global epidemic proportions and is recognised as a major cause of morbidity and mortality (97). Insulin resistance, characterised as the reduced sensitivity of glucose storage tissues to insulin-mediated biologic activity, is a condition that precedes the development of T2D (98), and often accompanies obesity, though not exclusively. Indeed, individuals with insulin resistance are not systematically overweight or obese (99) and the existence of metabolically-healthy obese and metabolically unhealthy normal-weight individuals has been described in previous studies (100). Insulin resistance is also related to poor glucose control and high levels of chronic inflammation, which together, comprise elements of the metabolic syndrome (101). The incidence of metabolic syndrome often parallels the incidence of obesity and incidence of T2D, and it has been estimated that more than 20% of adults worldwide are known to have metabolic syndrome (102). However, the prevalence of metabolic syndrome varies depending on environmental factors, the age and gender composition of the population, genetic differences, and lifestyle factors including physical activity level and eating habits (103,104). Despite the well-established epidemiological associations between metabolic syndrome and its related metabolic parameters (e.g. insulin resistance) and chronic diseases, the biological mechanisms underlying its development remain to be deciphered.

Insulin and insulin-like growth factor-1 mediate their biological effects via the insulin and insulin-like growth factor-1 receptors. Cellular actions of insulin and insulin-like growth factor-1 involve a wide variety of effects on post-receptor signalling pathways

within target cells (105). These signalling pathways contain several points of regulation, signal divergence, and cross talk with other signalling cascades that are essential to mediate the variety of insulin and insulin-like growth factor-1 biological responses (106). Mechanisms of insulin resistance are therefore multifactorial. In most cases, insulin resistance is caused by cellular perturbations, such as lipotoxicity, inflammation, glucotoxicity, mitochondrial dysfunction, and endoplasmic reticulum stress (105). These potential complex mechanisms lead to deregulation of genes and inhibitory protein modifications and result in impaired insulin and insulin-like growth factor-1 action (Table 3).

Previous studies have suggested that the gut microbiome might influence host metabolic health through several pathways that are directly or indirectly related to insulin resistance development (93,94,107,108). These pathways included energy extraction, intestinal barrier integrity, metabolism of bile acids and host metabolic and signalling pathways (109). Alterations of microbial composition and functionality in T2D patients compared to healthy participants have been described (93,94,110). For example, circulating levels of branched-chain amino acids have been positively associated with insulin resistance (111) and correlate with specific changes in gut microbiome composition (e.g. *P. copri* and *B. vulgatus*) and functions of individuals with insulin resistance (112). In addition, low bacterial diversity, characterised by the reduction in butyrate-producing bacteria (e.g. *Roseburia* and *Faecalibacterium prausnitzii*) and the increase in mucin-degrading bacteria, was associated with impaired gut integrity causing low-grade inflammation of visceral adipose tissue through endotoxemia (107,113,114) and may provide a link between obesity and insulin resistance (115). If animal and human studies have identified relevant differences in intestinal microbiota composition in subjects with insulin resistance related chronic diseases, the magnitude of effect of the gut microbiome on metabolic functions remains to be proven (116).

Table 3. Possible molecular mechanisms of insulin resistance pathogenesis

From Boucher et al, 2014 (105)

Cause	Mechanism	Effect
Lipotoxicity **Inflammation** **Hyperglycemia** **Mitochondrial dysfunction** **Endoplasmic reticulum stress**	Activation of Ser/Thr kinases	Inhibitory phosphorylation of insulin-signalling molecules
Genetic mutations	Point mutations in insulin receptor and insulin-signalling molecules	Increased protein turnover Reduced expression and ligand affinity Decreased signalling capacities
	Single-nucleotide polymorphism causing increased gene expression	Increased phosphatase and tensin homolog action leading to reduced phosphatidylinositol (3,4,5)-triphosphate levels
Lipotoxicity	Hyperactivation of protein phosphatase protein phosphatase 2A	Reduced phosphorylation of insulin receptor and insulin-signalling molecules
Inflammation	Cytokine-induced suppressor of cytokine signalling-3 protein activation	Inhibition or insulin receptor tyrosine kinase activity Competition for insulin receptor substrate binding to insulin receptor Increased insulin receptor substrate degradation
	Cytokine-induced reduction in gene expression	Decreased expression of insulin-signalling molecules
Hyperglycaemia	Glycation of insulin-signalling molecules	Reduced affinity for insulin receptor Decreased DNA-binding capacities of transcription factors
	Hyperactivation of protein phosphatase protein phosphatase 2A	Reduced phosphorylation of insulin receptor and insulin-signalling molecules
Hyperinsulinemia	Hyperactivation of leucine-rich repeat protein phosphatases-1 and Grb14	Decreased Akt Ser 473 phosphorylation Competition for insulin receptor substrate binding to insulin receptor

Insulin resistance and subsequent hyperinsulinemia and T2D are conditions which are associated with increased risk of cancer at a number of anatomical sites (117). For example, metabolic changes such as hyperinsulinemia and insulin resistance are associated with colorectal cancer (118–120). The mechanisms linking insulin resistance and metabolic dysfunction with colorectal cancer are not fully understood but may reflect enhanced signalling through insulin receptors and insulin-like growth factors (i.e. insulin-like growth factor-1 and 2) and signal transduction to pro-mitogenic and anti-apoptotic pathways that enable cell proliferation and survival of transformed cells (121). Insulin stimulates cell proliferation by binding with low affinity to the insulin-like growth factor-1 receptor, inducing high serum levels of insulin-like growth factor-1 and increasing risk of cancer, including colorectal cancer (122,123). Furthermore, some cancer cells promote tumor proliferation by locally producing insulin-like growth factor-2 which binds with high affinity to foetal isoforms of the insulin receptor in tumor cells (124). Since insulin resistance, T2D and obesity are known cancer-promoting conditions, understanding the relationship between the gut microbiome and metabolic syndrome might help to understand the mechanisms of cancer development and open new possibilities for preventive strategies as well as screening, diagnosis, and treatment (125).

1.3.2. Role of the gut microbiome in cancer

There is growing evidence that the gut microbiota potentially influence the initiation and progression of tumor development (126,127). Since disturbances of the host-microbiome homeostasis have been associated with many established hallmarks of cancer (128), the gut microbiome likely influences cancer risk at different levels, including cancer initiation, promotion, dissemination, and response to therapy (129). Broadly, these hallmarks comprise dysregulations of host metabolism (93,107,108) and immune function (59,79), perturbations of homeostatic processes promoting chronic inflammation (130–132) and promotion of genome instability and mutations (133,134) (Figure 3).

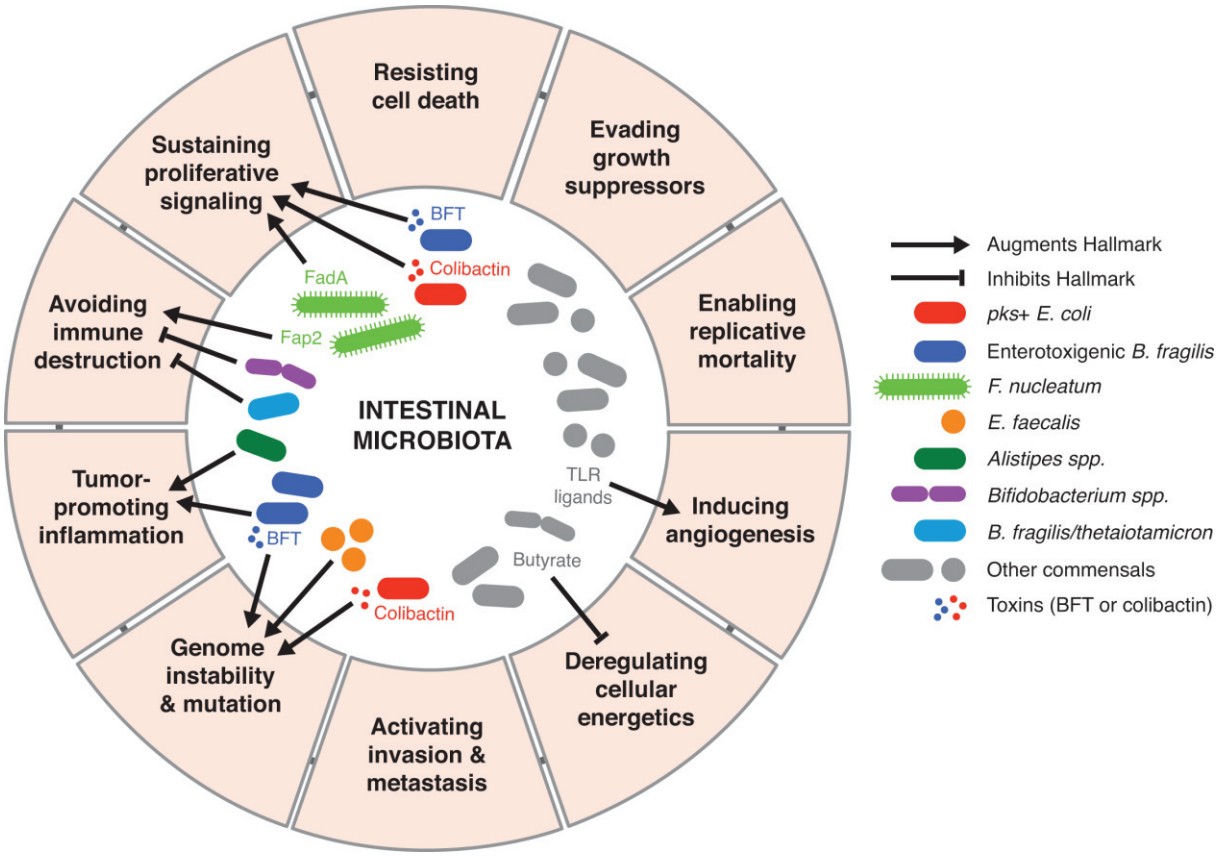


Figure 3. The gut microbiome influences several hallmarks of cancer through diverse pathways

From Fulbright et al, 2017 (128)

2. Colorectal cancer

Although recent studies have implicated the gut microbiota in the development of cancers at distal sites (135,136), the earliest observations linking the microbiota with the hallmarks of cancer have primarily focused on colorectal cancer as the most complex and dense populations of endogenous microorganisms are found within the gastrointestinal tract.

2.1. Colorectal cancer: pathophysiology, epidemiology and etiology

2.1.1. Pathophysiology

Colorectal cancer develops through four main stages: initiation, promotion progression and metastasis (137). Briefly, initiation involves irreversible genetic damage to target cells that allow neoplastic transformations occur. The initiated cells proliferate during the promotion stage, inducing abnormal growth. During the progression phase, through

the gradual accumulation of genetic and epigenetic alterations, benign tumour cells transform into malignant cancer cells and acquire aggressive characteristics and metastatic potential. Finally, metastasis is marked by the spread of cancer cells from the primary organ to other organs or tissues through the bloodstream or the lymphatic system. The duration of each phase is difficult to estimate but these mechanisms generally take a long time and decades may be required for all stages to be completed in colorectal cancer (138).

Colorectal cancer occurs through at least three distinct carcinogenic pathways (139). The adenoma-carcinoma sequence pathway accounts for 85 to 90% of sporadic colorectal cancer. Normal cells progress to small adenoma, to large adenoma and, finally, to cancer. This pathway is predominantly associated with chromosomal instability (140). In addition, the serrated pathway accounts for 10 to 15% of sporadic colorectal cancer. This model is characterised by the progression from normal cells, to hyperplastic polyp, to sessile serrated adenomas and, finally, to cancer (141). This pathway is predominantly associated with CpG island methylator phenotype (142). Finally, a twofold increased risk of colorectal cancer was previously observed in patients with inflammatory bowel disease (143). Due to the low incidence of inflammatory bowel disease, the inflammatory pathway explains less than 2% of all colorectal cancers. Driven by chronic inflammation, normal cells progress to indefinite dysplasia, to low-grade dysplasia, to high-grade dysplasia and, finally, to cancer (144).

2.1.2. Epidemiology

Worldwide, colorectal cancer is the third most common cancer in men and the second most common cancer in women, and is also the second leading cause of cancer-related death in the world among both sexes combined (145). Over the next 15 years, the number of cases of colorectal cancer is expected to increase by 72% to more than 3 million (146). While the incidence of colorectal cancer in many countries with a high human development index has decreased among those older than 50 years, a recent study established that rates of colorectal cancer in individuals younger than 50 years increased by up to 4% per year (147). In addition, in many countries that underwent major development transitions, particularly in Eastern Europe, Asia and South America, increases in both colorectal cancer incidence and mortality have been

observed (146,148). The rise in incidence in countries in transition and the generational changes may reflect the role of secular changes in colorectal cancer risk factors that have influenced incidence rates.

2.1.3. Etiology

Although genetic factors, including family history of colorectal cancer and hereditary cancer syndromes (143,149), as well as chronic inflammation induced by inflammatory bowel diseases, contribute to colorectal cancer development, there is strong evidence that many modifiable dietary and lifestyle factors affect colorectal cancer risk (Table 4) (150,151). These factors include obesity, low physical activity and sedentary lifestyle, poor diets (such as higher consumption of red and processed meat, high fat, low fibre, low whole grain and low calcium), smoking tobacco and excessive alcohol consumption (152–154).

Table 4. Dietary and lifestyle factors associated with strong evidence to colorectal cancer risk

Adapted from World Cancer Research Fund, 2017 (152)

	Decreases risk	Increases risk
Convincing	Physical activity	Processed meat Alcoholic drinks Body fatness Adult attained height
Probable	Wholegrains Foods containing dietary fibre Dairy products Calcium supplements	Red meat

Further, two dietary patterns have been consistently observed to be associated with colorectal cancer. The first pattern indicates a potential protective effect of a diet characterised by high intake of fruit and vegetables as well as fibre and whole grains, nuts or legumes, fish or other seafood, calcium and dairy products (152). The other pattern, considered as unhealthy and mainly represented by the so-called Western diet, is characterised by high intake of red and processed meat, sugar-sweetened beverages, refined grains, desserts and potatoes (155). Biological mechanisms linking dietary patterns and colorectal cancer are probably multifaceted, reflecting a complex

interplay of various dietary components instead of the effect of individual nutrients or specific foods (155).

Physical activity might also reduce colorectal cancer risk through its beneficial effects on gut motility, the immune system, inflammation and metabolic hormones (156). In addition, regardless of the level of physical activity, a sedentary lifestyle, as characterised by prolonged sitting, is emerging as a risk factor for colorectal cancer by impairing skeletal muscle function, which facilitates insulin resistance and promotes colorectal carcinogenesis (157,158).

As previously indicated, overweight, obesity and T2D are established risk factors for colorectal cancer and may directly account for more than 10% of colorectal cancer cases worldwide (159,160). In obesity there is dysregulation of adipose tissue metabolism and immunology characterised by high levels of proinflammatory adipokines, which could contribute to the development of chronic low-grade inflammation promoting tumour growth and progression (161,162). However, important questions remain regarding the mechanisms underlying the link between obesity and colorectal cancer. For example, it is not known if obesity-related inflammation increases colorectal cancer risk directly or secondarily through processes such as insulin resistance (163).

2.2. Gut microbiome and colorectal cancer

Advances in characterising the gut microbiome have offered potential insights into the etiology of colorectal cancer. In fecal samples collected within a case-control study, colorectal cancer risk was positively associated with decreased bacterial diversity with enrichment of *Bacteroidetes* and depletion of *Firmicutes* in colorectal cancer cases compared to non-cancer controls (95). In another case-control study, the composition of the gut microbiota was investigated in colorectal cancer precursors and revealed differences in the relative distribution of eight phyla between adenomatous and control colorectal tissues (164). Accumulating evidence supports the likelihood that the collective activity of the gut microbiota might influence colorectal cancer development and progression (165,166). However, over the past decade, several species have garnered specific interest for the potential roles in colorectal carcinogenesis.

Data from *in vitro* and murine models have demonstrated that *Fusobacterium nucleatum* might promote colorectal cancer cell proliferation and increase tumor growth rates (167). Several mechanisms have been proposed to explain these findings (Figure 4). *F. nucleatum* promotes cell proliferation by binding to E-cadherin on the colorectal cancer cell surface through FadA adhesin and activates the oncogenic Wnt/ β-catenin signalling pathway (168). Through the effect of Fap2 adhesin, *F. nucleatum* facilitates colorectal cancer enrichment by binding to a disaccharide sugar motif and alters the function of tumor-infiltrating lymphocytes and natural killer cells by binding to the inhibitory immune receptor TIGIT (169,170). *F. nucleatum* also facilitates resistance to chemotherapy by activating autophagy through LPS and Toll-like receptor 4 expressed on colorectal cancer cells (171).

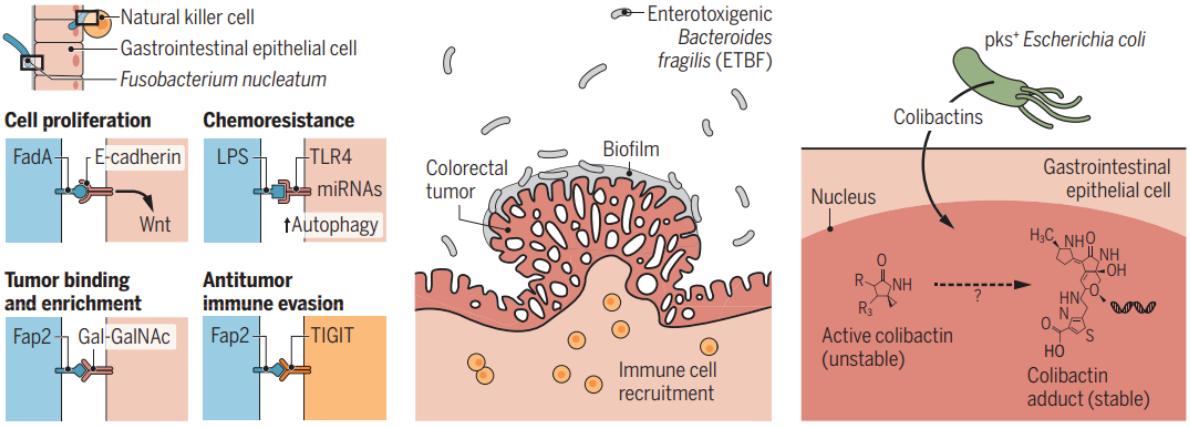


Figure 4. Three examples of microorganisms that might drive colorectal cancer
From Garrett, 2015 (129)

In addition, in mice, enterotoxigenic *Bacteroides fragilis*, a pathogen that causes diarrhoea and inflammation, potentiates colorectal carcinogenesis and was detected in biofilms covering human adenocarcinomas and adenomas (172). The importance of colibactin-producing *Escherichia coli* in colorectal cancer tumorigenesis has been confirmed, when the *E. coli* genotoxin-producing strains (known to induce double-strand DNA breaks and cell cycle arrest) were found in more than 50% of colorectal cancer cases (173).

For all the aforementioned examples, it is apparent that variation in the gut microbiota and potential link with colorectal neoplasia might represent a promising hypothesis that could offer insights to unexplained features of the epidemiology of colorectal cancer. However, it remains unclear whether bacterial features of the gut microbiota have a causative role in colorectal carcinogenesis and due to several limitations, evidence on the potential mechanisms involved are still too weak to draw definitive conclusions.

2.3. Colorectal cancer screening and gut microbiome

As described above, colorectal cancers typically develop from normal epithelium through sequential phases of adenomatous dysplasia over a period that can span decades. Colorectal cancer survival was therefore strongly correlated with stage at diagnosis, which supports the importance of colorectal cancer screening programs where neoplastic lesions can be detected at an early stage (174). The recommended screening for the average-risk population (i.e. individuals aged 50 years or older, with no additional risk factors) is one of the following: an annual or biennial stool test, sigmoidoscopy every 5 years, or colonoscopy every 10 years (Figure 5) (175).

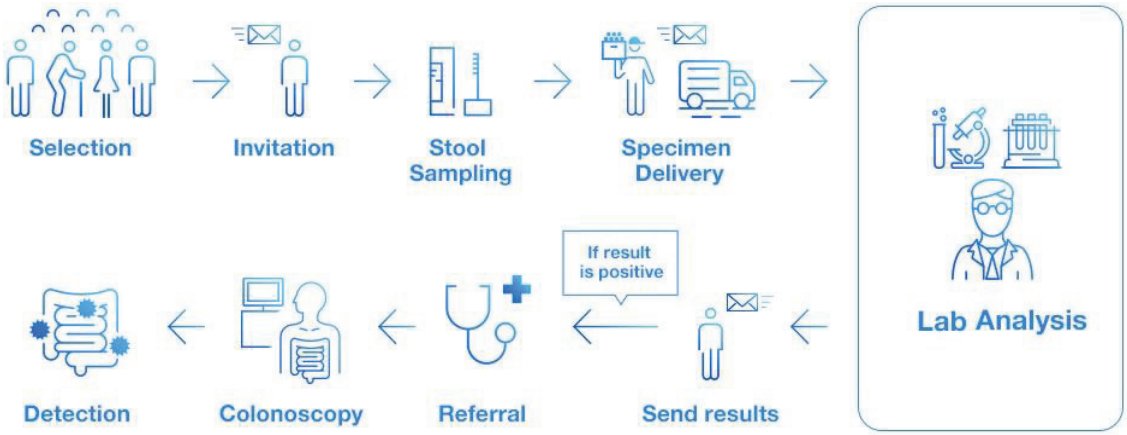


Figure 5. Schematic representation of colorectal cancer screening protocol using FIT

From Eiken Chemical Co., Ltd. (176)

Several stool tests are being used in countries where colorectal cancer screening registries are implemented, each of these modalities differing by specificity, sensitivity, acceptance, and economic impact. One screening tool, which has been widely implemented for its excellent specificity, is the fecal occult blood test (FOBT). Despite

its poor sensitivity for the detection of colorectal adenomas, significant reductions in colorectal cancer mortality have been reported with the use of FOBT in screening-based randomised trials (174,177,178). Subsequently developed after FOBT, the fecal immunochemical test (FIT) for the detection of human haemoglobin in stool has higher sensitivity for the detection of colorectal cancers and early stages compared with FOBT (179,180). Observational studies on the evaluation of FIT in colorectal cancer screening programs are very consistent and one, reported relative risks of death from colorectal cancer 10% to 40% lower among individuals screened by FIT (181–183). Therefore, the FIT has been recommended as the first option for detecting fecal occult blood in colorectal cancer screening (184) and is increasingly being used in countries with an organised colorectal cancer screening program (185). For example, in France, the FIT has been offered every two years to average-risk individuals, aged 50 to 74, since 2015. According to Santé Publique France, approximately 5,300,000 individuals were screened by FIT between 2018 and 2019, in France. Of these, around 198,000 (3.8%) were FIT-positives (186). These colorectal cancer screening registries in which millions of FOBT or FIT samples are being collected each year might provide rich opportunities to establish cohorts of individuals with pre-diagnostic collected fecal samples.

Indeed, epidemiological studies of the association between the gut microbiome and colorectal cancer development have so far been very limited, usually with small sample sizes and case-control and other cross-sectional designs without repeated measurements. Thus, due to potential reverse causality, caution is required in the interpretation of colorectal cancer-microbiome associations (187). In data based on case–control studies, it is not possible to determine whether the carcinogenic process changes the local environment and creates a new niche for microorganisms or whether alterations in the microbial population contribute to carcinogenesis. Repeated, prospectively collected samples from population-based cohort studies are necessary for better understanding of the temporal nature of microbial associations with colorectal cancer.

3. Thesis aim and objectives

The main aim of this thesis was to contribute to unravelling the putative associations of the gut microbiome with colorectal cancer development by strengthening epidemiologic evidence on the relationship between the gut microbiome and established colorectal cancer risk factors, including insulin resistance and chronic inflammation, and by evaluating fecal sample collection methods used in ongoing screening programs with a view to providing the necessary evidence for establishing future cohorts with prediagnostic fecal samples. To address this over-arching aim, this thesis has been divided into three objectives.

3.1. The human microbiome in relation to cancer risk

The first part of the thesis aimed to summarize the current epidemiological literature that has been published in the past decade on the association of the microbiome with cancer. Indeed, while some reviews and meta-analyses focusing on microbiome profiles for single cancer sites have already been published, no systematic overview of the most significant associations between microbiome composition (e.g. taxa and diversity) and different cancer sites was published to date. Due to the diversity of published data, an overview of available epidemiologic evidence linking the microbiome with cancer was needed.

Objective I:
Evaluate the strengths and weaknesses of the current evidence on the relationship between the human microbiome and cancer in epidemiologic studies.

3.2. Markers of metabolic health and gut microbiome diversity

The second part of the thesis aimed to investigate the association between insulin resistance and inflammation, recognised colorectal cancer risk factors, and the gut microbiome in population-based epidemiological studies. While the gut microbiome has been hypothesised to influence insulin resistance development, human data on insulin resistance and other metabolic parameters and the gut microbiome are still limited as very few cohorts have collected both blood and fecal specimens. Insulin resistance lies on the causal pathways of diabetes and other metabolic diseases, and

is an established risk factor for colorectal cancer. Therefore, deciphering the pathophysiology of insulin resistance and inflammation using population-based data might help to understand the mechanisms of colorectal cancer development and strengthen evidence for potential preventive strategies.

Objective II:
Investigate the association of gut bacteria 16S rRNA gene sequence data with serologic levels of homeostatic model for insulin resistance, glycated haemoglobin and C-reactive protein in two distinct population-based cohorts, the Northern Finland Birth Cohort 1966 and TwinsUK.

3.3. Comparison of fecal sample collection methods for microbial analysis

The third part of the thesis evaluated microbiome stability and accuracy in fecal samples collected using different methods employed in ongoing colorectal cancer screening programs. Indeed, to potentially inform the establishment of international studies with repeated, prospectively collected fecal specimens based in colorectal cancer screening programs, methodologic work was needed to test FIT and specimen collection cards used in countries where different screening devices are employed.

Objective III:
Compare the impact of different fecal collection methods currently used in ongoing colorectal cancer screening programs on microbiome technical reproducibility, accuracy and stability.

Chapter II: The human microbiome in relation to cancer risk: a systematic review of epidemiological studies

1. Background

The gut microbiome has potentially been associated with cancer development (128) but due to the diversity of published data, the exact mechanisms of this relationship are not clear (187). A large number of individual observational studies that have examined the association of the microbiome with single cancer types have been published in the past decade, however, no systematic overview of the most significant associations between microbiome composition and diversity, and different cancer sites was published to date.

2. Objective

The objective of this study was to evaluate the strengths and weaknesses of the current evidence on the relationship between the human microbiome and cancer in epidemiologic studies.

3. Methods

Inclusion and exclusion criteria

A systematic review was conducted to identify peer-reviewed publications related to the human microbiome and cancer. Observational studies (i.e. case–control, including case-crossover design or cohort studies) and randomized controlled trials were considered for inclusion. Reviews, case reports, and other studies without a comparison group, as well as *in vitro* and animal studies were excluded. Adult humans with or without any type of cancer were included while underage persons (<18 years) were excluded. Studies investigating the prevalence or incidence of cancer and its association with the human microbiome were considered for inclusion. Studies investigating only precancerous lesions/conditions (no malignant conditions) were excluded. Studies focusing on single bacteria instead of communities of microorganisms were excluded.

Data collection and analysis

Tailored search strings (see Supplementary Material Table S1 in annex) were used to search for relevant articles from database conception until December 2019 via the PubMed and EMBASE databases. Studies were reviewed and selected independently by four reviewers. Significant differences in relative abundance of taxa and/or diversity measures between cancer cases and controls were extracted and reported. If multiple analyses were run with different levels of adjustment, then the one with maximal adjustment for confounders was chosen. The methodologic quality of the included articles was assessed independently by two reviewers using the Newcastle-Ottawa quality assessment scale (188). An association was considered as “strong” when three or more publications reported a statistically significant association in the same direction and none were in the opposite direction. The association was deemed “suggestive” when two publications reported a statistically significant association in the same direction and none in the opposite direction.

4. Results

A total of 124 articles (including 15,764 subjects in total; 7,652 cancer cases vs. 8,112 controls) were included and used for this review. Among all the anatomical sites evaluated, the gut microbiome followed by the oral microbiome were most frequently studied. The majority of articles reported specific differences in microbiome diversity and/or composition between cases and control groups (including not only non-cancer patients but sometimes tissues). Fifty studies reported differences in gut microbiome composition between colorectal cancer patients and various controls. Overall, consistent findings showed that abundance of pro-inflammatory opportunistic taxa, such as *Fusobacterium*, *Parvimonas* and *Porphyromonas*, was increased in colorectal cancer patients while a depletion of butyrate-producing bacteria was observed. Sixteen studies described differences in oral microbiome composition between oral cancer patients and controls. Enrichment of *Fusobacterium* and depletion of *Streptococcus* were reported in patients with oral cancer compared with controls. The quality assessment demonstrated rather low quality for the “ascertainment of the exposure of interest” because studies often used different methods of ascertainment for cases and controls, for example, resection for cases while biopsy for controls. The large diversity

of parameters used to describe the microbial composition made it impossible to harmonize the different studies in a way that would allow meta-analysis.

5. Conclusion

Although strong evidence was available for certain taxa of the gut microbiome and colorectal cancer risk, and for the oral microbiome and oral cancer risk, for most of the microbiome taxa/indicators the evidence was still too weak to draw firm conclusions in relation to their role in cancer. Most findings now require validation in population-based prospective studies in which standardised methods are applied.

6. Scientific article

Huybrechts I*, **Zouiouich S***, Loobuyck A, Vandenbulcke Z, Vogtmann E, Pisanu S, Iguacel I, Scalbert A, Indave I, Smelov V, Gunter MJ, Michels N (**2020**) The human microbiome in relation to cancer risk: a systematic review of epidemiological studies. *Cancer Epidemiology, Biomarkers & Prevention*. doi: 10.1158/1055-9965.EPI-20-0288.*co-first authors

Role: Co-first author, contribution to data extraction and quality controls in text and tables, contribution to original draft preparation and revised version.

The Human Microbiome in Relation to Cancer Risk: A Systematic Review of Epidemiologic Studies

Inge Huybrechts[1], Semi Zouiouich[1], Astrid Loobuyck[2], Zeger Vandenbulcke[2], Emily Vogtmann[3], Silvia Pisanu[1,4], Isabel Iguacel[1,5], Augustin Scalbert[1], Iciar Indave[1], Vitaly Smelov[1,6], Marc J. Gunter[1], and Nathalie Michels[2]

ABSTRACT

The microbiome has been hypothesized to play a role in cancer development. Because of the diversity of published data, an overview of available epidemiologic evidence linking the microbiome with cancer is now needed. We conducted a systematic review using a tailored search strategy in Medline and EMBASE databases to identify and summarize the current epidemiologic literature on the relationship between the microbiome and different cancer outcomes published until December 2019. We identified 124 eligible articles. The large diversity of parameters used to describe microbial composition made it impossible to harmonize the different studies in a way that would allow meta-analysis, therefore only a qualitative description of results could be performed. Fifty studies reported differences in the gut microbiome between patients with colorectal cancer and various control groups. The most consistent findings were for *Fusobacterium, Porphyromonas*, and *Peptostreptococcus* being significantly enriched in fecal and mucosal samples from patients with colorectal cancer. For the oral microbiome, significantly increased and decreased abundance was reported for *Fusobacterium* and *Streptococcus,* respectively, in patients with oral cancer compared with controls. Overall, although there was a large amount of evidence for some of these alterations, most require validation in high-quality, preferably prospective, epidemiologic studies.

Introduction

The human microbiome defines either the microorganisms (bacteria, archaea, lower and higher eukaryotes, and viruses) found in and on the human body or their collective genomes (i.e., genetic material; refs. 1–3). Since the development of high-throughput approaches using next-generation sequencing, our understanding of the human microbiome, even the noncultivable microorganisms, has increased enormously, shedding light on its complexity (1, 4, 5). Although the gut microbiome has been most extensively studied, microorganisms also inhabit all of the barrier surfaces of the human body including the skin, the oral cavity, the nasopharynx, the esophagus and stomach, and also the vagina, the urinary tract, the lungs and others (4, 6). The composition of the microbiome varies depending on the anatomic site and it also differs between individuals (4). The microbiota, the ensemble of microorganisms living in a specific environment, form a dynamic entity, and can change within an individual in response to diet and other external factors such as medication, environment, and lifestyle (1, 7, 8).

In addition to the role that microbial imbalance (9) may play in infectious and autoimmune conditions; there is a growing appreciation of potential influences on cancer development (3, 6). For example, the microbiota and its metabolic products influence the development and maintenance of inflammation, which has been broadly recognized to be one of the hallmarks of cancer (10–12). Inflammation has been demonstrated to promote cancer by promoting genetic instability and, once the tumor has been established, through the creation of a tumor-promoting microenvironment. In addition, the microbiota may also influence the development of cancer-promoting conditions, such as adiposity and insulin resistance (13–15). Because of these effects on host metabolism, cellular proliferation, inflammation, and immunity, the microbiome may influence cancer risk at different levels, including cancer initiation, promotion, dissemination, and response to therapy (6, 16). The effects of the microbiota on cancer risk can be local, situated at the interface where the organism interacts with the host tissue in which the cancer originates, or systemic, via the physiologic communication of the host organism and the microbiota through intact membranes or following alteration of barrier permeability in pathologic situations. This potential link between the microbiome and cancer development might offer new opportunities for cancer prevention by understanding etiologic pathways for screening, diagnosis, and treatment (2, 3, 17).

A large number of individual studies on the association of the microbiome with cancer have been published in the past decade, and an in-depth review of the literature is now needed. While some published reviews and meta-analyses focusing on microbiome profiles for single cancer sites were recently published (18–20), no systematic overview of the most significant associations between microbiome composition (e.g., taxa and diversity) and different cancer sites has been published to date. Therefore, the objective of this systematic review was not only to summarize the published data, but also to evaluate the strengths and weaknesses of the current evidence on the

[1]International Agency for Research on Cancer, Lyon, France. [2]Department of Public Health and Primary Care, Ghent University, Ghent, Belgium. [3]Division of Cancer Epidemiology & Genetics, NCI, Bethesda, Maryland. [4]Department of Biomedical Sciences, Section of Microbiology and Virology, University of Cagliari, Cagliari, Italy. [5]GENUD (Growth, Exercise, NUtrition and Development) Research Group, Faculty of Health Sciences, University of Zaragoza, Zaragoza, Spain. [6]Division of Noncommunicable Diseases and Promoting Health through the Life-course, WHO Regional Office for Europe, Copenhagen, Denmark.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

I. Huybrechts and S. Zouiouich contributed equally as the co-first authors of this article.

Corresponding Author: Inge Huybrechts, International Agency for Research on Cancer, 150 Cours Albert Thomas, Lyon 69372, France. Phone: 334-7273-8148; Fax: 334-7273-8361; E-mail: huybrechtsi@iarc.fr

Cancer Epidemiol Biomarkers Prev 2020;29:1856–68

doi: 10.1158/1055-9965.EPI-20-0288

©2020 American Association for Cancer Research.

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

relationship between the human microbiome and cancer in epidemiologic studies.

Materials and Methods

A systematic review was conducted to identify peer-reviewed publications related to the human microbiome and cancer. The review considered the recommendations of the Meta-analysis of Observational Studies in Epidemiology (MOOSE) group (21) and followed the Preferred Reporting Items for Systematic Reviews and Metaanalyses (PRISMA; ref. 22). A protocol (Prospero registration number: CRD42018105860) was prepared following these guidelines (22).

Eligibility (inclusion and exclusion) criteria considered in this review

Type of studies

Studies that utilized the following designs were considered for inclusion:

(i) Observational studies: case–control, including case-crossover design or cohort studies.
(ii) Interventional studies: Randomized Controlled Trials (RCT).

Reviews, case reports, and other studies without a comparison group, cross-sectional studies, reports from conferences or annual meetings, editorials, opinions, *in vitro* studies and studies of animals were excluded from this review. Studies not written in English were also excluded from this review.

Type of participants

This review focused on adult humans with or without any type of cancer. Studies focusing on children (<18 years) were excluded.

Type of outcome and exposure measurements

Studies investigating the prevalence or incidence of cancer and its association with the human microbiome were considered for inclusion. Our primary interest was cancer development and not cancer survival/mortality. Studies investigating only precancerous lesions/conditions (no malignant conditions) were excluded.

Studies that target specific microbes (e.g., *Helicobacter pylori*) instead of studying a microbiome (community of microorganisms) were also excluded. Next-generation sequencing approaches, including 16S ribosomal RNA gene amplicons and genome shotgun metagenomics, are the most common methods for investigating communities of microorganisms. However other methods such as aerobic and anaerobic subculturing onto selective agar media were also considered in this review, provided that communities of microorganisms were investigated. Exposures of interest were relative abundancy of taxa and diversity measures (alpha and beta diversity) in cancer cases versus controls.

Data collection and analysis

Tailored search strings (see Supplementary Material Table S1) were used to search relevant articles up from database conception until December 2019 via the PubMed and EMBASE databases. Also, references in the selected publications were checked for further studies.

Selection of studies

Four investigators (A. Loobuyck and Z. Vandenbulcke for first search; I. Huybrechts and N. Michels for second search) independently reviewed titles and abstracts. In case no consensus was reached, the article was maintained for the next selection phase (e.g., selection by full-text article). When the disagreement was still present at full-text selection ($n = 6$), the opinion of the remaining co-authors involved in this review was requested.

Data extraction

Data were extracted by three researchers independently (I. Huybrechts, N. Michels, and S. Zouiouich). Considering the important differences in outcome measures reported in the different publications, it was decided to report the significant differences in relative abundancy of taxa and/or diversity measures (alpha and beta diversity) between cancer cases and controls. If multiple analyses were run with different levels of adjustment, then the one with maximal adjustment for confounders was chosen.

Methodologic quality and strength of association

Two reviewers (I. Huybrechts and N. Michels) independently assessed the methodologic quality of the included articles via the Newcastle –Ottawa quality assessment scale (ref. 23; See Supplementary Table S2). Each study was evaluated using three broad criteria: (i) appropriate selection of the study population (up to four stars); (ii) comparability of the study groups (up to two stars); and (iii) ascertainment of the exposure or outcome of interest (up to three stars). The full score was 9, and a high-quality study in our analysis was defined as a study with at least 7 points. The performance of a meta-analysis to calculate adjusted pooled estimates was ruled out because of the large variation in the definition of outcome variables and exposures. We performed instead a qualitative synthesis of results and reported them in categories according to microbiome and study design. An association was considered as "strong" when three or more publications reported a statistically significant association in the same direction and none were in the opposite direction. The association was deemed "suggestive" when two publications reported a statistically significant association in the same direction and none in the opposite direction. All analyses (e.g., counts of cases and controls and articles) were performed in Microsoft Excel.

Availability of data and material

All data generated or analyzed during this study are included in this published article (and its Supplementary Information files).

Results

Study design and population characteristics of the included studies

A total of 124 articles (including 15,764 subjects in total; 7,652 cancer cases vs. 8,112 controls) were included and used for this review (see Materials and Methods and **Fig. 1:** flowchart). An in-depth overview of the different studies and their study design and methods is given in Supplementary Table S3. 119 studies were case–control studies, including three nested case–control studies (24–26), and five were cohort studies (19, 27–30). No RCTs were identified. The gut microbiome was the microbiome most frequently studied ($n = 50$), followed by the oral microbiome ($n = 16$) and then some studies of the bile duct, cervical and intrauterine, esophagus and gastric, laryngeal, lung, skin, breast, urinary, and prostate microbiome (see **Table 1**).

Thirty-nine of the studies were conducted in North America (n subjects = 5,999), 5 in South America (n of subjects = 300), 53 in Asia (n subjects = 6,783), 27 in Europe (n subjects = 2,485), 2 in Africa (n subjects = 124) and one in Australia ($n = 83$). The year of

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

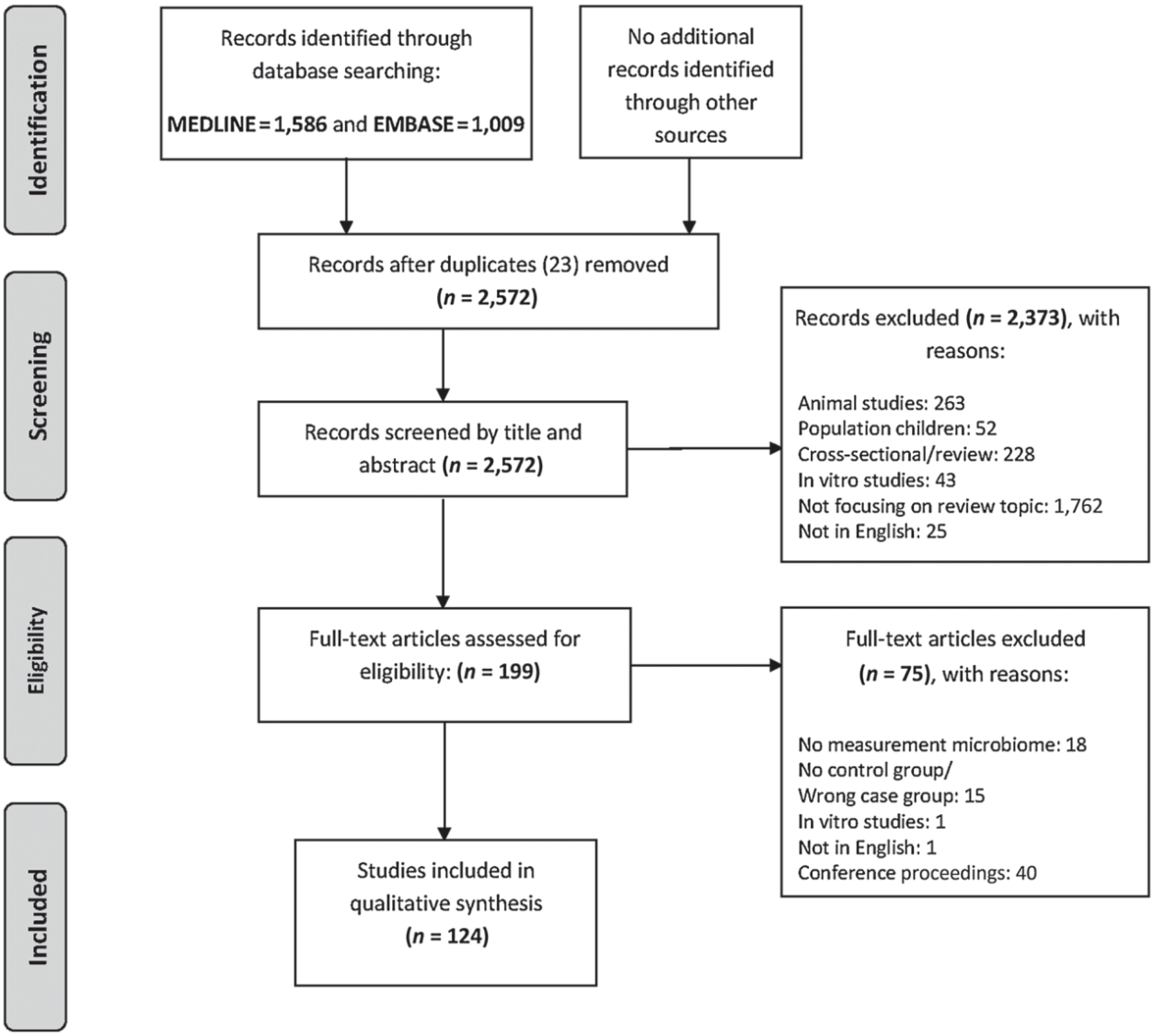


Figure 1.
PRISMA flow diagram. Adapted from: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7):e1000097. doi:10.1371/journal.pmed1000097. For more information, visit www.prisma-statement.org.

publication ranged from 1983 to 2019. The sample size of the studied populations showed large variation. Four articles had a sample size smaller than 10, 17 articles had a sample size ranging from 10 to 29, 27 articles had a population size from 30 to 49, 32 articles included a population size from 50 to 99, and 46 articles studied over 100 subjects. The largest number of subjects was included for the gut microbiome in relation to colorectal cancer risk (n = 5,761), followed by the oral microbiome–oral cancer associations (n = 1,912), and the oral microbiome–pancreatic cancer associations (n = 1,729). All studies included a population with an overall age ranging from 18 to 96 years old. In most of the studies, the cases tended to be slightly older than the healthy controls.

Quality of the studies

The methodologic quality assessments are presented in Supplementary Table S4. Eighty-eight articles had a good quality score of 7/9 or more of which 51 articles with a total quality score above 8 and two articles reaching the maximum score of 9 stars. For selection, 70 articles received the maximum score, (31), while the most frequent problem was lack of information about the selection of cases and controls and thus their representativeness for the community. For comparability, 82 articles scored the maximum score, while other articles often did not perform any correction for confounders between cases and controls of any kind. This decreases the comparability between the patient group and the healthy controls. For exposure, 20 articles received the maximum score, while low scores were mostly because the studies used another method of ascertainment between cases and controls (e.g., resection for cases while biopsy for controls).

Overall significant findings

As demonstrated in **Table 1** and Supplementary Table S4, a large number of samples from different anatomic sites have been investigated in relation to different cancer types. Supplementary Table S5 gives an overview of the different bacteria described in these

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

Table 1. Number of articles retrieved by microbiome type and the possible links with cancer.

Microbiome	Linked with the following cancer(s)	Articles, *n*	Subjects, *n*		
			Total subjects	Cases	Controls
Airway and lung microbiome	Lung cancer	7	309	189	120
Bile duct microbiome	Cholangiocarcinoma	2	260	160	100
Breast microbiome	Breast cancer	2	109	73	36
Cervical microbiome	Cervical cancer	3	289	143	146
Esophagus microbiome	Esophageal cancer	1	101	50	51
Gastric microbiome	Gastric cancer	10	999	687	312
	Esophageal cancer	1	91	37	54
Gut microbiome	Colorectal cancer	50	5,751	2,667	3,084
	Gastric cancer	2	204	116	88
	Liver cancer	1	321	150	171
	Prostate cancer	1	155	90	65
	Breast cancer	3	325	158	167
	Thyroid cancer	1	65	30	35
Laryngeal microbiome	Laryngeal cancer	1	60	29	31
Pharyngeal microbiome	Laryngeal cancer	1	96	68	28
Oral microbiome	Oral cancer	16	1,912	724	1,188
	Head and neck cancer	4	580	301	279
	Esophageal cancer	2	551	193	358
	Pancreatic cancer	5	1,729	819	910
	Gastric cancer	1	51	34	17
	Colorectal cancer	3	1,019	377	642
	Colorectal, lung, and gastric cancer	1	386	286	100
	Liver cancer	1	60	35	25
	Lung cancer	1	66	51	15
Ovarian microbiome	Ovarian cancer	1	119	99	20
Prostate microbiome	Prostate cancer	1	16	16	0
Skin microbiome	Skin cancer	1	32	15	17
Urinary microbiome	Bladder cancer	2	78	41	37
	Prostate cancer	1	30	14	16
All microbiomes	All cancers	124[a]	15,764	7,652	8,112

[a]Two articles included several cancers and as such were included twice in the table, namely for each cancer site. As such, the sum of the different articles based upon cancer site is higher (126) than the total number of articles used in this review (124).

studies and their direction of association with (site-specific) cancer risk. Some of the alterations in the site-specific microbiota were consistent between studies while others were only significant in one study or conflicting between studies (see also overview of significant findings for each study included in this review in Supplementary Table S4).

In addition, study results were not always comparable due to high heterogeneity in outcome definitions. Many of the articles used different parameters to describe and compare the microbiome including: alpha diversity, beta diversity, relative abundance, absolute abundance, presence/absence of taxa, and some articles clustered microbiota into communities or calculated ratios for the abundance of two microbiota taxonomic groups. Furthermore, the taxonomic level reported differed between articles (mainly the phyla, the families and genera were studied) and some of the articles showed no significant differences in presence and relative abundance of certain phyla while they did show differences at lower taxonomic levels. For example, one study (32) found the same five most abundant phyla within the cases with esophageal cancer or dysplasia and the healthy controls, but different abundances of the orders Clostridiales and Erysipelotrichales. Finally, there are also many different measures and statistical methods used to analyze alpha and beta diversity what often limits the comparability of studies reporting these diversity measures.

Because of the large amount of data presented in the different studies included in this review, we have focused on results for which at least two studies reported findings (e.g., diversity measures or taxa) in the same direction in the description below (and visualized in **Table 2**). A full overview of the results found in the different studies is given in the Supplementary Table S4 and the study design of the included studies is presented in Supplementary Table S3.

Cholangiocarcinoma has been investigated in relation to the bile duct microbiome in two articles (33, 34), although one article only focused on the extrahepatic variant. Both studies reported microbiome differences between cases and controls, although not for the same operational taxonomic units (OTU) and a clear separation between cases and controls when analyzing beta diversity.

Two studies investigated the breast tissue microbiome in relation to invasive breast cancer (using benign breast disease as controls), reporting increased relative abundance of certain genera, although these differed between the two studies. Increased abundance was reported in the following low-abundant family and genera in the breast tissue of women with invasive breast cancer: Unclassified Bacteroidetes, Comomonadaceae Enterobacteriaceae, *Bacillus and Staphylococcus* (35), *Fusobacterium, Atopobium, Hydrogenophaga, Gluconacetobacter*, and *Lactobacillus* (unadjusted $P < 0.05$; ref. 36). In both studies, microbial diversity was not considered.

Three studies investigated associations between the cervical microbiome and cervical cancer risk. Conflicting results were found for *Lactobacillus iners;* however, all studies reported a decrease in

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

Table 2. Summary table presenting associations between microbiome composition and colorectal [number of studies (n) = 50] and oral cancer risk (n = 16).

	Taxa and diversity	Colorectal cancer	*N* (%) studies for that finding[a]	Oral cancer	*N* (%) studies for that finding[a]
Phylum	Bacteroidetes			S & T	3 (19)
	Firmicutes			S	2 (13)
	Fusobacteria	F & T	10 (20)	S & T	2 (13)
Family	Fusobacteriaceae	F & T	2 (4)		
	Lachnospiraceae	F & T	6 (12)		
	Peptostreptococcaceae	F	2 (4)		
	Porphyromonadaceae	F	2 (4)		
	Ruminococcaceae	F &T	5 (10)		
Genus	*Actinomyces*			S	2 (13)
	Acinetobacter	F	3 (6)		
	Atopobium	F & T	4 (8)		
	Bifidobacterium	F & T	4 (8)		
	Blautia	F & T	2 (4)		
	Capnocytophaga			S & T	2 (13)
	Desulfovibrio	F	2 (4)		
	Dialister			S & T	2 (13)
	Fusobacterium	F & T	16 (32)	S & T	4 (25)
	Gemella	F & T	3 (6)		
	Leptotrichia	F & T & S	3 (6)		
	Odoribacter	F	2 (4)		
Species	*Oscillibacter*	F	2 (4)		
	Parvimonas	F	6 (12)	S & T	2 (13)
	Peptostreptococcus	F & T	5 (10)	S & T	2 (13)
	Porphyromonas	F & T	8 (16)		
	Prevotella	F	2 (4)		
	Staphylococcus	F	2 (7)		
	Streptococcus			S	4 (25)
	Bacteroides fragilis	F & T	4 (8)		
	Blautia coccoides	F	2 (4)		
	Escherichia coli	F & T	2 (4)		
	Fusobacterium nucleatum	F & T	11 (22)		
	Parvimonas micra	F	4 (8)		
	Peptostreptococcus stomatis	F	3 (6)		
	Porphyromonas assaccharolytica	F	2 (4)		
	Prevotella intermedia	F & S	2 (4)		
Diversity	Beta diversity	Sign. in 7 studies	7 (14)		

Abbreviations: F, faecal samples; S, saliva samples; T, tissue samples (cancer site specific).
[a]Example n (%): 10 and 2 studies for Fusobacteria for colorectal cancer and oral cancer, respectively, so 20% and 13% of the total number of studies on colorectal cancer and oral cancer.

Strong positive association: (≥3 publications in same direction; none in opposite direction)
Suggestive positive association: (2 publications in same direction; none in opposite direction)
Strong negative association: (≥3 publications in same direction; none in opposite direction)
Suggestive negative association: (2 publications in same direction; none in opposite direction)

abundance of the beneficial *Lactobacillus crispatus* in cervical cancer cases (37, 38). All three studies also suggested higher alpha-diversity (Shannon index and/or OTU number) among cervical cancer cases compared with controls (37, 38, 39).

One study investigated esophageal cancer risk in relation to the esophageal microbiome (40) while another study investigated its association with the gastric microbiome (32). No significant differences were found between the esophageal microbiome of the patients with esophageal cancer and the healthy controls with no evidence of esophageal disease (40). Differences were found in the gastric microbiome composition between controls and cases with esophageal cancer on the order, family, and genus level, though the abundances were different and not always in the same direction in the two studies (32, 41). Microbial diversity was considered in one study, in which the number of bacteria between cases and controls was not significant (40).

Ten studies investigated associations between the gastric microbiome and gastric cancer risk (42–51). The gastric microbiome of the cases with gastric cancer showed a significant difference compared with control subjects and in the cancer tissues compared with the adjacent normal tissue in the same patient with cancer. Three studies showed an increased abundancy of Fusobacteria among patients with gastric cancer or cancer tissue samples (44, 47, 52) and two studies an increased abundancy of Neisseria among patients with gastric cancer when compared with control subjects (46, 47). In two of these studies, the analysis of microbial diversity was included, showing a trend to diminish going from nonatrophic gastritis to intestinal type of gastric cancer (42, 47).

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

Seven studies evaluated associations between the airway-lung microbiome and lung cancer risk, using different methodologies [e.g., sputum samples (53, 54) vs. exhaled breath condensate and bronchial brushing (29, 30, 55–57)]. All studies demonstrated clear differences between the microbiome samples of patients with lung cancer in comparison with healthy controls, though different taxa were studied (Carpagnano only focused on Aspergillus species). Streptococcus were increased among patients with lung cancer in three studies (30, 54, 56).

Thirty-four studies investigated associations between the oral microbiome and several cancers, of which 16 considered the association with oral cancer (27, 31, 58–71) and five with pancreatic cancer (24, 72–74). The sixteen studies considering the associations between the oral microbiome and oral cancer risk obtained inconsistent results. Consistent results have been visually presented as strong or suggestive positive/negative associations in **Table 2**. From the 19 studies that reported results on the phylum *Bacteroidetes*, three studies confirmed a significant positive association between *Bacteroidetes* abundance and oral cancer risk. Similarly, four studies confirmed a positive association between the abundance of the genus *Fusobacterium* and oral cancer risk. Several studies also observed an altered abundance of *Streptococcus* in patients with oral cancer versus controls (27, 31, 58–60, 67, 69) with four studies confirming a significant negative association between *Streptococcus* abundance and oral cancer risk. Furthermore, six studies reported an altered abundance of *Prevotella*, but results differed in the direction of the association. Two of the five studies investigating associations between the oral microbiome and pancreatic cancer risk reported consistently a lower abundance of *Neisseria* in the oral microbiome of pancreatic cancer cases compared with controls (72, 73). Two studies found a lower abundance of *Neisseria* and *Haemophilus* when investigating the tongue coating samples of patients with colorectal, lung, and gastric cancer compared with controls (75, 76). Among the 34 studies analyzing the oral microbiome, 13 studies presented results on the microbial diversity and most of them showed a significant lower bacterial diversity than the healthy controls.

Two studies investigated associations between the urinary microbiome and bladder cancer risk (77, 78). Both studies reported a decreased abundancy of *Streptococcus* and Bucevic reported also increased abundancy of *Fusobacterium* among patients with bladder cancer. Bi and colleagues also reported a higher alpha diversity among patients with bladder cancer.

Fifty studies linked differences in the gut microbiome with colorectal cancer, three with breast cancer risk (79–81) and two with gastric cancer risk (82, 83). Thirty-three articles used fecal samples as a proxy for the gut microbiome, while 22 articles used mucosal biopsies. Three articles investigated both feces and mucosal biopsies (84–86). To compare within consistent specimen types, use of fecal samples or mucosal samples is clearly specified in the description of the results regarding associations between the gut microbiome and colorectal cancer risk here below. A summary of the strong and suggestive associations reported for the gut microbiome in relation to colorectal cancer risk has been visualized in **Table 2**.

The phylum Fusobacteria was more often detected in colorectal cancer tumor biopsies compared with adjacent normal tissue (25, 87–89). Seven studies also found significantly higher levels of the genus *Fusobacterium* in mucosal samples of colorectal cancer cases in comparison with controls (25, 84, 86, 88–92). In five studies, the species *Fusobacterium nucleatum* was enriched in mucosal samples in colorectal cancer cases (85, 87, 93–95). Thirteen studies also detected a significant enrichment of the genus *Fusobacterium* in fecal samples of patients with colorectal cancer (19, 52, 96–106).

Eight studies found that the genus *Porphyromonas* was significantly enriched in fecal samples of patients with colorectal cancer compared with controls (96, 99–102, 104–107).

Two studies showed that *Lactobacillus* was significantly more present in fecal samples from colorectal cancer cases compared with controls, while the studies of Mira-Pascual and colleagues and Ohigashi and colleagues found nonsignificantly higher abundance of *Lactobacillus* in controls (85, 108–110).

The genus *Prevotella* was more commonly detected in fecal samples of colorectal cancer cases compared with controls in the study of Amiot and colleagues and two other studies reported the same trend (85, 108, 109). *Peptostreptococcus* had significantly higher abundance in fecal samples from colorectal cancer cases compared with controls in five studies (52, 84, 100, 101, 104, 110). Two studies also found a significant increase of *Peptostreptococcus* in colorectal cancer tissue compared with healthy controls (84, 88).

Six studies demonstrated an increased abundance of *Parvimonas* in colorectal cancer cases (52, 95, 97, 100, 101, 105) and three studies a significantly higher presence of *Gemella* and *Leptotrichia* in fecal samples of colorectal cancer cases compared with controls (25, 84, 91, 100, 101).

In the Spanish cohort of Allali and colleagues *Blautia* was significantly more abundant in adjacent normal tissue compared with colorectal cancer tumor biopsies and Chen and colleagues also observed higher *Blautia* levels in healthy controls compared with colorectal cancer cases (25, 84).

At the family level, Ruminococceae and *Lachnospiraceae* were significantly more abundant in controls and/or adjacent/"off-tumor" mucosae compared with colorectal cancer tumor samples in four and six studies, respectively (84, 90, 91, 95, 104, 106, 111, 112).

In summary, the most consistent findings for the gut microbiome were for *Fusobacterium* (most important species *Fusobacterium nucleatum*), *Parvimonas, Porphyromonas*, and *Peptostreptococcus* being significantly enriched in fecal and mucosal samples from patients with colorectal cancer compared with controls or adjacent normal tissue.

Among the 50 studies investing associations between cancer risk and gut microbiome composition, results on microbial diversity were often reported. While findings on alpha diversity were inconsistent (2, 25, 52, 84, 85, 87, 89, 90, 92, 96, 97, 100, 101, 108, 110, 113, 114), five studies on colorectal cancer found significant results for beta diversity (84–86, 98, 100).

Three articles had investigated associations between the gut microbiome and breast cancer risk using a case–control design. All three studies reported microbiome differences between breast cancer cases and controls, although not for the same operational taxonomic units (OTU). The Shannon index was significantly lower in breast cancer cases and the beta-diversity was significantly different between cases and controls (79, 80).

Discussion

In total, 124 studies have been included in this systematic review to evaluate the relation between the human microbiome and cancer. The gut microbiome was the microbiome most frequently studied, followed by the oral microbiome and then studies of the bile duct, cervical, and intrauterine, esophagus and gastric, laryngeal, lung, skin, breast,

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

urinary and prostate microbiome. The quality assessment demonstrated rather low quality for the "ascertainment of the exposure of interest" because studies often used different methods of ascertainment for cases and controls, for example, resection for cases while biopsy for controls. The large diversity of parameters used to describe the microbial composition made it impossible to harmonize the different studies in a way that would allow meta-analysis. All articles, apart from one, showed specific differences in microbiome distribution between cases and controls. Some findings were consistent between studies while others were conflicting or only reported in one single study (see Supplementary Table S5).

Concerning the esophageal microbiome, two studies investigated different populations with other sample types, and reported different results, though both found positive associations between *Corynebacterium* and *Peptococcus* in the mouth and esophageal cancer risk. Bacteria from the *Corynebacterium* and *Peptococcus* genera represent an important component of skin and mucosal membranes, being part of the oral and upper respiratory tract microbiome and exist through commensal interactions with their hosts (115, 116). However, some species from the *Corynebacterium* genus can act as opportunistic pathogens and secrete exotoxins associated with a number of diseases such as diphtheria, pleuropneumonia, and sepsis (115). Similarly, *Peptococcus,* a Gram-positive anaerobic coccus, has been observed to be enriched in different infections such as periodontitis and in wounds such as chronic ulcers (116). Immunocompromised patients with cancer patients might be privileged targets for these opportunistic pathogens. More research and standardized protocols for sample collection are needed to clarify this potential link between the esophageal microbiome and cancer risk.

Many bacteria can be found in the human oral cavity and these are not only associated with oral cancers but also with cancers of the lung, the esophagus, the stomach, the pancreas and the colorectum. While the oral microbiome has been suggested to play a crucial role in carcinogenesis, the oral microbial community is complex and may be impacted by many other environmental and genetic factors and potentially due to this complexity, published research has often shown inconsistent results. However, some potential pathways have been suggested for the putative role between the oral microbiome and cancer risk. For instance, the observed decrease of *Streptococcus* in precancer and oral cancer cases might reflect early changes of the oral mucosa surface related to tumorigenesis (31, 58). Indeed, *Streptococci* might lose their ability to adhere to mucosa undergoing tumorigenesis while other species, such as *Fusobacterium,* which are often increased in oral cancer cases (58) might adhere better. *Fusobacterium nucleatum* is reported to activate the nuclear translocation of NFκB leading to a proinflammatory environment. However, this translocation is inhibited by *Streptococcus* thus attenuating the proinflammatory responses induced by *Fusobacterium*. This suggests that a decrease of *Streptococcus* combined with an increase of *Fusobacterium* might play a role in oral cancers (117). In the gut, *F. nucleatum* is associated with colorectal cancer, potentially because it might increase reactive oxygen species (ROS) and IL10 production leading to an inhibition of the T cells and the antitumor immunity (118). Furthermore, some *Streptococci* can inhibit the oral colonization of certain bacteria such as *Aggregatibacter actinomycetemcomitans, Prevotella intermedia*, and *Porphyromonas gingivalis* on the oral epithelial surfaces (117). This highlights the importance of interactions between the bacteria themselves and the human mucosa (117).

The gut microbiome was analyzed in both fecal and colorectal mucosal samples. In general, fecal samples had a higher microbial diversity when compared with tissue-associated microbiomes. The results included in this review clearly demonstrate that gut microbiota composition differ between patients with colorectal cancer and noncancer. The most consistent findings were for *Fusobacterium* (with most important species *Fusobacterium nucleatum*), *Parvimonas, Porphyromonas*, and *Peptostreptococcus*, which were significantly enriched in both fecal and mucosal samples of patients with colorectal cancer compared with controls across multiple studies. Several of our reviewed articles and a recent meta-analysis (119) reported that *Fusobacterium nucleatum*, already known as an invasive and proinflammatory agent that can cause acute and chronic oral and gastrointestinal infections, was significantly enriched in both fecal and biopsy samples of colorectal cancer patients. With its unique FadA adhesin, *Fusobacterium nucleatum* adheres to, invades, and induces inflammatory and oncogenic responses to stimulate growth of colorectal cancer cells (120). Another study suggested that *Fusobacterium nucleatum* induces NFκB to promote colorectal cancer (98). Alternatively, *Fusobacterium* may act as a "passenger" that multiplies in the more favorable conditions caused by the malignant tumor rather than itself being a causal factor in colorectal cancer development (103).

Species of *Parvimonas* were consistently reported with elevated abundance in colorectal cancer tumor biopsies. *Parvimonas micra*, the only species described at the genus level, is known to cause bacteremia, abdominal abscesses, endocarditis, and other infections (95). Its tumor-promoting effect has been suggested to be associated with altered immune responses and enhanced inflammation in the gut (121).

Previously, it has been shown in malignant and primary human oral epithelial cells that *Porphyromonas gingivalis* and its membrane fraction induces upregulation of a number of genes involved in inflammation and cell proliferation and control (122); however, specific data for colorectal cancer are lacking.

A possible explanation for the higher abundance of *Peptostreptococcus* species in patients with colorectal cancer could be due specifically to the presence of *Peptostreptococcus anaerobius*. *P. anaerobius* has been shown to promote colon dysplasia in a mouse model, specifically interacting with TLR2 and TLR4 on colon cells to increase levels of reactive oxidative species, which promotes cholesterol synthesis and cell proliferation (123).

Overall, many other differences in the abundance of certain bacteria were found between cancer cases and controls, although often not consistent between studies. Differences in results could partly be due to variation in study design and methods used such as tissue sampling (different method or different site), techniques used to characterize the microbiome, or different study populations (e.g., genetic and environmental differences including nutritional intakes and habits, oral hygiene, air pollution etc.), exclusion criteria, or other unrecognized factors. Furthermore, the microbiome may differ depending on tumor stage which could also be a cause of inconsistent results if stage was not considered in the analysis.

Barriers and challenges in microbiome studies

The number of epidemiologic studies that investigate the association between the human microbiome with cancer risk is rising; however, it is challenging to obtain definitive and high quality evidence. First, it remains difficult to establish the causality of the cancer–microbe associations (2, 16). When a bacterium is enriched at a tumor site, there are other possible reasons for that association

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

besides a causal relation. For example, the microbe can take advantage of the tumor's oxygen tension or carbon sources, or find an underused nutritional niche where they can grow freely (16). It is not clear whether the carcinogenic process changes the local environment and creates new niches for microbes, or if an alteration of the microbial composition and its function contribute to carcinogenesis. As such, reverse causation is an important concern (2). Given that many of the previous studies use a case–control design, prospective studies are urgently needed to help understand the temporal nature of the association between the microbiome and cancer development.

Another challenge is that different microbes might contribute to different stages of carcinogenesis (6, 16, 17). In the initiation stage, the microbiome may promote specific genetic mutations and chronic inflammation, but may also be involved in creating other tumor-promoting environments such as the development of obesity and the metabolic syndrome (6). After the initiation phase, other microbiota could be responsible for tumor growth, angiogenesis, and metastasis (6, 16). Because cancer development is a process that continues over many years, it is possible that by the time the cancer is diagnosed, the initial causal microbes are no longer present due to the later stage tumor environment (16, 17).

It is also unlikely that many of these cancers are due to an undetected "one microbe-one disease" association (17). It is well established that *H. pylori* infection causes gastric cancer and there are nine other microorganisms that are designated by the International Agency for Research on Cancer to be carcinogenic to humans (16, 17). However, some microbes might have modest and subtle contributions to cancer development, which will be harder to detect. It is also likely that some of those modest contributions depend on the genetic background of the host, sex, and age; as well as lifestyle and other environmental exposures (2, 17).

In addition, the microbiome composition at one specific anatomic site may be related with cancer development localized elsewhere (e.g., relationship between oral microbiome and pancreatic cancer risk; refs. 24, 72, 73), making it complex to investigate causal relationships. Indeed, symbiotic bacteria, especially in the gut, are able to influence host metabolism through producing and modifying a plethora of biologically active compounds, including hormones, carbohydrates, and lipids that reach the systemic circulation. For instance, high concentrations of secondary bile acids, derived from host-produced primary bile acids, were associated with increased risk for colorectal, pancreatic, and liver cancer (124–126). Another example is the important role of indole derivatives, produced by gut bacteria from the degradation of tryptophan, such as aryl hydrocarbon receptor ligands, to maintain intestinal homeostasis, limit inflammation and thus, indirectly reduce the risk of colorectal, prostate and breast cancer (127–129). Finally, short-chain fatty acids are produced by the gut microbiome through fermentation of nondigestible fiber and starch and help maintain balance between intestinal immunity and inflammation (130, 131). As such, microbiota may influence the risk of many cancers, independent of anatomic location due to effects of microbial metabolites on inflammation, an important hallmark of cancer, and other metabolic conditions associated with cancer risk.

Finally, the study design and methods used may limit the validity, generalizability and comparability of results. Because many exogenous and intrinsic host factors influence the human microbiome (>125 factors collectively explain 18.7% of the variation in human gut microbiome composition; ref. 132), a large sample size and an important battery of information on host factors are needed to investigate potential causal relationships between the human microbiome and cancer risk with sufficient statistical power. Furthermore, restrained physical access to certain body sites limits the possibilities for large-scale sample collections. For example, to avoid invasive procedures such as biopsy, other specimen available for studies of prostate cells such as expressed prostate secretions or urine samples containing prostatic cells are often used (133, 134). In addition, different sample handling/storage strategies (e.g., room temperature storage versus immediate freezing, different collection media, etc.; refs. 135, 136), laboratory methods, sterilization procedures, sequencing strategies (e.g., 16S rRNA gene or shotgun sequencing), and various processing pipelines for the raw sequencing data, can all impact the comparability of different studies (137). Indeed, the techniques used to characterize the microbiome composition may also importantly impact the quality of the results. While bacteriological cultures (in 8 of the 124 included studies) only allow the analysis of certain specific communities of microorganisms as (e.g., <30% of gut microbiota have been cultured to date), most studies were based on sequence analysis of the 16S rRNA gene, which is sufficiently divergent to provide a phylogenetic signal and identification of unknown bacteria. More recent culture-independent techniques such as next-generation sequencing have the advantages that they are high throughput, phylogenetically characterize the microbiota components, and quantify the relative proportions of organisms present (138). In addition, for determining alpha and beta diversity, there are many different measures and statistical methods that have been used that may further limit comparability between studies. Gail and colleagues are currently working on a method to have a reference-based distance for beta-diversity to be able to combine beta diversity analyses across studies in future (139).

Multi-omics analysis (140) could potentially assist in disentangling complex relationships between the human microbiome and cancer risk by integrating the microbiome data with other omics data of interest for investigating causal factors of cancer development. However, microbiome multi-omics data analysis remains challenging as few tools exist to integrate these data. Current statistical approaches, such as correlation tests between microbiome taxa and specific host metabolites, do not necessarily meet the assumptions of the sparse compositional nature of microbiome data (141, 142).

Limitations of our review

A limitation of our review is that tumor stage and grade were not considered because of limited evidence in the studies included in this systematic review. Different stages of the cancer might harbor different microbiomes and this could be a cause of nonconsistent results across studies. Tumor severity has been shown to have an impact on microbiome composition (85). On the other hand, another study reported that the gut microbiome did not change as the cancer progressed and that any change might have occurred in the early stage of carcinogenesis (108). On the basis of these opposing findings, it would be important to further investigate whether tumor stage has an impact on the composition of the microbiome.

Another limitation is that the location of the colorectal cancer tumors was not considered although differences in microbiome composition between distal and rectal versus proximal cancers have been shown (86). In addition, microbiome composition in the colorectum can also be influenced by bowel cleansing which is not considered in every study and which was not part of our exclusion criteria.

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

The microbiome of patients from different countries was also compared with one another (see Supplementary Tables S3 and S4). The study of Allali and colleagues showed significantly higher phylogenetic diversity in tumor and adjacent tissues from patients of the United States compared with a Spanish cohort (25). So it is not certain whether gut microorganisms linked to cancer are the same across different countries and populations.

As mentioned above, we compared studies that made use of different techniques to characterize the gut microbiome. These different techniques may lead to different findings. Every study has its own exclusion criteria and although many studies matched their cases and controls for age, sex, body mass index (BMI), and other factors there are some differences in potential confounding variables across reviewed articles, which may have impact on our findings (see Supplementary Tables S3 and S4). Specific study design issues should be considered when conducting longitudinal studies for microbiome research (143). Biological sample collection, processing (DNA extraction, amplification, sequencing, and bioinformatics) and storage as well as quality control standards still require improvement. At the same time, guidelines need to be elaborated to report results and share data to ensure replication and reproducibility of these studies (143).

Finally, a limitation of this review is the lack of quantitative results from meta-analysis due to substantial heterogeneity in the microbiome parameters assessed.

Clinical implications and future perspectives

This review has shown some specific microorganisms appear to have a different prevalence in cancer cases than in controls. Tumor stage was not considered so it is not clear whether these specific microorganisms are present from the initiation of carcinogenesis. Therefore, more research, in particular via prospective studies, is needed to investigate if the microbiome could be a potential tool for cancer early detection and/or prevention strategies. Few studies already indicated that the gut microbiome, with its specific microorganisms could potentially be used as an effective tool in early screening for colorectal cancer (113).

Knowledge on microorganisms linked to cancer may also play an important role in cancer therapy. There is limited evidence suggesting that dysbiosis of the gut microbiome can be improved using probiotic and prebiotic supplementations in children with cancer (144). There is limited evidence for cancer prevention, but it is used for many other health outcomes such as improved digestion (17).

More understanding of the specific roles of the microbiome is required before promoting the microbiome as an adjuvant therapy that enhances efficacy or attenuates the toxicity of chemotherapies (6, 16). Herein, stronger evidence derived from studies with more optimal study design and methods and from different geographic regions is needed. To detect strong associations between microbiome profiles and various types of cancer, adequate power could be reached by pooling or meta-analyzing data across multiple studies. However, heterogeneity of data hinders meta-analysis. Therefore, standardization across studies is needed. Future prospective studies should coordinate and collect fecal samples using at least one common method in addition to the method of their choice. Recent studies have evaluated multiple fecal and oral sample collection methods and a variety of methods appear to be reproducible, stable, and accurate compared to an immediately frozen fecal sample. (145–147). DNA extraction and sample handling were also reported to contribute to variability in microbiome data (135, 136), it is therefore important for future studies to use standardized methods including negative and positive controls at all stages of sample processing and, when possible, use automation when processing samples. To have homogeneity of data, data analysis methods need to be transparent, consistently reported and reproducible, which still remains a major challenge in current microbiome reports. Microbiome bioinformatics platforms such as Quantitative Insights Into Microbial Ecology 2 (QIIME 2; refs. 148, 149) improve reproducibility of bioinformatics analysis by providing automated, decentralized data provenance tracking. Another bioinformatics challenge for future prospective studies is the low access to long-term archives of raw microbiome sequence data with sample and study metadata in common formats (150). Finally, future study designs should focus on the assessment of potential confounding factors by collecting high-quality clinical and nutritional data that are known to influence the microbiome.

Conclusions

Although strong evidence was available for certain taxa of the gut microbiome and colorectal cancer risk, and for the oral microbiome and oral cancer risk, for most of the microbiome taxa/indicators the evidence was still too weak to draw firm conclusions in relation to their role in cancer. In addition, based on the state of the field, it is currently not possible to have an overall ascertainment of cancer risk due to the microbiome because the data generated is heterogeneous and merging studies for meta-analysis is extremely difficult. Therefore, future prospective studies with prediagnostic specimen collection using standardized methods, consistent laboratory methodology and bioinformatics, and inclusion of quality control samples are required to establish causal links. Also "integromics" analysis (integrating microbiome data with other –omics data; e.g., metabolomics, epigenomics) may further assist in disentangling these complex relationships between the human microbiome and cancer risk although these will require significantly increased sample sizes and likely pooling across multiple cohorts.

Disclosure of Potential Conflicts of Interest

N. Michels reports grants from Research Foundation Flanders (FWO) during the conduct of the study. No potential conflicts of interest were disclosed by the other authors.

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Acknowledgments

E. Vogtmann was supported by the Intramural Research Program of the NCI at the NIH. N. Michels is financially supported by Research Foundation—Flanders 12H1519N (FWO3EO2015004301).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 22, 2020; revised May 6, 2020; accepted July 22, 2020; published first July 29, 2020.

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

References

1. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome 2015;3:31.
2. Vogtmann E, Goedert JJ. Epidemiologic studies of the human microbiome and cancer. Br J Cancer 2016;114:237–42.
3. Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. BMJ 2017;356:j831.
4. Blum HE. The human microbiome. Adv Med Sci 2017;62:414–20.
5. Gevers D, Knight R, Petrosino JF, Huang K, McGuire AL, Birren BW, et al. The Human Microbiome Project: a community resource for the healthy human microbiome. PLoS Biol 2012;10: e1001377.
6. Dzutsev A, Goldszmid RS, Viaud S, Zitvogel L, Trinchieri G. The role of the microbiota in inflammation, carcinogenesis, and cancer therapy. Eur J Immunol 2015;45:17–31.
7. Daliri EB, Tango CN, Lee BH, Oh DH. Human microbiome restoration and safety. Int J Med Microbiol 2018;308:487–97.
8. Karkman A, Lehtimaki J, Ruokolainen L. The ecology of human microbiota: dynamics and diversity in health and disease. Ann N Y Acad Sci 2017;1399:78–92.
9. Proal AD, Lindseth IA, Marshall TG. Microbe-microbe and host-microbe interactions drive microbiome dysbiosis and inflammatory processes. Discov Med 2017;23:51–60.
10. Francescone R, Hou V, Grivennikov SI. Microbiome, inflammation, and cancer. Cancer J 2014;20:181–9.
11. Fouad YA, Aanei C. Revisiting the hallmarks of cancer. Am J Cancer Res 2017;7: 1016–36.
12. Hanahan D, Weinberg RA.The hallmarks of cancer. Cell 2000;100:57–70.
13. Karlsson FH, Tremaroli V, Nookaew I, Bergstrom G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature 2013;498:99–103.
14. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BA, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature 2016;535:376–81.
15. Aron-Wisnewsky J, Prifti E, Belda E, Ichou F, Kayser BD, Dao MC, et al. Major microbiota dysbiosis in severe obesity: fate after bariatric surgery. Gut 2019;68: 70–82.
16. Garrett WS. Cancer and the microbiota. Science 2015;348:80–6.
17. Bultman SJ. The microbiome and its potential as a cancer preventive intervention. Semin Oncol 2016;43:97–106.
18. 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;25:679–89.
19. Yachida S, Mizutani S, Shiroma H, Shiba S, Nakajima T, Sakamoto T, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. Nat Med 2019;25:968–76.
20. Weng MT, Chiu YT, Wei PY, Chiang CW, Fang HL, Wei SC. Microbiota and gastrointestinal cancer. J Formos Med Assoc 2019;118:S32–S41.
21. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008–12.
22. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700.
23. The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-analyses; [about 2 screens]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp.
24. Fan X, Alekseyenko AV, Wu J, Peters BA, Jacobs EJ, Gapstur SM, et al. Human oral microbiome and prospective risk for pancreatic cancer: a population-based nested case-control study. Gut 2018;67:120–7.
25. Allali I, Delgado S, Marron PI, Astudillo A, Yeh JJ, Ghazal H, et al. Gut microbiome compositional and functional differences between tumor and non-tumor adjacent tissues from cohorts from the US and Spain. Gut Microbes 2015;6:161–72.
26. Hayes RB, Ahn J, Fan X, Peters BA, Ma Y, Yang L, et al. Association of oral microbiome with risk for incident head and neck squamous cell cancer. JAMA Oncol 2018;4:358–65.
27. Guerrero-Preston R, Godoy-Vitorino F, Jedlicka A, Rodríguez-Hilario A, González H, Bondy J, et al. 16S rRNA amplicon sequencing identifies microbiota associated with oral cancer, human papilloma virus infection and surgical treatment. Oncotarget 2016;7:51320–34.
28. Alanee S, El-Zawahry A, Dynda D, Dabaja A, McVary K, Karr M, et al. A prospective study to examine the association of the urinary and fecal microbiota with prostate cancer diagnosis after transrectal biopsy of the prostate using 16sRNA gene analysis. Prostate 2019;79:81–7.
29. Lee SH, Sung JY, Yong D, Chun J, Kim SY, Song JH, et al. Characterization of microbiome in bronchoalveolar lavage fluid of patients with lung cancer comparing with benign mass like lesions. Lung Cancer 2016;102:89–95.
30. Liu HX, Tao LL, Zhang J, Zhu YG, Zheng Y, Liu D, et al. Difference of lower airway microbiome in bilateral protected specimen brush between lung cancer patients with unilateral lobar masses and control subjects. Int J Cancer 2018; 142:769–78.
31. Henrich B, Rumming M, Sczyrba A, Velleuer E, Dietrich R, Gerlach W, et al. Mycoplasma salivarium as a dominant coloniser of Fanconi anaemia associated oral carcinoma. PLoS One 2014;9:e92297.
32. Nasrollahzadeh D, Malekzadeh R, Ploner A, Shakeri R, Sotoudeh M, Fahimi S, et al. Variations of gastric corpus microbiota are associated with early esophageal squamous cell carcinoma and squamous dysplasia. Sci Rep 2015;5:8820.
33. Chng KR, Chan SH, Ng AHQ, Li C, Jusakul A, Bertrand D, et al. Tissue microbiome profiling identifies an enrichment of specific enteric bacteria in opisthorchis viverrini associated cholangiocarcinoma. EBioMedicine 2016;8: 195–202.
34. Aviles-Jimenez F, Guitron A, Segura-Lopez F, Mendez-Tenorio A, Iwai S, Hernandez-Guerrero A, et al. Microbiota studies in the bile duct strongly suggest a role for Helicobacter pylori in extrahepatic cholangiocarcinoma. Clin Microbiol Infect 2016;22:178 e11–22.
35. Urbaniak C, Gloor GB, Brackstone M, Scott L, Tangney M, Reid G. The microbiota of breast tissue and its association with breast cancer. Appl Environ Microbiol 2016;82:5039–48.
36. Hieken TJ, Chen J, Hoskin TL, Walther-Antonio M, Johnson S, Ramaker S, et al. The microbiome of aseptically collected human breast tissue in benign and malignant disease. Sci Rep 2016;6:30751.
37. Audirac-Chalifour A, Torres-Poveda K, Bahena-Román M, Téllez-Sosa J, Martínez-Barnetche J, Cortina-Ceballos B, et al. Cervical microbiome and cytokine profile at various stages of cervical cancer: a pilot study. PLoS One 2016;11:e0153274.
38. Oh HY, Kim BS, Seo SS, Kong JS, Lee JK, Park SY, et al. The association of uterine cervical microbiota with an increased risk for cervical intraepithelial neoplasia in Korea. Clin Microbiol Infect 2015;21:674.
39. Fang RL, Chen LX, Shu WS, Yao SZ, Wang SW, Chen YQ. Barcoded sequencing reveals diverse intrauterine microbiomes in patients suffering with endometrial polyps. Am J Transl Res 2016;8:1581–92.
40. Mannell A, Plant M, Frolich J. The microflora of the oesophagus. Ann R Coll Surg Engl 1983;65:152–4.
41. Half EE, Amir I, Naftali T, Gophna U, Konikoff FM. Gastric microbiota is altered in esophagitis and Barrett's esophagus and further modified by proton pump inhibitors. Gastroenterology 2012;142:S442.
42. Aviles-Jimenez F, Vazquez-Jimenez F, Medrano-Guzman R, Mantilla A, Torres J. Stomach microbiota composition varies between patients with non-atrophic gastritis and patients with intestinal type of gastric cancer. Sci Rep 2014;4:4202.
43. Seo I, Jha BK, Suh SI, Suh MH, Baek WK. Microbial profile of the stomach: Comparison between normal mucosa and cancer tissue in the same patient. J Bacteriol Virol 2014;44:162–9.
44. Castano-Rodriguez N, Goh KL, Fock KM, Mitchell HM, Kaakoush NO. Dysbiosis of the microbiome in gastric carcinogenesis. Sci Rep 2017;7:15957.
45. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G, et al. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut 2018;67:1024–32.
46. Hu YL, Pang W, Huang Y, Zhang Y, Zhang CJ. The gastric microbiome is perturbed in advanced gastric adenocarcinoma identified through shotgun metagenomics. Front Cellular Infect Microbiol 2018;8:433.
47. Li TH, Qin Y, Sham PC, Lau KS, Chu KM, Leung WK. Alterations in gastric microbiota after H. pylori eradication and in different histological stages of gastric carcinogenesis. Sci Rep 2017;7:44935.
48. Liu X, Shao L, Liu X, Ji F, Mei Y, Cheng Y, et al. Alterations of gastric mucosal microbiota across different stomach microhabitats in a cohort of 276 patients with gastric cancer. EBioMedicine 2019;40:336–48.

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

49. Wang L, Zhou J, Xin Y, Geng C, Tian Z, Yu X, et al. Bacterial overgrowth and diversification of microbiota in gastric cancer. Eur J Gastroenterol Hepatol 2016;28:261–6.
50. Yu G, Hu N, Wang L, Wang C, Han XY, Humphry M, et al. Gastric microbiota features associated with cancer risk factors and clinical outcomes: a pilot study in gastric cardia cancer patients from Shanxi, China. Int J Cancer 2017;141:45–51.
51. Yu G, Torres J, Hu N, Medrano-Guzman R, Herrera-Goepfert R, Humphrys MS, et al. Molecular characterization of the human stomach microbiota in gastric cancer patients. Front Cell Infect Microbiol 2017; 7:302.
52. Yu J, Feng Q, Wong SH, Zhang D, Liang QY, Qin Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 2017;66:70–8.
53. Hosgood HD 3rd, Sapkota AR, Rothman N, Rohan T, Hu W, Xu J, et al. The potential role of lung microbiota in lung cancer attributed to household coal burning exposures. Environ Mol Mutagen 2014;55:643–51.
54. Cameron SJS, Lewis KE, Huws SA, Hegarty MJ, Lewis PD, Pachebat JA, et al. A pilot study using metagenomic sequencing of the sputum microbiome suggests potential bacterial biomarkers for lung cancer. PLoS One 2017;12: e0177062.
55. Carpagnano GE, Lacedonia D, Palladino GP, et al. Aspergillus spp. colonization in exhaled breath condensate of lung cancer patients from Puglia region of Italy. BMC Pulm Med 2014;14:22.
56. Tsay JJ, Wu BG, Badri MH, Clemente JC, Shen N, Meyn P, et al. Airway microbiota is associated with upregulation of the PI3K pathway in lung cancer. Am J Respir Crit Care Med 2018;198:1188–98.
57. Wang K, Huang Y, Zhang Z, Liao J, Ding Y, Fang X, et al. A preliminary study of microbiota diversity in saliva and bronchoalveolar lavage fluid from patients with primary bronchogenic carcinoma. Med Sci Monit 2019;25:2819–34.
58. Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, Queiroz ELS, et al. Changes in abundance of oral microbiota associated with oral cancer. PLoS One 2014;9:e98741.
59. Berkovits C, Toth A, Szenzenstein J, Deak T, Urban E, Gacser A, et al. Analysis of oral yeast microflora in patients with oral squamous cell carcinoma. Springerplus 2016;5:1257.
60. Homann N, Tillonen J, Meurman JH, Rintamaki H, Lindqvist C, Rautio M, et al. Increased salivary acetaldehyde levels in heavy drinkers and smokers: a microbiological approach to oral cavity cancer. Carcinogenesis 2000;21:663–8.
61. Banerjee S, Tian T, Wei Z, Peck KN, Shih N, Chalian AA, et al. Microbial signatures associated with oropharyngeal and oral squamous cell carcinomas. Sci Rep 2017;7:4036.
62. Bornigen D, Ren B, Pickard R, Li J, Ozer E, Hartmann EM, et al. Alterations in oral bacterial communities are associated with risk factors for oral and oropharyngeal cancer. Sci Rep 2017;7:17686.
63. Hsiao JR, Chang CC, Lee WT, Huang CC, Ou CY, Tsai ST, et al. The interplay between oral microbiome, lifestyle factors and genetic polymorphisms in the risk of oral squamous cell carcinoma. Carcinogenesis 2018;39:778–87.
64. Hu X, Zhang Q, Hua H, Chen F. Changes in the salivary microbiota of oral leukoplakia and oral cancer. Oral Oncol 2016;56:e6–8.
65. Lee WH, Chen HM, Yang SF, Liang C, Peng CY, Lin FM, et al. Bacterial alterations in salivary microbiota and their association in oral cancer. Sci Rep 2017;7:16540.
66. Lim Y, Fukuma N, Totsika M, Kenny L, Morrison M, Punyadeera C. The performance of an oral microbiome biomarker panel in predicting oral cavity and oropharyngeal cancers. Front Cell Infect Microbiol 2018;8:267.
67. Mok SF, Karuthan C, Cheah YK, Ngeow WC, Rosnah Z, Yap SF, et al. The oral microbiome community variations associated with normal, potentially malignant disorders and malignant lesions of the oral cavity. Malays J Pathol 2017;39: 1–15.
68. Perera M, Al-Hebshi NN, Perera I, Ipe D, Ulett GC, Speicher DJ, et al. Inflammatory bacteriome and oral squamous cell carcinoma. J Dent Res 2018;97:725–32.
69. Wolf A, Moissl-Eichinger C, Perras A, Koskinen K, Tomazic PV, Thurnher D. The salivary microbiome as an indicator of carcinogenesis in patients with oropharyngeal squamous cell carcinoma: a pilot study. Sci Rep 2017; 7:5867.
70. Yost S, Stashenko P, Choi Y, Kukuruzinska M, Genco CA, Salama A, et al. Increased virulence of the oral microbiome in oral squamous cell carcinoma revealed by metatranscriptome analyses. Int J Oral Sci 2018;10:32.
71. Zhao H, Chu M, Huang Z, Yang X, Ran S, Hu B, et al. Variations in oral microbiota associated with oral cancer. Sci Rep 2017;7:11773.
72. Farrell JJ, Zhang L, Zhou H, Chia D, Elashoff D, Akin D, et al. Variations of oral microbiota are associated with pancreatic diseases including pancreatic cancer. Gut 2012;61:582–8.
73. Torres PJ, Fletcher EM, Gibbons SM, Bouvet M, Doran KS, Kelley ST. Characterization of the salivary microbiome in patients with pancreatic cancer. PeerJ 2015;3:e1373.
74. Olson SH, Satagopan J, Xu Y, Ling L, Leong S, Orlow I, et al. The oral microbiota in patients with pancreatic cancer, patients with IPMNs, and controls: a pilot study. Cancer Causes Control 2017;28:959–69.
75. Han S, Yang X, Qi Q, Pan Y, Chen Y, Shen J, et al. Potential screening and early diagnosis method for cancer: tongue diagnosis. Int J Oncol 2016;48:2257–64.
76. Hu J, Han S, Chen Y, Ji Z. Variations of tongue coating microbiota in patients with gastric cancer. Biomed Res Int 2015;2015:173729.
77. Bi H, Tian Y, Song C, Li J, Liu T, Chen Z, et al. Urinary microbiota - a potential biomarker and therapeutic target for bladder cancer. J Med Microbiol 2019;68: 1471–8.
78. Bucevic Popovic V, Situm M, Chow CT, Chan LS, Roje B, Terzic J. The urinary microbiome associated with bladder cancer. Sci Rep 2018;8:12157.
79. Goedert JJ, Jones G, Hua X, Xu X, Yu G, Flores R, et al. Investigation of the association between the fecal microbiota and breast cancer in postmenopausal women: a population-based case-control pilot study. J Natl Cancer Inst 2015; 107:djv147.
80. Goedert JJ, Hua X, Bielecka A, Okayasu I, Milne GL, Jones GS, et al. Postmenopausal breast cancer and oestrogen associations with the IgA-coated and IgA-noncoated faecal microbiota. Br J Cancer 2018; 118:471–9.
81. Zhu J, Liao M, Yao Z, Liang W, Li Q, Liu J, et al. Breast cancer in postmenopausal women is associated with an altered gut metagenome. Microbiome 2018;6:136.
82. Qi YF, Sun JN, Ren LF, Cao XL, Dong JH, Tao K, et al. Intestinal microbiota is altered in patients with gastric cancer from Shanxi Province, China. Dig Dis Sci 2019;64:1193–203.
83. Youssef O, Lahti L, Kokkola A, Karla T, Tikkanen M, Ehsan H, et al. Stool microbiota composition differs in patients with stomach, colon, and rectal neoplasms. Dig Dis Sci 2018;63:2950–8.
84. Chen W, Liu F, Ling Z, Tong X, Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One 2012;7:e39743.
85. Mira-Pascual L, Cabrera-Rubio R, Ocon S, Costales P, Parra A, Suarez A, et al. Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol 2014;50:167–79.
86. Flemer B, Lynch DB, Brown JM, Jeffery IB, Ryan FJ, Claesson MJ, et al. Tumour-associated and non-tumour-associated microbiota in colorectal cancer. Gut 2017;66:633–43.
87. Alexander JL, Scott A, Mroz A, Perdones-Montero A, McKenzie J, Rees DN, et al. Mass spectrometry imaging (MSI) of microbiome-metabolome interactions in colorectal cancer. Gastroenterology 2016;150:S23.
88. Gao Z, Guo B, Gao R, Zhu Q, Wu W, Qin H. Probiotics modify human intestinal mucosa-associated microbiota in patients with colorectal cancer. Mol Med Rep 2015;12:6119–27.
89. McCoy AN, Araújo-Pérez F, Azcárate-Peril A, Yeh JJ, Sandler RS Keku TO. Fusobacterium is associated with colorectal adenomas. PLoS One 2013;8:e53653.
90. Burns MB, Lynch J, Starr TK, Knights D, Blekhman R. Virulence genes are a signature of the microbiome in the colorectal tumor microenvironment. Genome Med 2015;7:55.
91. Warren RL, Freeman DJ, Pleasance S, Watson P, Moore RA, Cochrane K, et al. Co-occurrence of anaerobic bacteria in colorectal carcinomas. Microbiome 2013;1:16.
92. Marchesi JR, Dutilh BE, Hall N, Peters WHM, Roelofs R, Boleij A, et al. Towards the human colorectal cancer microbiome. PLoS One 2011;6:e20447.
93. Tahara T, Yamamoto E, Suzuki H, Maruyama R, Chung W, Garriga J, et al. Fusobacterium in colonic flora and molecular features of colorectal carcinoma. Cancer Res 2014;74:1311–8.
94. Hale VL, Jeraldo P, Chen J, Mundy M, Yao J, Priya S, et al. Distinct microbes, metabolites, and ecologies define the microbiome in deficient and proficient mismatch repair colorectal cancers. Genome Med 2018;10:78.

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

95. Shah MS, DeSantis T, Yamal JM, Weir T, Ryan EP, Cope JL, et al. Re-purposing 16S rRNA gene sequence data from within case paired tumor biopsy and tumor-adjacent biopsy or fecal samples to identify microbial markers for colorectal cancer. PLoS One 2018;13:e0207002.
96. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst 2013;105:1907–11.
97. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat Commun 2015;6:6528.
98. Kasai C, Sugimoto K, Moritani I, Tanaka J, Oya Y, Inoue H, et al. Comparison of human gut microbiota in control subjects and patients with colorectal carcinoma in adenoma: terminal restriction fragment length polymorphism and next-generation sequencing analyses. Oncol Rep 2016; 35:325–33.
99. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, et al. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. PLoS One 2016;11:e0155362.
100. Wu N, Yang X, Zhang R, Li J, Xiao X, Hu Y, et al. Dysbiosis signature of fecal microbiota in colorectal cancer patients. Microb Ecol 2013;66: 462–70.
101. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 2014;10:766.
102. Allali I, Boukhatem N, Bouguenouch L, Hardi H, Boudouaya HA, Cadenas MB, et al. Gut microbiome of Moroccan colorectal cancer patients. Med Microbiol Immunol 2018;207:211–25.
103. Amitay EL, Werner S, Vital M, Pieper DH, Hofler D, Gierse IJ, et al. Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis 2017;38:781–8.
104. Baxter NT, Ruffin MTt, Rogers MA, Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med 2016;8:37.
105. Dai Z, Coker OO, Nakatsu G, Wu WKK, Zhao L, Chen Z, et al. Multi-cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. Microbiome 2018;6:70.
106. Sinha R, Ahn J, Sampson JN, Shi J, Yu G, Xiong X, et al. Fecal microbiota, fecal metabolome, and colorectal cancer interrelations. PLoS One 2016;11: e0152126.
107. Zackular JP, Rogers MA, Ruffin MTt, Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res 2014;7:1112–21.
108. Ohigashi S, Sudo K, Kobayashi D, Takahashi O, Takahashi T, Asahara T, et al. Changes of the intestinal microbiota, short chain fatty acids, and fecal pH in patients with colorectal cancer. Dig Dis Sci 2013;58:1717–26.
109. Amiot A, Dona A, Wijeyesekera A, Tournigand C, Lebaleur Y, Sobhani I, et al. 1H HR NMR spectroscopy of fecal extracts enables detection of advanced colorectal neoplasia. United European Gastroenterol J 2015;3:A266.
110. Kanazawa K, Konishi F, Mitsuoka T, Terada A, Itoh K, Narushima S, et al. Factors influencing the development of sigmoid colon cancer. Bacteriologic and biochemical studies. Cancer 1996;77:1701–6.
111. Richard ML, Liguori G, Lamas B, Brandi G, da Costa G, Hoffmann TW, et al. Mucosa-associated microbiota dysbiosis in colitis associated cancer. Gut Microbes 2018;9:131–42.
112. Mori G, Rampelli S, Orena BS, Rengucci C, De Maio G, Barbieri G, et al. Shifts of faecal microbiota during sporadic colorectal carcinogenesis. Sci Rep 2018;8: 10329.
113. Gao R, Gao Z, Huang L, Qin H. Gut microbiota and colorectal cancer. Eur J Clin Microbiol Infect Dis 2017;36:757–69.
114. Weir TL, Manter DK, Sheflin AM, Barnett BA, Heuberger AL, Ryan EP. Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PLoS One 2013;8:e70803.
115. Bratcher DF. Other corynebacteria. In: Long SS, Prober CG, Fischer M, editors. In: Principles and practice of pediatric infectious diseases. 5th ed. Philadelphia: Elsevier; 2018. p. 778–81.
116. Murphy EC, Frick IM. Gram-positive anaerobic cocci–commensals and opportunistic pathogens. FEMS Microbiol Rev 2013;37:520–53.
117. Zhang G, Chen R, Rudney JD. Streptococcus cristatus modulates the Fusobacterium nucleatum-induced epithelial interleukin-8 response through the nuclear factor-kappa B pathway. J Periodontal Res 2011;46:558–67.
118. Nosho K, Sukawa Y, Adachi Y, Ito M, Mitsuhashi K, Kurihara H, et al. Association of Fusobacterium nucleatum with immunity and molecular alterations in colorectal cancer. World J Gastroenterol 2016;22:557–66.
119. Wang D, Li Y, Zhong H, Ding Q, Lin Y, Tang S, et al. Alterations in the human gut microbiome associated with Helicobacter pylori infection. FEBS Open Bio 2019;9:1552–60.
120. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. Cell Host Microbe 2013;14:195–206.
121. Yu J, Zhao L, Zhao R, Long X, Coker OO, Sung JJY. The role of Parvimonas micra in intestinal tumorigenesis in germ-free and conventional APCmin/+ mice. J Clin Oncol 2019;37:531.
122. Groeger S, Jarzina F, Domann E, Meyle J. Porphyromonas gingivalis activates NFkappaB and MAPK pathways in human oral epithelial cells. BMC Immunol 2017;18:1.
123. Tsoi H, Chu ESH, Zhang X, Sheng J, Nakatsu G, Ng SC, et al. Peptostreptococcus anaerobius induces intracellular cholesterol biosynthesis in colon cells to induce proliferation and causes dysplasia in mice. Gastroenterology 2017; 152:1419–33.
124. Jia W, Xie G, Jia W. Bile acid-microbiota crosstalk in gastrointestinal inflammation and carcinogenesis. Nat Rev Gastroenterol Hepatol 2018; 15:111–28.
125. Xie G, Wang X, Huang F, Zhao A, Chen W, Yan J, et al. Dysregulated hepatic bile acids collaboratively promote liver carcinogenesis. Int J Cancer 2016;139: 1764–75.
126. Li P, Shu Y, Gu Y. The potential role of bacteria in pancreatic cancer: a systematic review. Carcinogenesis 2020;41:397–404.
127. Xie G, Raufman JP. Role of the Aryl hydrocarbon receptor in colon neoplasia. Cancers 2015;7:1436–46.
128. Jin UH, Lee SO, Pfent C, Safe S. The aryl hydrocarbon receptor ligand omeprazole inhibits breast cancer cell invasion and metastasis. BMC Cancer 2014;14:498.
129. Fritz WA, Lin TM, Cardiff RD, Peterson RE. The aryl hydrocarbon receptor inhibits prostate carcinogenesis in TRAMP mice. Carcinogenesis 2007;28: 497–505.
130. Descamps HC, Herrmann B, Wiredu D, Thaiss CA. The path toward using microbial metabolites as therapies. EBioMedicine 2019;44:747–54.
131. Neumann S, Peyroux EM, Woodall MJ, Shields NJ, Young SL, Pattison ST. The influence of microbial metabolites in the gastrointestinal microenvironment on anticancer immunity. Curr Cancer Treat 2019.
132. Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 2016; 352:565–9.
133. Smelov V, Arroyo Muhr LS, Bzhalava D, Brown LJ, Komyakov B, Dillner J. Metagenomic sequencing of expressed prostate secretions. J Med Virol 2014;86: 2042–8.
134. Smelov V, Bzhalava D, Arroyo Muhr LS, Eklund C, Komyakov B, Gorelov A, et al. Detection of DNA viruses in prostate cancer. Sci Rep 2016; 6:25235.
135. Choo JM, Leong LE, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. Sci Rep 2015;5:16350.
136. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol 2017;35:1069–76.
137. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol 2017;35: 1077–86.
138. Hiergeist A, Glasner J, Reischl U, Gessner A. Analyses of intestinal microbiota: culture versus sequencing. ILAR J 2015;56:228–40.
139. Maziarz M, Pfeiffer RM, Wan Y, Gail MH. Using standard microbiome reference groups to simplify beta-diversity analyses and facilitate independent validation. Bioinformatics 2018;34:3249–57.
140. Le Cao KA, Gonzalez I, Dejean S. integrOmics: an R package to unravel relationships between two omics datasets. Bioinformatics 2009; 25:2855–6.
141. Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ. Microbiome datasets are compositional: and this is not optional. Front Microbiol 2017;8: 2224.
142. Filzmoser P, Hron K. Correlation analysis for compositional data. Math Geosci 2008;41:905.
143. Sinha R, Ahsan H, Blaser M, Caporaso JG, Carmical JR, Chan AT, et al. Next steps in studying the human microbiome and health in

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

prospective studies, Bethesda, MD, May 16–17, 2017. Microbiome 2018;6:210.
144. Bai J, Behera M, Bruner DW. The gut microbiome, symptoms, and targeted interventions in children with cancer: a systematic review. Support Care Cancer 2018;26:427–39.
145. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, et al. Collecting fecal samples for microbiome analyses in epidemiology studies. Cancer Epidemiol Biomarkers Prev 2016;25:407–16.
146. Vogtmann E, Chen J, Kibriya MG, Chen Y, Islam T, Eunes M, et al. Comparison of fecal collection methods for microbiota studies in Bangladesh. Appl Environ Microbiol 2017;83:e00361–17.
147. Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, et al. Comparison of collection methods for fecal samples in microbiome studies. Am J Epidemiol 2017;185:115–23.
148. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 2019;37:852–7.
149. Minot SS, Krumm N, Greenfield NB. One codex: a sensitive and accurate data platform for genomic microbial identification. BioRxiv 027607 [Preprint]. 2015. Available from: https://doi.org/10.1101/027607.
150. Overcoming hurdles in sharing microbiome data. Nat Microbiol 2017;2: 1573.

Downloaded from https://bloodcancerdiscov.aacrjournals.org at IARC Library (WHO) on October 15, 2020. Copyright 2020 American Association for Cancer Research.

Chapter III: Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies

1. Background

While the gut microbiome has been hypothesised to influence insulin resistance and chronic inflammation development (93,94,107,108), human data on insulin resistance and other metabolic parameters and the gut microbiome are still limited as very few cohorts have collected both blood and fecal specimens. Insulin resistance and inflammation lie on the causal pathways of diabetes and other metabolic diseases, and are established risk factors for colorectal cancer (118–120), and therefore, deciphering the pathophysiology of insulin resistance and inflammation using population-based data might help to understand the mechanisms of colorectal cancer development and strengthen evidence for potential preventive strategies.

2. Objective

The aim of this study was to investigate the association of gut bacteria 16S rRNA gene sequence data with serologic levels of homeostatic model for insulin resistance (HOMA-IR), glycated haemoglobin (HbA1c) and C-reactive protein (CRP) in two distinct population-based cohorts, the Northern Finland Birth Cohort 1966 (NFBC) and TwinsUK.

3. Methods

Study design and samples

This study was carried out among 506 participants from the NFBC and 1,140 participants from TwinsUK. In the NFBC, at age 46 years, anthropometric measures, including body mass index (BMI), and lifestyle information were collected using self-administered questionnaires and clinical examinations. All the metabolic markers were measured during clinical examination after an overnight fasting period of 12 hours. Additionally, at 46 years of age, the participants were asked to collect a fecal sample at home. In TwinsUK, height and weight were measured during each participant's annual clinic visit, allowing BMI to be calculated. Smoking status was obtained by

questionnaires. Glucose and insulin levels were measured for all individuals after a 10 hours overnight fast. TwinsUK participants were also asked to collect a fecal sample at home and bring it to the clinical research facility. In both cohorts, HOMA-IR was calculated based on fasting insulin and fasting glucose levels. The gut microbiome was characterized using the same procedure in both cohorts. DNA was extracted using MO-BIO PowerSoil® kit, and the V4 region of the 16S rRNA gene was amplified and sequenced using the MiSeq. The sequence quality control was performed with QIIME2 and DADA2, and diversity metrics calculated using R.

Statistical analysis

We estimated associations between gut microbiome diversity metrics and serologic levels of HOMA-IR, glycated haemoglobin HbA1c and C-reactive protein CRP using multivariable linear regression models adjusted for gender, smoking status and BMI. We used general linear models with a quasi-Poisson distribution and Microbiome Regression-based Kernel Association Tests (MiRKAT) to estimate associations of metabolic parameters with alpha- and beta-diversity metrics, respectively, and generalized additive models for location scale and shape (GAMLSS) fitted with the zero-inflated beta distribution to identify taxa associated with the metabolic markers.

4. Results

Overall, increased levels of HOMA-IR and CRP were associated with decreasing richness or diversity of microbiome composition in the two cohorts. Most importantly, the relationship between high levels of metabolic parameters and lower microbiome diversity was still significant, even after adjustment for BMI. In addition, in the NFBC, the association of BMI and microbiome diversity was attenuated when adjusting the model for insulin resistance. Finally, the beta component of the zero-inflated beta regression model indicated associations between HOMA-IR, CRP and HbA1c with several bacteria at the genus level.

5. Conclusion

Our findings indicate, in two distinct population-based cohorts, that individuals with worse metabolic control have lower gut microbial diversity and may provide insight into the pathophysiological mechanisms that underlie poor metabolic health. However, due to the cross-sectional nature of this study, it is impossible to conclude if these

differences in microbial composition and taxa associations are a cause or a consequence of metabolic dysfunction.

6. Scientific article

Zouiouich S, Loftfield E, Huybrechts I, Viallon V, Louca P, Vogtmann E, Wells P, Steves C, Herzig KH, Menni C, Jarvelin MR, Sinha R, Gunter MJ (**2020**) Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies. *Diabetologia*. (**Provisionally accepted - Under revision**)

Role: First author, contribution to analytical strategy, conducted analysis in NFBC data and provided code for TwinsUK data, writing original draft and submission, writing revised version.

Markers of metabolic health and gut microbiome diversity: findings from two population-based cohort studies

Semi Zouiouich[1] (zouiouichs@students.iarc.fr), Erikka Loftfield[2] (erikka.loftfield@nih.gov), Inge Huybrechts[1] (huybrechtsi@iarc.fr), Vivian Viallon[1] (viallonv@iarc.fr), Panayiotis Louca[3] (panayiotis.louca@kcl.ac.uk), Emily Vogtmann[2] (emily.vogtmann@nih.gov), Philippa M Wells[3] (philippa.wells@kcl.ac.uk), Claire J Steves[3] (claire.j.steves@kcl.ac.uk), Karl-Heinz Herzig[4] (karl-heinz.herzig@oulu.fi), Cristina Menni[3] (cristina.menni@kcl.ac.uk), Marjo-Riitta Jarvelin[5,6,7,8] (m.jarvelin@imperial.ac.uk), Rashmi Sinha[2] (sinhar@exchange.nih.gov), Marc J Gunter[1] (gunterm@iarc.fr)

Affiliations: [1]Nutrition and Metabolism Section, International Agency for Research on Cancer-WHO; [2]Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute; [3]Department of Twin Research, King's College London, London, UK; [4]Research Unit of Biomedicine, Medical Research Center (MRC), University of Oulu, University Hospital, Oulu, Finland and Department of Gastroenterology and Metabolism, Poznan University of Medical Sciences, Poznan, Poland; [5]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; [6]Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland; [7]Unit of Primary Health Care, Oulu University Hospital, OYS, Oulu, Finland; [8]Department of Life Sciences, College of Health and Life Sciences, Brunel University London, London, United Kingdom

Corresponding author: Semi Zouiouich; 150 cours Albert Thomas, 69372, Lyon CEDEX 08, France; zouiouichs@students.iarc.fr

Word count: 3,769

Abstract

Aims/hypothesis: The gut microbiome is hypothesized to influence development of insulin resistance and other metabolic parameters, however, data from population-based studies are limited. We investigated associations between serologic measures of metabolic health and the gut microbiome in the Northern Finland Birth Cohort 1966 (NFBC1966) and in TwinsUK.

Methods: Among 506 individuals from the NFBC1966, with available fecal microbiome (16S rRNA gene sequence) data, we estimated associations between gut microbiome diversity metrics and serologic levels of homeostatic model for insulin resistance (HOMA-IR), glycated haemoglobin (HbA1c) and C-reactive protein (CRP) using multivariable linear regression models adjusted for gender, smoking status and body mass index (BMI). We used general linear models with a quasi-Poisson distribution and Microbiome Regression-based Kernel Association Tests (MiRKAT) to estimate associations of metabolic parameters with alpha- and beta-diversity metrics, respectively, and generalized additive models for location scale and shape (GAMLSS) fitted with the zero-inflated beta distribution to identify taxa associated with the metabolic markers. Associations between gut microbiome diversity and HOMA-IR and CRP were also investigated in 1,140 adult participants from TwinsUK.

Results: In the NFBC1966, alpha diversity was significantly lower in individuals with higher HOMA-IR with an average of 76.5 (standard deviation [SD] 18.9) sequence variants (ASVs) for the first quartile of HOMA-IR and 65.9 (SD 18.4) for the fourth quartile of HOMA-IR. Alpha-diversity was also lower with higher HbA1c (number of ASVs and Shannon's diversity, P-values <0.001 and =0.003 respectively) and higher CRP (number of ASVs, P-values=0.025), even after control for BMI and other potential confounders. In TwinsUK, alpha-diversity measures were also significantly lower among participants with higher measures of HOMA-IR and CRP. When considering beta-diversity, we found that microbial community profiles were associated with HOMA-IR in the NFBC1966 and TwinsUK, using multivariate MiRKAT models, with binomial deviance dissimilarity P-values of <0.001.

Conclusions/interpretation: Overall, higher levels of HOMA-IR, CRP and HbA1c were associated with statistically significantly lower microbiome diversity in both the NFBC1966 and TwinsUK, even after control for BMI and other parameters. These results from two distinct population-based cohorts provide evidence that individuals with worse metabolic control have lower gut microbial diversity and may provide insight into the pathophysiological mechanisms that underlie poor metabolic health.

Key words: Insulin resistance, metabolic health, HOMA-IR, fecal microbiome

Research in context

What is already known about this subject?

- Insulin resistance lies on the causal pathways of diabetes and other metabolic diseases, however, the pathophysiological origins of insulin resistance are not fully understood.
- The gut microbiome has been hypothesized to influence insulin resistance development. However, human data on insulin resistance and the gut microbiome are limited as very few population-based cohorts have collected both blood and fecal specimens.

What is the key question?

- What is the relationship between the gut microbiome and serologic measures of metabolic health in population-based studies?

What are the new findings?

- Higher levels of insulin resistance and other markers of metabolic dysfunction were associated with lower richness and/or evenness of microbiome composition in two population-based cohort studies, even after control for obesity.

How might this impact on clinical practice in the foreseeable future?

- Modification of the gut microbiota among individuals with worsening insulin senstivity and metabolic control may offer a potential preventive strategy for Type 2 diabetes.

Introduction

Obesity and Type 2 diabetes (T2D) have reached global epidemic proportions and are recognised as major causes of morbidity and mortality (97,189). Insulin resistance is a pathophysiological condition that precedes the development of T2D (98), however its aetiology is not fully understood. Obesity is a recognised risk factor for insulin resistance but not all insulin resistant individuals are overweight or obese (99); indeed, the existence of metabolically-healthy obese and metabolically unhealthy normal-weight individuals has been described in previous studies (100). Despite the well-established epidemiological links between insulin resistance, its related parameters (i.e. poor glucose control and high levels of inflammation), and chronic diseases including obesity and T2D, the potential role of the gut microbiome in the development of insulin resistance and T2D is not fully understood.

Previous studies have suggested a link between the gut microbiome and metabolic health (94,107) and in particular, have described differences in microbial composition and functionality in T2D patients compared to healthy participants, suggesting that the microbiome may play a key role in the pathogenesis of the disease (93,94). The gut microbiome might influence host metabolic health through several pathways including energy extraction, intestinal barrier integrity, metabolism of bile acids and host metabolic and signalling pathways, which are directly or indirectly related to insulin resistance development (109). For example, circulating levels of branched-chain amino acids (BCAA) have been positively associated with insulin resistance (111) and BCAA levels in insulin-resistant individuals correlate with specific changes in gut microbiome composition and functions, such as enriched potential for BCAA biosynthesis, and deprivation of genes encoding for BCAA transport into bacterial cells (112).

Despite an important expansion of research on the role of the human gut microbiome and metabolic disorders in recent years, human data on insulin resistance from population-based studies are lacking. In this study, we investigated the relationship between measures of metabolic health (i.e., insulin resistance, glucose control and inflammation) and gut microbial diversity independent to BMI, among individuals in the Northern Finland Birth Cohort (NFBC1966) and then sought to replicate these findings in the TwinsUK cohort.

Methods

Ethics approval and consent to participate: Informed written consent was obtained from all NFBC1966 participants, and the research protocols were approved by the Ethics Committee of Northern Ostrobotnia Hospital District, Finland. A written informed consent was obtained from all the TwinUK volunteers upon registration and also during their clinical visits, and the research protocols had ethical approval as parts of the Twins UK (EC04/015) study from Local Research Ethics Committee at the Department of Twin Research and Genetic Epidemiology, King's College London. The study was approved by the International Agency for Research on Cancer Ethic Committee.

Northern Finland Birth Cohort

Study design and samples: The NFBC1966 was established in 1965 and included 12,055 mothers, with expected date of delivery between 1st of January to 31st of December 1966, and subsequently 12,058 children representing 96% of live births in two Finnish provinces, Oulu and Lapland (https://www.oulu.fi/NFBC1966/) (190). Pregnancies were followed prospectively and children were followed through childhood, adolescence and early adulthood up to 46 years of age. At 31 and 46 years of age, information about health, behaviour, work and social background were collected using self-administered questionnaires and clinical examinations were performed. At 31 and 46 years of age, measures of metabolic health were generated using blood samples. Additionally, at 46 years of age, the participants were asked to collect a stool sample at home using a collection tube without additive that was provided by study investigators. Participants were asked to return the sample to the study facility the same day; if the specimen was taken 1-2 days in advance, the participants were asked to store the stool sample at -20°C. At the study laboratory, the stool samples were first transferred to -20°C freezers and after, to long-term -70°C freezers without any additive within days of collection. Written informed consent was obtained from all participants either via postal inquiry or clinical examination at 31-year and 46-year time points. Ethical approval for the NFBC1966 project was obtained from the Ethical Committee of the Medical Faculty of University of Oulu and Northern Ostrobothnia Hospital District.

This investigation builds upon a related study conducted by Loftfield et al within the NFBC1966, which aimed to explore associations of body mass index (BMI) history and adult BMI with fecal microbial diversity (n=565) and microbial metabolite levels (n=340) (191). In the current analysis, participants were selected from the same cohort with fecal microbial diversity data and were included if measures of metabolic health (i.e. insulin [μU/L], glucose [mmol/L], high sensitivity C-reactive protein [mg/L], and glycated haemoglobin [mmol/mol]) were available at 46 years of age. In total, 506 participants were included in our analytic sample, including 187 males (37.0%) and 319 females (63.0%).

At age 46 years, participants had a clinical examination, during which height and weight were measured, and they completed self-administered questionnaires reporting lifestyle and demographic characteristics. BMI was estimated as measured weight (kg) divided by measured height (m) squared (kg/m^2). BMI was categorized according to the WHO International Classification system: normal weight (18.5–24.9), overweight/pre-obese (25.0–29.9), obese (≥ 30.0). Smoking status was categorized into three groups: never smoker, former smoker or current smoker.

Laboratory methods: The laboratory methods have previously been described in detail (192). Metabolic markers were measured for all individuals after an overnight fasting period (12 hours). Fasting plasma glucose was analysed by an enzymatic dehydrogenase method (Advia 1800, Siemens Healthcare Diagnostics, Tarrytown, New York, USA). Fasting serum insulin was analysed by a chemiluminometric immunoassay (Advia Centaur XP, Siemens Healthcare Diagnostics, Tarrytown, New York, USA). Homeostasis model assessment for insulin resistance (HOMA-IR) is a commonly used measure of insulin resistance and is calculated as fasting insulin [μU/mL] x fasting glucose [mmol/L] /22.5 (192). HOMA-IR values were subdivided into quartiles based on its distribution across all included participants: Q_1: <1.20, Q_2: 1.20-1.91, Q_3: 1.91-3.1, Q_4: >3.11 for categorical analyses. The concentration of glycated haemoglobin (HbA1c) were measured using immunochemical assay methods. High-sensitivity C-reactive protein (CRP) was analysed by an immune-nephelometric assay (BN ProSpec, Siemens Healthcare Diagnostics, Newark, Delaware, USA). HOMA-IR, HbA1c and CRP were log-transformed for numerical analyses.

DNA extraction, amplification and sequencing: Fecal samples were processed at the University of California, San Diego (La Jolla, California). DNA extraction, PCR amplification, and sequencing were completed as described by Loftfield et al (193) , and previously by Vogtmann et al (194), using the universal bacterial primer set 515F/806R (195). In brief, study samples were randomly ordered and distributed within the batches. For technical reproducibility, replicate fecal samples from three individuals (n=62) were distributed within and across batches. Four quality control (QC) samples were also included within each DNA extraction batch: artificial community, chemostat community (196), extraction blank and PCR blank. DNA was extracted with the MO-BIO PowerSoil® DNA isolation kit and the V4 region of the 16S rRNA gene was PCR amplified and 2 × 150 bp paired-end sequencing was performed on the MiSeq (Illumina, San Diego, CA). After removing singletons and reads with read errors, the average coverage was approximately 112,000 reads per sample.

Bioinformatics: Bioinformatics processing was performed using QIIME 2 2017.8 (197). Sequences were demultiplexed, and quality control on forward reads was performed with DADA2 (23). Paired end reads were not joined, because shorter 16S rRNA gene sequences would be dropped because they cannot be joined with 150 base reads, resulting in systematic bias in community composition. Taxonomy was assigned to amplicon sequence variants (ASVs) using q2-feature-classifier (198) and the Greengenes 13_8 reference database (199). A phylogenetic tree was built by aligning ASVs with MAFFT (200), filtering highly variable positions via q2-alignment, and applying FastTree (201) to construct an unrooted tree, followed by midpoint rooting using q2-phylogeny midpoint-root. Diversity metrics (i.e Shannon index, observed sequence variants, binomial deviance dissimilarity, Jaccard, weighted and unweighted UniFrac) were computed using the “vegan” (202) and “microbiome” packages in R (verion 3.6.0) at a depth of 10,000 reads per sample to represent the diversity of unique sequences in each sample. As described by Loftfield et al, inspection of QC data suggested good reproducibility within and across batches (193).

TwinsUK

Study design and samples: The TwinsUK cohort is one of the largest adult twin registries in the world comprising over 14,000 volunteers (203). In the current analysis

we analysed data from 1,140 female participants with data on the gut microbiome and both HOMA-IR and CRP.

Anthropometric and lifestyle**:** Height and weight were measured during each participant's annual clinic visit, allowing BMI to be calculated. Smoking status was obtained by questionnaires.

Laboratory methods: Plasma glucose and insulin were measured for all individuals after a 10 hours overnight fast. Immunoassay (Abbott Laboratories, Maidenhead, U.K.) quantified insulin and glucose was measured by Ektachem 700 multichannel analyzer using an enzymatic colorimetric slide assay (Johnson and Johnson Clinical Diagnostic Systems, Amersham, U.K), as previously described (204) and HOMA-IR was calculated. High sensitive CRP was measured by latex-enhanced nephelometry on a Siemens Prospec Nephelometer.

Microbiome: Fecal samples were collected at home and brought or sent on ice to the clinical research facility where they were stored at -80°C. Samples were then processed to determine gut microbial composition by 16S rRNA gene sequencing, as previously described (205). Briefly, the V4 variable region of the 16S rRNA was amplified and sequenced using a multiplexed approach on the Illumina MiSeq. 16S sequences were demultiplexed in QIIME. Amplicon sequencing variants were then generated using the DADA2 package in R using the pipeline as described elsewhere (206).

Statistical analyses: Descriptive characteristics of the participants (i.e. measures of metabolic health and covariates) were presented by BMI groups. Pearson and Spearman correlations were calculated to estimate correlations between the different measures of metabolic health and BMI. To estimate the association of the measures of metabolic health as categorical and numerical variables with the alpha-diversity metrics, general linear models with a quasi-Poisson distribution were used. For the association between microbial community profiles using beta-diversity metrics and the measures of metabolic health, MiRKAT tests were calculated (207). To identify the taxa associated with HOMA-IR, CRP and HbA1c, after grouping ASVs at the genus level, a generalized additive model for location scale and shape (GAMLSS) fitted with

the zero-inflated beta distribution (BEZI) was computed using the “gamlss” package in R (208). The GAMLSS-BEZI model is a mixture model with two components including a zero model accounting for excess zeros and a count model to capture the remaining component by beta regression, allowing for overdispersion effects. The first component of this mixture model is linked by the nu parameter that models the probability at zero, while the second component is indexed by the mu and sigma parameters, respectively the mean and precision parameters. Likelihood-ratio tests between models including adjustment factors only and models including the measures of metabolic health and adjustment factors were performed. P-values < 0.05 were considered as indicators of the effect of the analysed measures of metabolic health in affecting the distribution of the taxonomic component. The p-values were then corrected using Benjamini-Hochberg false discovery rate (FDR) < 0.05. All statistical tests described above were adjusted for BMI, gender and smoking status in the NFBC1966 and for BMI, age and smoking status in TwinsUK. Other confounding factors such as alcohol consumption and physical activity were considered but not selected.

Results

As presented in Table 1, the majority of NFBC1966 participants were female, and 29.2%, 45.9% and 24.9% were normal weight, overweight and obese, respectively. The TwinsUK cohort was entirely female and older compared to the NFBC1966 (mean age 62.9 and 46.6, respectively). In the NFBC1966, 31.9% of currents smokers were obese against 16.5% in TwinsUK. In the NFBC1966, BMI was strongly correlated with HOMA-IR (R=0.64, 95%CI [0.58; 0.68]) and more moderately correlated with CRP (R=0.37, 95%CI [0.29; 0.44]) and HbA1c (R=0.23, 95%CI [0.14; 0.31]) (Figure S1). Compared to NFBC1966, in TwinsUK, BMI was moderately correlated with HOMA-IR (R=0.43, 95%CI [0.38; 0.47]) and CRP (R=0.37, 95%CI [0.31; 0.41]).

Prior to conducting multivariable analyses, variance inflation factors (VIF) were calculated to evaluate potential multicollinearity between alpha diversity metrics, BMI and HOMA; moderate VIF were found with these two covariates (VIF BMI 1.37, VIF HOMA-IR 1.43), suggesting that these predictors are not correlated with other variables. In NFBC1966 multivariable models that adjusted for BMI, gender and smoking status, alpha diversity was significantly reduced with increasing quartiles of HOMA-IR (Table 2). For example, Shannon’s diversity was lower in the 4th quartile of

HOMA-IR (2.59, 95%CI [2.50; 2.68]) compared to the 1st quartile (2.80, 95%CI [2.71; 2.89]). Similar results were found in the TwinsUK cohort where Shannon's diversity was significantly lower in the 4th quartile of HOMA-IR (3.72, 95%CI [3.65; 3.79]) compared to the 1st quartile (3.91, 95%CI [3.84; 3.98]). Quasi-Poisson models in NFBC1966 showed significant inverse associations of alpha diversity with numerical HOMA-IR, CRP and HbA1c levels (Table 2). In the TwinsUK cohort, Shannon's diversity was significantly reduced with increasing HOMA-IR (Estimate=-0.062, P-value<0.001) and CRP (Estimate=-0.018, P-value= 0.031) as well as the number of observed ASVs. In the NFBC1966, in multivariable models including numerical HOMA-IR or CRP and BMI, the association between BMI and the number of observed ASVs was no longer significant after adjustment (P-value without HOMA-IR: 0.001, P-value in model including HOMA-IR: 0.489, P-value in model including CRP: 0.082) (Table S1). In multivariable models including HbA1c and BMI, the association between BMI and the number of observed ASVs remained significant. In the TwinsUK cohort, BMI remained strongly inversely associated with alpha diversity, regardless of the presence of HOMA-IR or CRP in the model (Table S1).

In the NFBC1966, using the first three PCoA vectors from four beta diversity matrices, no visual clustering was observed by measures of metabolic health (results not shown). In the NFBC1966, when considering community composition using measures of beta-diversity, significant differences were detected by numerical HOMA-IR, CRP and HbA1c for three measures (e.g. binomial, Jaccard, Unweighted UniFrac) in adjusted MiRKAT models (Table 3); however, no association was seen for weighted UniFrac. In the TwinsUK cohort, similar results were found with significant differences detected by numerical HOMA-IR for three measures (e.g. binomial, Jaccard, Unweighted UniFrac) but not for weighted UniFrac. However, by numerical CRP, a significant difference was observed only for Jaccard. In MiRKAT models including BMI with adjustment only for gender and smoking status in the NFBC1966 and for age and smoking status in TwinsUK, the associations with BMI were significant for binomial, Jaccard and unweighted UniFrac in the NFBC1966 and for the four measures in TwinsUK (Table S2). In the NFBC1966, after inclusion of HOMA-IR in the models, the associations between BMI and microbial composition for binomial and unweighted UniFrac metrics were were no longer statistically significant. In the TwinsUK cohort,

BMI remained strongly associated with four beta diversity matrices, regardless of the inclusion of HOMA-IR or CRP in the model (Table S2).

In GAMLSS models adjusted for BMI, gender and smoking status, sixteen individual genera met FDR-adjusted statistical significance with HOMA-IR, ten with CRP and nine with HbA1c in the NFBC1966. Higher values of HOMA-IR were strongly associated with higher average relative abundances of members from the Peptococcaceae (Estimate=11.62, P-value<0.001), Bifidobacteriaceae (genus *Gardnerella*) (Estimate=2.64, P-value<0.001), Veillonellaceae, Lachnospiraceae (genus *Blautia*) and Paraprevotellaceae families and lower average relative abundances of members from the Peptostreptococcaceae (Estimate=-0.52, P-value<0.001), Peptococcaceae (Estimate=-0.23, P-value<0.001), and Prevotellaceae families (Table S3a). Increased CRP was associated with lower average relative abundances of members from the Paraprevotellaceae, Peptococcaceae, Veillonellaceae and Peptostreptococcaceae families and higher average relative abundance of one ASVs, *Peptococcus* (Table S3b). Higher levels of HbA_{1c} were associated with lower average relative abundances of members from the Bifidobacteriaceae, Peptococcaceae, Veillonellaceae and Peptostreptococcaceae families and higher average relative abundance of one ASVs, *Oxalobacter* (Table S3c). For the presence/absence of taxa, the zero models accounting for excess zeros showed that the metabolic biomarkers-taxa associations were never significant except for the genus *Gardneralla* and HbA1c (Estimate=3.87, P-value=0.013).

Although the NFBC1966 results presented in this paper include both men and women, sensitivity analyses were run including women only (n=319) with similar results (data not shown).

Discussion

In this analysis of gut microbiome profiles and metabolic parameters in two population-based cohorts, as levels of HOMA-IR, CRP and HbA1c increased, we observed decreasing richness and/or evenness of the gut microbial taxa, even after adjustment for BMI. In addition, HOMA-IR, CRP and HbA1c were associated with microbial community as indicated by significant associations with three beta-diversity measures, but were not associated with weighted UniFrac, suggesting that there were differences

in community composition when ignoring phylogenetic similarity, but when this was accounted for, the communities were quite similar. Further, HOMA-IR, CRP and HbA1c were significantly associated with the relative abundance of specific taxa, including members of the Peptococcaceae, Bifidobacteriaceae and Veillonellaceae families, but not with their presence or absence. In the NFBC1966, we observed that when including HOMA-IR or CRP in the models, the association between BMI and some measures of alpha- and beta-diversity lost statistical significance. These findings support a potential relationship between insulin resistance and markers of metabolic health and gut microbial diversity that may be independent of BMI.

Our data support the hypothesis that greater gut microbiome diversity is associated with better insulin sensitivity (107). A clinical trial studying the effect of fecal transplant from lean donors to men diagnosed with metabolic syndrome found that insulin sensitivity had improved and gut bacterial diversity had increased 6 weeks post transplantation (209). These results corroborate other studies identifying gut microbial alterations and attenuation of metabolic syndrome after various weight-loss interventions (210,211). Perturbation of homeostatic interactions between the gut microbiome and the host might promote metabolic disturbances but the mechanisms underlying this relationship are not well understood (212) (Figure 1). As described previously, the gut microbiome can trigger inflammatory processes associated with obesity and insulin resistance by stimulating immune cells through lipopolysaccharides derived from bacterial membranes (213,214). Furthermore, microbial-derived short-chain fatty acids, including butyrate (215), can enhance insulin sensitivity (216) and suppress insulin-mediated fat accumulation (217). CRP is a marker of chronic low-grade systemic inflammation associated with obesity and insulin resistance (218). Previous studies have found a relationship between systematic inflammation, through high-sensitivity CRP plasma levels, and microbial composition, supporting our findings (219). Our results suggest that higher levels of HbA1c were associated with a decreased bacterial diversity. Results from the literature seem less clear regarding the association between HbA1c level and gut microbiome composition. Some articles reported evidence of an association between HbA1c level and bacterial group counts (220) while others did not find any significant relationship (221). These findings could be explained by the fact that elevated insulin levels or HOMA-IR appeared to identify

certain traits of metabolic syndrome changes, especially abdominal obesity, earlier than both HbA1c and measures of glucose (222).

Consistent with other studies, BMI was inversely associated with gut microbiome diversity (193). However, following adjustment for HOMA-IR and CRP, this association lost statistical significance with some measures of alpha- and beta-diversity, suggesting that the observed relationship between BMI and gut microbiome diversity might be indirect and influenced by measures of insulin resistance and inflammation. This hypothesis needs to be further explored in other cohorts as results from TwinsUK suggested that associations between BMI and gut microbial diversity were attenuated but remained strongly significant.

Sixteen individual taxa met FDR-adjusted statistical significance with HOMA-IR, ten with CRP and nine with HbA1c. In particular, higher HOMA-IR was associated with higher average relative abundances of members from the Peptococcaceae, Bifidobacteriaceae (genus *Gardnerella*) and Veillonellaceae families while higher CRP levels were associated with lower relative abundances of Paraprevotellaceae, Peptococcaceae and Peptostreptococcaceae. Higher levels of HbA1c were associated with higher presence of the family Bifidobacteriaceae (genus *Gardnerella*). In a previous study, Schott et al. found that the relative abundance of certain pro-inflammatory bacteria, including members from the Peptococcaceae (*rc4-4 sp* genus) and Peptostreptococcaceae, was higher in obese mice (223). Our findings support an association between HOMA-IR and Peptococcaceae *rc4-4 sp* but also show differences in CRP-taxa associations. Insulin-resistant individuals had a significantly higher abundance of Veillonellaceae, which support our findings (224). Relative abundance of succinate-producing Veillonellaceae was also found to be higher in obese individuals (225). No previous studies were found reporting the positive association between HOMA-IR and *Gardnerella* or the inverse association between HbA1c and *Gardnerella* which were found in our study (Tables S1a and S1c). Overall, the GAMLSS models showed that metabolic biomarkers were more associated with the relative abundance of the taxa rather than their presence or absence.

A limitation of our study was the cross-sectional nature of the analyses. Repeated, prospectively collected samples would be needed to study the causal relationships

between gut microbiome and markers of metabolic health and to assess their respective association with risk of diseases such as T2D or cancer. Other limitations of the current study were that TwinsUK was only females, even if sensitivity analyses including females only in the NFBC1966 showed similar results, and that data on HbA1c were not available for the replication. Differences between NFBC1966 and TwinsUK populations have provided interesting insight to explore the association between microbiome and metabolic health. Indeed, a previous study has shown that associations between the microbiome and various risk factors varied by geographic location, it is a potential limitation that some associations may differ in other populations (226). Therefore, being able to see associations between gut microbiome composition and measures of metabolic health in two different populations showed consistency, independently of other potential factors. Finally, we recognize the limited taxonomic resolution and accuracy of the 16S rRNA gene sequencing methodology, preventing us to adequately perform species-level associative analysis and to unambiguously discriminate low-abundance taxa from noise and further replication and extension (e.g., metagenomics) of results in other population-based studies is warranted.

Our study suggests that general population cohorts are valuable in identifying potential associations between microbial features and measures of metabolic health, particularly with low contrasts in exposure and disease. However, the use of cross-sectional data does not allow us to causally interpret these associations, since interactions between metabolic biomarkers and the gut microbiome are complex and dynamic, and can be strongly affected by behavioural changes. Therefore, it is now critical for future studies to collect longitudinal data of both lifestyle exposures and the microbiome to help understand the dynamic relationship between the gut microbiome and host metabolism. Integration of microbiome data paired with fecal metabolomics data will provide a more complete picture of the metabolomics-microbial mechanisms that contribute to metabolic balance between the host and the gut microbiome.

In conclusion, insulin resistance, poor control of blood glucose levels and systemic inflammation were associated with statistically significantly lower gut microbiome diversity and distinct microbial community structures in both the NFBC1966 and TwinsUK, even after control for BMI and other parameters. These results from two

distinct population-based cohorts provide evidence that individuals with worse metabolic control have lower gut microbial diversity and may provide insight into the pathophysiological mechanisms that underlie poor metabolic health. It is, however, impossible to conclude if these differences in microbial composition and taxa associations are a cause or a consequence of metabolic dysfunction. Thus, large-scale, prospective studies with collection of fecal samples and longitudinal data on lifestyle and metabolic biomarkers at several time-points are now needed to validate and extend these observations.

Acknowledgements: Not applicable

Data availability: The datasets used and/or statistical analysis code for the current study are available from the authors (Marc Gunter, gunterm@students.iarc.fr) on reasonable request.

Funding: This work was supported by the International Agency for Research on Cancer, WHO, France and the National Cancer Institute, USA.

Authors' relationships and activities: Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Contribution statement: SZ, IH, MJG planned the analyses. SZ analysed the NFBC1966 data. PW and CJS prepared the ASVs data in the TwinsUK cohort. PL analysed the TwinsUK data. VV, EL, EV provided support with the NFBC1966 data analysis. CM provided support with the TwinsUK data analysis. KHH, MRJ contributed to consensus discussions regarding the draft. IH, RS, MJG supervised the study. SZ wrote the manuscript. All authors read and approved the final manuscript.

Tables

Table 1 - Description of the study population stratified by BMI category in NFBC1966 and TwinsUK

	NFBC1966 (n=506) BMI, kg/m^2			TwinsUK (n=1140) BMI, kg/m^2		
	18.5 to 25	**25 to 30**	**over 30**	**18.5 to 25**	**25 to 30**	**over 30**
n (%)	148 (29.2)	232 (45.9)	126 (24.9)	499 (43.7)	411 (36.1)	230 (20.2)
Male, n (%)	32 (17.1)	108 (57.8)	47 (25.1)	0	0	0
Female, n (%)	116 (36.4)	124 (38.9)	79 (24.7)	499 (43.7)	411 (36.1)	230 (20.2)
Age, mean (SD)	46.6 (0.6)	46.6 (0.5)	46.7 (0.6)	62.9 (8.9)	64.2 (8.5)	61.8 (8.4)
Never smoker, n (%)	98 (33.9)	128 (44.3)	63 (21.8)	300 (43.7)	235 (34.2)	152 (22.1)
Former smoker, n (%)	29 (23.6)	61 (49.6)	33 (26.8)	152 (44.9)	127 (37.6)	59 (17.5)
Current smoker, n (%)	21 (22.3)	43 (45.8)	30 (31.9)	47 (40.8)	49 (42.6)	19 (16.5)
HOMA-IR, mean (SD)	1.32 (0.7)	2.26 (1.3)	4.61 (4.4)	0.84 (1.32)	1.12 (1.46)	1.56 (1.37)
CRP, mean (SD)	1.08 (1.7)	1.45 (2.7)	2.95 (6.6)	1.88 (2.95)	2.89 (6.29)	5.17 (8.41)
HbA1c, mean (SD)	5.35 (0.3)	5.49 (0.7)	5.65 (0.6)	-	-	-

Table 2 - Associations of metabolic parameters (numerical and categorical) with measures of alpha-diversity in NFBC1966 and TwinsUK

	NFBC1966 (n=506)[a]								TwinsUK (n=1140)[b]							
	Shannon's diversity				Observed ASVs				Shannon's diversity				Observed ASVs			
	Least-squares mean (95% CI)	Estimate	95% CI	*P*-value	Least-squares mean (95% CI)	Estimate	95% CI	*P*-value	Least-squares mean (95% CI)	Estimate	95% CI	*P*-value	Least-squares mean (95% CI)	Estimate	95% CI	*P*-value
HOMA-IR																
Numerical	-	-0.047	(-0.074; -0.019)	0.001	-	-0.072	(-0.116; -0.027)	0.002	-	-0.062	(-0.091; -0.032)	<0.001	-	-0.106	(-0.168; -0.042)	0.001
Quartile 1	2.80 (2.71; 2.89)	Ref	Ref	Ref	74.4 (70.7; 78.3)	Ref	Ref	Ref	3.91 (3.84;3.98)	Ref	Ref	Ref	214 (206; 223)	Ref	Ref	Ref
Quartile 2	2.74 (2.65; 2.83)	-0.022	(-0.062; 0.018)	0.282	71.9 (68.4; 75.6)	-0.033	(-0.096; 0.030)	0.304	3.83 (3.77; 3.98)	-0.021	(-0.042; 0.000)	0.051	211 (203; 220)	-0.019	(-0.063; 0.025)	0.413
Quartile 3	2.75 (2.67; 2.83)	-0.019	(-0.062; 0.023)	0.379	70.4 (67.1; 73.9)	-0.055	(-0.123; 0.013)	0.118	3.82 (3.75; 3.89)	-0.027	(-0.048; -0.004)	0.017	207 (199; 216)	-0.041	(-0.086; 0.005)	0.083
Quartile 4	2.59 (2.50; 2.68)	-0.078	(-0.129; -0.027)	0.003	66.6 (62.9; 70.4)	-0.111	(-0.192; -0.029)	0.008	3.72 (3.65; 3.79)	-0.055	(-0.077; -0.031)	<0.001	198 (190; 206)	-0.093	(-0.142; -0.042)	<0.001
CRP																
Numerical	-	-0.016	(-0.033; 0.002)	0.085	-	-0.032	(-0.060; -0.004)	0.025	-	-0.018	(-0.034; -0.001)	0.031	-	-0.041	(-0.075; -0.006)	0.019
Quartile 1	2.76 (2.67; 2.85)	Ref	Ref	Ref	71.6 (68.1; 75.3)	Ref	Ref	Ref	3.85 (3.78; 3.93)	Ref	Ref	Ref	211 (203; 220)	Ref	Ref	Ref
Quartile 2	2.67 (2.59; 2.76)	-0.031	(-0.074; 0.011)	0.154	71.1 (67.7; 74.7)	-0.007	(-0.074; 0.060)	0.843	3.83 (3.76; 3.90)	-0.007	(-0.028; 0.014)	0.523	210 (202; 218)	-0.010	(-0.055; 0.035)	0.657
Quartile 3	2.70 (2.62; 2.78)	-0.023	(-0.066; 0.020)	0.303	69.6 (66.2; 73.1)	-0.028	(-0.096; 0.039)	0.415	3.84 (3.77; 3.91)	-0.006	(-0.027; 0.015)	0.599	210 (202; 219)	-0.009	(-0.055; 0.036)	0.688
Quartile 4	2.69 (2.60; 2.78)	-0.027	(-0.072; 0.019)	0.263	68.2 (64.6; 72.0)	-0.049	(-0.122; 0.024)	0.195	3.77 (3.70; 3.84)	-0.028	(-0.050; -0.005)	0.016	201 (193; 209)	-0.063	(-0.111; -0.015)	0.010
$HbA1_c$																
Numerical	-	-0.042	(-0.069; -0.015)	0.003	-	-0.113	(-0.160; -0.067)	<0.001	-	-	-	-	-	-	-	-
Quartile 1	2.71 (2.62; 2.81)	Ref	Ref	Ref	73.6 (69.8; 77.6)	Ref	Ref	Ref	-	-	-	-	-	-	-	-
Quartile 2	2.82 (2.72; 2.92)	0.039	(-0.007; 0.084)	0.099	74.2 (70.2; 78.4)	0.008	(-0.063; 0.079)	0.822	-	-	-	-	-	-	-	-
Quartile 3	2.72 (2.65; 2.80)	0.005	(-0.037; 0.046)	0.832	70.9 (67.9; 73.9)	-0.038	(-0.102; 0.027)	0.256	-	-	-	-	-	-	-	-
Quartile 4	2.66 (2.58; 2.73)	-0.021	(-0.063; 0.022)	0.347	67.0 (64.1; 70.1)	-0.093	(-0.160; -0.025)	0.007	-	-	-	-	-	-	-	-

[a]Adjusted for BMI, gender and smoking status

[b]Adjusted for BMI, age and smoking status

Table 3 – Association of metabolic parameters with community composition using measures of beta-diversity in NFBC1966 and TwinsUK

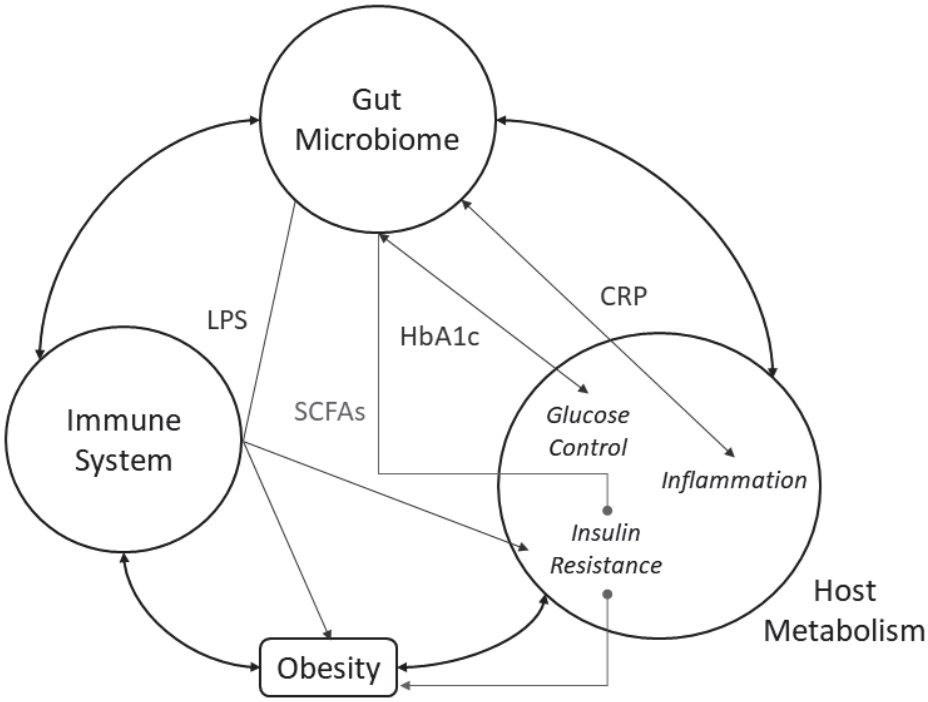
	NFBC1966 (n=506)[a]				TwinsUK (n=1140)[b]			
	Binomial P-value	**Jaccard P-value**	**Weighted Unifrac P-value**	**Unweighted Unifrac P-value**	**Binomial P-value**	**Jaccard P-value**	**Weighted Unifrac P-value**	**Unweighted Unifrac P-value**
HOMA-IR	<0.001	<0.001	0.563	0.004	<0.001	<0.001	0.072	<0.001
CRP	0.112	0.007	0.384	0.047	0.345	0.043	0.241	0.099
HbA1c	<0.001	<0.001	0.112	<0.001	-	-	-	-

[a]Adjusted for BMI, gender and smoking status

[b]Adjusted for BMI, age and smoking status

Figure

Figure 1 - Schematic depicting possible interactions between the gut microbiome and host metabolic components



LPS: Lipopolysaccharides; CRP: C-reactive protein; SCFAs: Short-chain fatty acid (including butyrate); HbA1c: Glycated haemoglobin

The gut microbiome might trigger inflammatory processes associated with obesity and insulin resistance by stimulating immune cells through LPS derived from bacterial membranes. In addition, systematic inflammation, through CRP plasma levels, was associated with lower microbial diversity. Microbial-derived SCFAs, including butyrate, may enhance insulin sensitivity and suppress insulin-mediated fat accumulation. The level of HbA1c was inversely associated with gut microbiome composition, however the mechanisms underlying this relationship are not well understood. The association between the gut microbiome and metabolic health parameters is likely complicated by the fact that the relationship between each component could be bidirectional or even cyclical.

Supplementary material

Figure S1 - Heatmap of Spearman correlation matrix of the measures of metabolic health and BMI in the NFBC1966

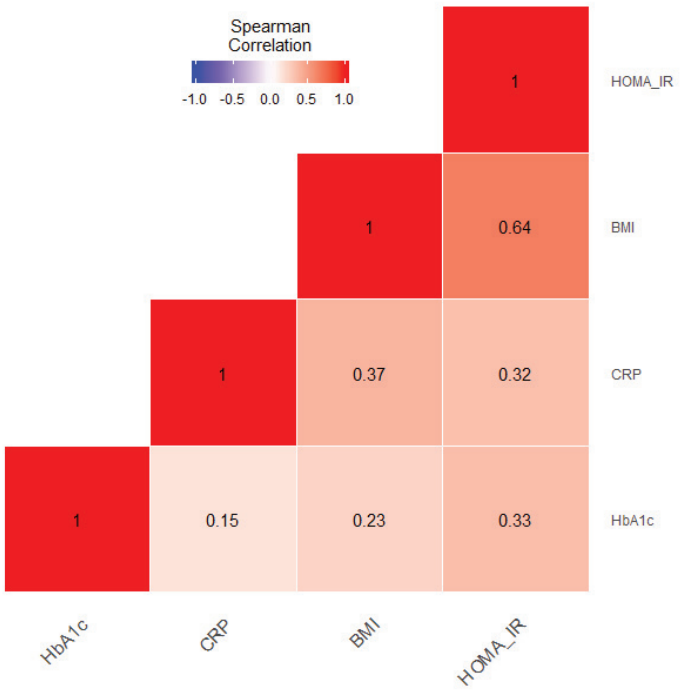


Table S1 – Association of measures of alpha-diversity with BMI before and after adjustment for metabolic markers in NFBC1966 and the TwinsUK

Model adjusted for	NFBC1966 (n=506)[α]				TwinsUK (n=1140)[β]			
	Shannon's diversity		Observed ASVs		Shannon's diversity		Observed ASVs	
	Estimate	P-value[α]	Estimate	P-value[α]	Estimate	P-value[β]	Estimate	P-value[β]
No adjustment	-0.002	0.258	-0.008	0.001	-0.004	<0.001	-0.009	<0.001
HOMA-IR	0.002	0.253	-0.002	0.489	-0.002	0.006	-0.007	<0.001
CRP	0.000	0.972	-0.005	0.082	-0.003	<0.001	-0.008	<0.001
HbA1c	-0.001	0.564	-0.006	0.015	-	-	-	-

[α]Adjusted for BMI, gender and smoking status

[β]Adjusted for BMI, age and smoking status

Table S2 – Association of BMI, before and after adjustment for metabolic markers, with community composition using measures of beta-diversity in NFBC1966 and TwinsUK

Model adjusted for	NFBC1966 (n=506)[α]				TwinsUK (n=1140)[β]			
	Binomial P-value	Jaccard P-value	Weighted Unifrac P-value	Unweighted Unifrac P-value	Binomial P-value	Jaccard P-value	Weighted Unifrac P-value	Unweighted Unifrac P-value
No adjustment	0.004	<0.001	0.113	0.004	<0.001	<0.001	<0.001	<0.001
HOMA-IR	0.099	0.004	0.241	0.258	<0.001	<0.001	0.001	<0.001
CRP	0.015	<0.001	0.011	0.004	<0.001	<0.001	<0.001	<0.001
HbA1c	0.015	<0.001	0.09	0.037	-	-	-	-

[α]Adjusted for gender and smoking status

[β]Adjusted for age and smoking status

Table S3 (a) - Association of HOMA-IR with genera in NFBC1966

ASVs		Mu (mean parameter)		Sigma (precision parameter)		Nu (probability at zero)	
Family	**Genus**	**Estimate**	**Adj. p-value**	**Estimate**	**Adj. p-value**	**Estimate**	**Adj. p-value**
Desulfovibrionaceae	**Desulfovibrio**	-0.0223	<0.0001	0.6482	<0.0001	-0.0833	0.5113
Veillonellaceae	**Megamonas**	1.3963	<0.0001	-0.9280	0.9982	0.0402	0.9982
Peptostreptococcaceae	**Unknown**	-0.5189	<0.0001	-0.0328	0.0048	-0.0112	0.9563
Peptococcaceae	**rc4-4**	11.6284	<0.0001	-0.7770	<0.0001	0.8720	0.3907
Prevotellaceae	**Paraprevotella**	0.0494	<0.0001	0.0433	<0.0001	-0.1222	0.1463
Peptococcaceae	**Peptococcus**	-0.2263	<0.0001	-0.0220	0.9760	0.1305	0.7036
Mogibacteriaceae	**Unknown**	-0.0570	<0.0001	0.1721	0.0005	0.0988	0.0943
Bifidobacteriaceae	**Gardnerella**	2.6445	<0.0001	0.1682	<0.0001	-0.0415	0.9995
Veillonellaceae	**Mitsuokella**	0.1647	<0.0001	0.1890	0.0001	-0.1227	0.1025
Prevotellaceae	**Prevotella**	-0.1753	0.0006	0.2179	0.0006	-0.0214	0.8095
Ruminococcaceae	**Oscillospira**	-0.0779	0.0056	0.0436	0.3736	-0.0953	0.7753
Lachnospiraceae	**Blautia**	0.0633	0.0121	-0.0590	0.1427	0.0475	1.0000
Ruminococcaceae	**Butyricicoccus**	0.0308	0.8515	0.7611	<0.0001	-0.1017	0.3426
Odoribacteraceae	**Odoribacter**	-0.0069	0.8671	-0.6237	<0.0001	-0.1058	0.1664
Corynebacteriaceae	**Corynebacterium**	0.0409	0.9995	-2.8626	<0.0001	0.0415	0.9995
Actinomycetaceae	**Arcanobacterium**	1.0614	0.9997	1.7224	<0.0001	-3.9031	0.9997

Adjusted for BMI, gender and smoking status

Table S3 (b) - Association of CRP with genera in NFBC1966

ASVs		Mu (mean parameter)		Sigma (precision parameter)		Nu (probability at zero)	
Family	**Genus**	**Estimate**	**Adj. p-value**	**Estimate**	**Adj. p-value**	**Estimate**	**Adj. p-value**
Veillonellaceae	**Megamonas**	-0.2106	<0.0001	-0.0659	0.0000	-0.0192	0.9747
Peptococcaceae	**rc4-4**	-0.8197	<0.0001	-0.4562	0.0000	0.4157	0.9999
Paraprevotellaceae	**Prevotella**	-0.1915	<0.0001	0.1622	0.0000	-0.0188	0.9970
Paraprevotellaceae	**Unknown**	-1.2657	<0.0001	0.0302	0.9490	0.1757	0.3926
Peptostreptococcaceae	**Unknown**	-0.6372	<0.0001	-0.7206	0.0000	0.1023	0.7983
Peptococcaceae	**Peptococcus**	0.2825	<0.0001	-0.0758	0.1787	0.0449	0.8639
Alcal	**Sutterella**	-0.4148	0.0001	0.1893	0.7581	0.0514	0.7581
Odoribacteraceae	**Odoribacter**	-0.1705	0.2070	-1.5997	<0.0001	0.4615	0.2738
Ruminococcaceae	**Butyricicoccus**	7.6385	0.2334	0.3027	<0.0001	0.2016	0.8018
Actinomycetaceae	**Arcanobacterium**	0.0864	1.0000	0.2245	<0.0001	-0.4791	1.0000

Adjusted for BMI, gender and smoking status

Table S3 (c) - Association of HbA1c with genera in NFBC1966

ASVs		Mu (mean parameter)		Sigma (precision parameter)		Nu (probability at zero)	
Family	**Genus**	**Estimate**	**Adj. p-value**	**Estimate**	**Adj. p-value**	**Estimate**	**Adj. p-value**
Peptostreptococcaceae	**Unknown**	-1.9110	<0.0001	-0.3171	<0.0001	-0.0227	0.9729
Peptococcaceae	**rc4-4**	-0.7623	<0.0001	-0.9332	<0.0001	1.7686	0.5390
Peptococcaceae	**Peptococcus**	-2.4020	<0.0001	0.3281	<0.0001	0.7325	0.4174
Veillonellaceae	**Mitsuokella**	-2.0149	<0.0001	-6.8202	<0.0001	-0.0237	0.9547
Bifidobacteriaceae	**Gardnerella**	-34.2932	<0.0001	-1.1690	<0.0001	3.8717	0.0132
Oxalobacteraceae	**Oxalobacter**	1.4136	0.0014	0.3782	<0.0001	0.0874	0.9224
Ruminococcaceae	**Butyricicoccus**	0.8421	0.3082	2.7598	<0.0001	-0.4044	0.4060
Desulfovibrionaceae	**Desulfovibrio**	-0.5572	0.3199	1.6649	<0.0001	0.9462	0.3199
Odoribacteraceae	**Odoribacter**	-0.2741	0.5934	-2.6073	0.4788	-0.2817	0.4788

Adjusted for BMI, gender and smoking status

Chapter IV: Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs

1. Background

Population-based studies with gut microbiome data collected prior to chronic disease onset are extremely limited. Colorectal cancer screening programs, which provide screening through fecal tests such as the fecal immunochemical test (FIT) and fecal occult blood test (FOBT), may provide a platform for establishing cohorts with repeated, prospectively collected fecal specimens.

2. Objective

The aim of this study was to compare the impact of different fecal collection methods currently used in ongoing colorectal cancer screening programs on microbiome technical reproducibility, accuracy and stability.

3. Methods

Fecal sample processing

Fecal samples from nineteen healthy volunteers were aliquoted across seven different collection methods, including two cryotubes with no solution (considered to be the putative gold standard), two GenSaver specimen collection cards (Ahlstrom-Munksjö, Helsinki, Finland) used in Afghanistan, two GenCollect specimen collection cards (Ahlstrom-Munksjö, Helsinki, Finland) used in Afghanistan, four OC-Auto Sampling tubes (Eiken Chemical, Tokyo, Japan) used in France (and most other European countries), two Hemotrust tubes (Biosynex, Illkirch-Graffenstaden, France) used in Morocco, two One-Step FOB tubes (Padyabteb, Tehran, Iran) used in Iran, and two Specimen Collection Container A tubes (Alfresa Pharma, Osaka, Japan) used in Turkey. Aliquots were then directly frozen at -80°C, or stored at room temperature, 4°C or 30°C, or mailed to undergo colorectal cancer screening procedures before being frozen. Bacterial loads were calculated based on qPCR and microbiome composition were analysed for a total of 304 samples using 16S rRNA gene sequencing.

Statistical analysis

Intraclass correlation coefficients (ICCs) were calculated using the variance components from a one-way ANOVA to evaluate the stability and accuracy of the different fecal collection methods. The ICCs were computed based on (i) the square root of the relative abundances of the three most dominant phyla (Actinobacteria, Bacteroidetes, and Firmicutes) ant the most abundant genera (*Faecalibacterium*, *Bacteroides*, *UCG-002*, *Subdoligranulum*, *Roseburia*, *Eubacterium eligens* group, *Blautia*, *Christensenellaceae R-7* group, *Ruminococcus*), which were present in at least 50% of fecal samples with relative abundance of ≥0.1%, (ii) our three alpha diversity metrics, and (iii) the first multidimensional scaling axis, also called first principal coordinate (PC1), of our four beta diversity metrics.To calculate accuracy ICCs, we compared one replicate of samples without solution frozen immediately (considered as the gold standard), selected randomly, to one sample from each of the other collection methods for each participant. To calculate stability ICCs at different temperatures and procedures for each fecal collection method, we compared one sample frozen immediately to one stored at different conditions for each participant.

4. Results

Variability in diversity was largely explained by differences between participants and less by the collection method. Accuracy and stability ICCs were generally very high for PC1 of beta diversity matrices, except for OC-Auto Sampling tubes stored at 30°C and for Hemotrust tubes stored at room temperatures. Accuracy measures were very inconsistent for the relative abundance of the three phyla and alpha diversity, and in particular, very low for the inverse Simpson. At the genus level, we did not observe any differences in accuracy compared to the putative gold standard between Gram positive of Gram negative bacteria. Genera that represent a phylogenetically narrow group of species, *UGC-002* and *Christenellaceae RT7* group (≥0.75). Microbial profile stability was very high for specimen collection cards and seemed generally acceptable for FIT tubes (≥0.75), except for Hemotrust tubes (range, 0.06-0.94). Colorectal cancer screening tests did not impact microbiome stability in FIT tubes (≥0.75), however, exposure to summer temperatures (i.e. >30°C) (range, 0.41-0.90) did influence stability.

5. Conclusion

Our study supports previous findings indicating that microbial data obtained from FIT tubes and specimen collection cards are relatively stable and accurate and may be appropriate methods to collect fecal samples for gut microbiome analysis in population-based cohort studies. Furthermore, our findings suggest that opportunistic collection of fecal samples in FIT tubes after colorectal cancer screening is feasible, thereby permitting the potential establishment of cohorts within such screening programs.

6. Scientific article

Zouiouich S, Mariadassou M, Rué O, Vogtmann E, Huybrechts I, Severi G, Boutron-Ruault MC, Senore C, Naccarati A, Mengozzi G, Kozlakidis Z, Jenab M, Sinha R, Gunter MJ, Leclerc M (**2020**) Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs. *mSphere*. (**Submitted**)

Role: First author, contribution to study conceptualisation and logistic, conducted sample collection, conducted analysis, writing original draft and submission.

Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs

Semi Zouiouich (zouiouichs@students.iarc.fr)[1], Mahendra Mariadassou (mahendra.mariadassou@inrae.fr)[2,3], Olivier Rué (olivier.rue@inrae.fr)[2,3], Emily Vogtmann (emily.vogtmann@nih.gov)[4], Inge Huybrechts (huybrechtsi@iarc.fr)[1], Gianluca Severi (gianluca.severi@gustaveroussy.fr)[4,5], Marie-Christine Boutron-Ruault (marie-christine.boutron@gustaveroussy.fr)[4], Carlo Senore (carlo.senore@cpo.it)[6], Alessio Naccarati (alessio.naccarati@iigm.it)[7,8], Giulio Mengozzi (giulio.mengozzi@unito.it)[9], Zisis Kozlakidis (kozlakidisz@iarc.fr)[10], Mazda Jenab (jenabm@iarc.fr)[1], Rashmi Sinha (sinhar@exchange.nih.gov)[3], Marc J. Gunter (gunterm@iarc.fr)[1*], Marion Leclerc (marion.leclerc@inrae.fr)[11*]

Affiliations: [1]Nutrition and Metabolism Section, International Agency for Research on Cancer-WHO, Lyon, France; [2]INRAE, MaIAGE, Université Paris-Saclay, Jouy-en-Josas, France; [3]Université Paris-Saclay, INRAE, BioinfOmics, MIGALE Bioinformatics Facility, Jouy-en-Josas, France; [4]Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA; [4]Université Paris-Saclay, Université Paris-Sud, Université de Versailles Saint-Quentin-en-Yvelines, Centre de Recherche en Epidémiologie et Santé des Populations, National Institute for Health and Medical Research (INSERM), Villejuif, France; [5]Department of Statistics, Computer Science and Applications "G. Parenti", University of Florence, Italy; [6]Epidemiology and Screening Unit-CPO, University Hospital Città della Salute e della Scienza, Torino, Italy; [7]Italian Institute for Genomic Medicine (IIGM), c/o IRCCS Candiolo, Candiolo 10060, Turin, Italy; [8]Candiolo Cancer Institute - FPO IRCCS, Candiolo 10060, Turin, Italy; [9]Clinical biochemistry Unit, University Hospital Città della Salute e della Scienza, Turin , Italy; [10]Laboratory Services and Biobank, International Agency for Research on Cancer-WHO, Lyon, France; [11]Université Paris Saclay, INRAe, AgroParisTech, Micalis Institute, Jouy-en-Josas, France.

*Shared senior authorship

Corresponding author: Semi Zouiouich; 150 cours Albert Thomas, 69372, Lyon CEDEX 08, France; zouiouichs@students.iarc.fr

Key words: colorectal cancer, screening programs, microbiome, fecal sample collection

Abstract

Colorectal cancer screening programs with collection of fecal samples may provide a platform for population-based gut microbiome-disease research. We investigated the impact of fecal sample collection and storage method on the accuracy and stability of 16S rRNA and bacterial load data across seven different collection methods (i.e. no solution, two different specimen collection cards and four types of fecal immunochemical test (FIT) used in four different countries) among 19 healthy volunteers. Intraclass correlation coefficients (ICCs) were calculated for the relative abundance of the top three phyla, the most abundant genera, alpha diversity metrics and the first principal coordinates of the beta diversity matrices to estimate stability of fecal microbial profiles after storage for 7 days at room temperature, 4°C, 30°C and after screening for the presence of occult blood in the stool, and accuracy compared to samples frozen immediately with no solution (i.e. the putative gold standard). When compared to the putative gold standard, significant variation was observed for all collection methods, however, inter-individual variability was much higher than the variability introduced by the collection method. Stability ICCs were high (≥0.75) for FIT tubes that underwent colorectal cancer screening procedures, except for the relative abundance of Actinobacteria (0.65), and were lower for different FIT tubes stored at 30°C (range, 0.41-0.90) and at room temperature (range, 0.06-0.94). Our findings indicate that the use specimen collection cards and different types of FIT are acceptable tools for fecal sample collection methods and inform on their utility for developing microbiome-focused cohorts nested within screening programs.

Importance

Population-based studies with prediagnostic collection of fecal samples are lacking and are necessary for providing robust evidence on the role of the gut microbiome in disease development. Our study indicates that microbial profiles obtained from FIT and specimen collection cards are appropriate methods to collect fecal samples for gut microbiome analysis in population-based cohort studies. Furthermore, our findings suggest that opportunistic collection of fecal samples in FIT tubes after colorectal cancer screening is feasible, thereby permitting the potential establishment of cohorts within such screening programs.

Introduction

Evidence on the role of the human microbiome (i.e. the collection of microbes and microbial genes found in the human body) in the development of chronic diseases such as obesity, diabetes and, potentially cancer, is growing (108,227,228). However, most of the current literature on the association of the microbiome with chronic diseases is based on cross-sectional studies that used diverse methods (229). Most established prospective cohorts did not collect fecal samples and repeated, prospectively collected samples are likely necessary for advancing understanding of the relationship between the microbiome and chronic disease development. Colorectal cancer screening programs, which provide screening through fecal tests such as the fecal immunochemical test (FIT), may offer great potential for establishing population-based cohort studies with multiple fecal specimens and epidemiologic data.

The impact of fecal sample collection methods on gut microbiome parameters has been recently investigated (194,230–234). Microbial populations in fecal samples collected using fecal occult blood tests (FOBT) and FIT have been found to be stable at room temperatures for up to 4 to 7 days, with similar microbial communities compared to samples collected without an additive and frozen immediately. For example, in one study conducted among 52 healthy volunteers in the United States, all fecal sample collection methods (i.e. no additive, 95% ethanol, RNAlater, post-development FOBT cards, and FIT tubes) yielded microbial data that appeared relatively reproducible, stable, and accurate, when compared to the putative gold standard, and provided evidence that these collection methods can be employed for microbiome analyses in population-based studies (194).

However, to potentially inform on the establishment of international studies based in colorectal cancer screening programs, additional methodologic work is needed to test FIT and specimen collection cards used in other countries where different FIT methods and screening procedures are employed. Therefore, we designed a study to evaluate microbial stability in fecal specimens stored at room temperatures and accuracy of microbiome metrics from two different specimen collection cards and four different FIT tubes used in ongoing international colorectal cancer screening programs (in France and most European countries, Morocco, Turkey and Iran). Additionally, we investigated the impact of colorectal cancer screening procedures and alternative shipping

temperatures (e.g., summer and winter temperatures) on microbiome accuracy and stability ascertained from FIT samples.

Results

Microbial alpha diversity by collection method - The Shannon index and inverse Simpson index values were, on average, highest for the immediately frozen samples without solution, (i.e. the putative gold standard) and the number of observed OTUs appeared highest in the specimen collection cards (Figure 1). However, the different storage conditions did not seem to have a statistically significant impact on these alpha diversity metrics. From the linear mixed-effects model, we found that compared to the immediately frozen samples without solution (296, 95% CI [257; 335]), the observed number of OTUs was significantly higher in GenSaver cards at room temperatures (347, 95% CI [308; 387], P = 0.01) and in OC-Auto Sampling FIT tubes at -80°C (344, 95% CI [305; 383], P = 0.03). Compared to the defined gold standard (4.03, 95% CI [3.81; 4.24]), the Shannon index was significantly lower in OC-Auto Sampling FIT tubes at -80°C (3.73, 95% CI [3.51; 3.94], P = 0.005), One-Step FOB tubes at -80°C and at room temperatures (3.76, 95% CI [3.54; 3.98], P=0.03 and 3.49, 95% CI [3.27; 3.72], P < 0.001, respectively), and in Specimen Collection Container A tubes at -80°C and at room temperatures (3.60, 95% CI [3.38, 3.82], P < 0.001 and 3.70, 95% CI [3.48; 3.91], P < 0.001, respectively).

Percent variability explained by participant and collection method - All multidimensional scaling ordinations suggested that the between participant variability was higher than the technical variability (Figure S2 in supplementary material). In multidimensional scaling (MDS) plots, each point represents one microbiome sample for 19 participants and seven collection methods; all samples from a given participant tended to group together (as shown by ellipses, one per participant). Additionally, based on four distance matrices (i.e. Jaccard, Bray-Curtis, unweighted UniFrac and weighted UniFrac), the overall variability in diversity was largely explained by between participants (55% to 79%) and only marginally by collection methods (4.8% to 14.8%, Figure S3). For example, based on the Bray-Curtis distance matrix, the protocol variability accounted for roughly 6.9% of the overall variability whereas the biological variability accounted for 79.0%.

Relative abundance comparisons - At the phylum level, the distributions of relative abundances of each phylum were consistent for all the collection methods; samples were mainly represented by Firmicutes followed by Bacteroidetes (Figure 2). At the genus level, when compared to the putative gold standard, the relative abundance of *Faecalibacterium* was greater in the other collection methods, especially in the OC-Auto Sampling FIT tubes. Consequently, when compared to gold standard, the relative abundance of other genera, such as *Bacteroidetes* or *Blautia*, was lower in other collections methods. However, there was also substantial inter-individual variability at the genus level (Figure S4 in supplementary material).

Accuracy compared with putative gold standard – Samples collected without solution and immediately frozen at -80°C were considered as the putative gold standard and compared to samples collected using other methods in different storage conditions using ICCs (Figure 3, and Table S4 in supplementary material). Accuracy ICCs were variable for the relative abundance of Actinobacteria (range, 0.28–0.83), Bacteroidetes (range, 0.24–0.84), Firmicutes (range, 0.26–0.82) as well as for the inverse Simpson index (range, 0.23-0.73) and the weighted UniFrac (range, 0.46–0.93). Specifically, for the relative abundance of Actinobacteria, ICCs were ≥75% for OC-Auto Samples tubes stored at 30°C and those that went through screening, for One Step FOB tubes stored at room temperatures and for all the specimen collection cards. For the relative abundance of Bacteroidetes, ICCs were ≥75% for OC-Auto Samples tubes that went through screening, for One Step FOB tubes immediately frozen and for Specimen Collection Container A tubes immediately frozen. For the relative abundance of Firmicutes, ICCs were ≥75% for One Step FOB tubes immediately frozen and for Specimen Collection Container A tubes immediately frozen. ICCs were ≥75% for PC1 scores of weighted UniFrac for all the methods, except for OC-Auto Samples tubes stored at 30°C and for Hemotrust tubes stored at room temperatures. At the genus level, we did not detect any links between the accuracy differences and the Gram positive of Gram negative (Figure 4). As expected, genera that represent a phylogenetically narrow group of species, *UGC-002* and *Christenellaceae RT7* group, showed accuracies above 0.75. Conversely, ICCs were higher for observed OTUs (≥0.74) except for Hemotrust at room temperatures (0.31), OC-Auto Sampling tubes immediately frozen (0.66) and at 30°C (0.40), and One-Step FOB at room temperatures (0.60). Specimen collection cards showed the highest ICCs

for the Shannon index (range, 0.79–0.88). ICCs were ≥75% for PC1 scores of Jaccard, Bray-Curtis and unweighted UniFrac distances for all collection methods. Additionally, technical variability was also quantified based on beta diversity distances between the two replicates of the putative gold standard and used as baseline to assess the accuracy of each method. For each participant, the distance between the putative gold standard and each collection method was computed (Figure S5 in supplementary material). All those distances were slightly higher than the baseline, suggesting that the effect the collection methods exceeds the technical variability.

Stability – In each collection method, samples frozen immediately were compared to the samples stored in different conditions using ICCs (Figure 5, and Table S5 in supplementary material). When compared to specimen collection cards frozen directly, collection cards (i.e. GenCollect and GenSaver) kept at room temperatures showed high stability, with ICCs ≥89% for all seven metrics. Stability ICCs were generally lower and more variable in samples collected using Hemotrust tubes (range, 0.06-0.94). For One-Step FOB tubes the confidence intervals were wide, indicating high variability in stability at room temperatures. Specimen Collection Container A tubes at room temperatures showed stability ICCs ≥75% except for the relative abundance of Actinobacteria (0.58) and Firmicutes (0.66). Stability ICCs were lower and inconsistent for OC-Auto Sampling FIT tubes stored at 30°C (range, 0.41-0.90). However, stability ICCs were higher for OC-Auto Sampling FIT tubes stored at 4°C (≥0.75) except for the number of observed OTUs (0.65). Stability ICCs were also high for OC-Auto Sampling FIT tubes that went through colorectal cancer screening procedures, with ICCs ≥75%, except for the relative abundance of Actinobacteria (0.65). Interestingly, at the genus level, the stability remained very high, especially for Gencollect and Genesaver. It is worth pointing that *Blautia* genus, including the homoacetogen *Blautia hydrogenotrophica,* was significantly lower at room temperature (Figure 2B and Figure 6) In addition, *UGC-002*, but also *Faecalibacterium, Roseburia* and *Ruminococcus* were also subjected variations when stored at 30°C.

Discussion

In this study we compared microbiome stability and accuracy across different fecal sample collection methods (i.e. two different specimen collection cards and four different FIT tubes) used in ongoing colorectal cancer screening programs. We found

that the overall variability in diversity was largely explained by differences between participants and less by the collection method. In addition, accuracy and stability ICCs were generally very high for PC1 of beta diversity matrices, except for OC-Auto Sampling tubes stored at 30°C and for Hemotrust tubes stored at room temperatures. Accuracy measures were very inconsistent for the relative abundance of the three phyla and alpha diversity, and in particular, very low for the inverse Simpson. This highlights the importance of using one consistent method for study comparisons. Overall, microbial profile stability was very high for specimen collection cards and seemed generally acceptable for FIT tubes, except for Hemotrust tubes. Colorectal cancer screening tests did not impact microbiome stability in FIT tubes, however, exposure to summer temperatures (i.e. >30°C) did influence stability. These results are informative for the development of future population-based cohorts with fecal sample collection within colorectal cancer screening programs.

As shown in prior studies that have assessed microbiome diversity in human or animal samples, microbial composition and diversity were largely explained by between-participants differences and only marginally by the collection methods (194,233). Fecal specimen collection cards have been previously tested for microbial analysis in other studies and have shown moderate to excellent accuracy compared to the putative gold standard and stability at room temperatures (230,233,235–240). In this study, the specimen collection cards (i.e. GenCollect and GenSaver) stored at room temperatures for 10 weeks showed excellent stability when compared to the immediately frozen cards, although other studies detected lower amounts of DNA among fecal samples from humans and animals collected on FTA cards (Whatman) after several weeks (238,241). Also consistent with previous findings, we found that specimen collection cards tended to differ in bacterial taxa composition and have observed OTUs than when using the putative gold standard (194,235,241), but we did not observed higher Shannon and inverse Simpson. One hypothesis to explain this observation is that the card matrix may increase chemical cell lysis (235); as in our study, Firmicutes, often spore formers, were the most represented phylum in sample collected with specimen collection cards, supporting the hypothesis that cell lysis might be one explanation for these differences. Furthermore, *Blautia* genus, including the homoacetogen *Blautia hydrogenotrophica*, is also adequately represented in the collection cards, despite being highly sensitive to oxygen.

In previous studies, fecal samples stored in FIT tubes have also shown moderate to excellent accuracy compared to the gold standard and stability at room temperatures (234,242,243). However, the different types of FIT tubes in our study did not seem to perform equally. For the specimen cards and some of the FIT tubes, especially OC-Auto Sampling and Hemotrust tubes stored at -80°C, we detected significant differences in the relative abundance of phyla and genera with higher levels of Firmicutes and *Faecalibacterium* genus, including the anti-inflammatory butyrate producing *Faecalibacterium prausnitzii*, when compared to the gold standard, which support observations from a previous study (243). As described previously (237,242), when compared to the gold-standard, FIT tubes showed good to excellent accuracy for beta diversity metrics, however, some FIT tubes, including OC-Auto Sampling, Hemotrust and Specimen Collection Container A tubes, revealed lower accuracy for alpha diversity metrics, especially those stored for 7 days at room temperatures or at 30°C. Previous studies have found that FIT tubes showed good stability at room temperatures (194,234,243).

Regarding stability at room temperature, we found variable results for the different types of FIT tubes. When compared to those directly frozen, Hemotrust tubes stored at room temperature for 7 days showed poor stability for alpha diversity metrics and for One-Step FOB tubes the confidence intervals were wide, indicating high variability in stability at room temperature. In this study, we detected different stabilities at room temperature for *Eubacterium eligens* and *Roseburia*, other butyrate producers from the human gut microbiota. As OC-Auto Sampling tubes seemed to be less stable at 30°C, the collection and shipping of samples during high temperatures might have an impact on gut microbiome composition. These variations in accuracy and mostly in stability between the types of FIT tubes might be due to differences in DNA-stabilizing and anti-microbial properties of the solution (the composition of which remains proprietary for most methods) inside the tubes, impacting the stabilization of DNA, prevention of bacterial growth, and preservation of microbial profiles.

This study has several limitations. First, we included principally female, healthy participants, which might limit the inference of our results to general populations. However, previous studies have found that stability and accuracy for comparison of

samples that were frozen immediately without solution to other fecal sample collection methods were similar between different populations (194,232,234). Second, we used 16S rRNA gene sequencing to characterize the microbial composition, while other profiling methods such as whole-genome shotgun metagenomics are becoming more commonly employed in high-income settings(244). However, 16S rRNA gene sequencing remains the most affordable method to study the gut microbiome diversity, especially in the context of large epidemiologic cohorts. Additionally, in low-to-middle income countries, where sequencing technologies are not always available, these collection methods, and more specifically specimen collection cards, could be used to detect specific biomarker species that have been associated with specific diseases for example *Fusobacterium nucleatum* or *Parviromonas micra* which have been associated with colorectal cancer development(229,245). Finally, as previously described, freezing procedures might have an impact on relative abundance of Gram-positive and Gram-negative bacteria (246,247), and therefore, considering samples without solution frozen immediately as the gold standard method might be suboptimal. However, in the context of large population-based cohorts, immediate DNA extraction after defecation is likely unfeasible, and therefore, standardization of storage protocols is necessary.

Our study included many different collection methods, including two different specimen collection cards and four different FIT tubes, which are currently being used in ongoing colorectal cancer screening programs around the world. The different collection methods tested in this study has allowed us to highlight a number of important considerations for sample collections in large population settings such as acceptability from participants, safety for processing samples, the volume of kit necessary, the storage logistics and cost. For instance, due to the rigidity of the tubes, the solution in OC-Auto Sampling tubes was not easy to sample safely, however the amount of bacterial DNA obtained by qPCR indicated that sufficient material was available in these tubes. Based on qPCR data, the collected volume could be lowered from 500 microliters to 200 microliters to obtain sufficient DNA for several analyses and rendering it unnecessary to extract all of the liquid which might be logistically challenging. Additionally, storage of specimen collection cards is easier and cheaper than tubes, which might help low-to-middle income countries to develop infrastructure for microbiome research. Finally, because of the small size of the kits, the acceptance

is expected to be higher in the general population, compared to the gold standard, where provision of a whole fecal sample could make subjects uncomfortable and reluctant to participate. Furthermore, in this study, stability was assessed over the course of several days to several weeks at room temperature, but also for different conditions directly reflecting settings in colorectal cancer screening (i.e. shipping at different temperatures reflecting seasonal variation, colorectal cancer screening procedures including mailing and occult blood detection test). This is the first study in which the impact of colorectal cancer screening procedures on fecal samples using FIT tubes has been demonstrated. Importantly, OC-Auto Sampling FIT tubes that went through colorectal cancer screening procedures and tests had good stability, opening opportunities for establishing prospective cohorts within screening populations.

In conclusion, our study supports previous findings indicating that microbial data obtained from FIT tubes and specimen collection cards are relatively stable and accurate and may be appropriate methods to collect fecal samples for gut microbiome analysis in population-based cohort studies. Furthermore, our findings suggest that opportunistic collection of fecal samples in FIT tubes after colorectal cancer screening is feasible, thereby permitting the potential establishment of cohorts within such screening programs. Since different collection methods and high temperatures impact the stability and accuracy, it is important for future investigators, in the context of the implementation of large-scale epidemiologic studies, to coordinate their efforts and follow standardized protocols in order to accurately compare the microbiome between sites, groups or countries and be able to pool microbial data.

Methods

Study participants - Nineteen healthy participants from the International Agency for Research on Cancer (IARC) personnel were recruited in Lyon, France. Eligible participants were at least 18 years of age or older, had not taken antibiotics in the past 3 months and were willing to provide a fecal sample. All participants provided informed consent, and the study was approved by the IARC Ethics Committee. On average, participants were 40.2 years old and had a body mass index of 22.7 kg/m^2. The majority of participants were female (78.9%), had no weight variation in the past six months (73.7%) and reported having a regular bowel movement at least once per day (94.8%) (Table S1 in supplementary material).

Fecal sample collection - At recruitment, participants were provided with a fecal sample collection kit including all materials needed for the fecal sample collection at the workplace or at home along with collection instructions. To collect the fecal sample, participants were asked to provide a fecal sample in the provided fecal collection containers (Sarstedt, Nümbrecht, Germany) by filling the scoop contained in each tube. The participant collected the fecal sample at the workplace (n=9) or at home (n=10) before going to work and returned it to the study coordinator within a few hours after collection. The study coordinator then delivered it to the laboratory for immediate processing. The participants then completed a questionnaire that was used to obtain information on the time and date of sample collection, typical bowel movements and general information.

Following the collection step, the fecal samples were manually homogenized and aliquoted for the different collection methods (Figure S1 in supplementary material). For each participant, fecal samples were aliquoted into two cryotubes without solution (considered to be the putative gold standard, average weight per cryotube 208.1 mg), two GenSaver specimen collection cards (Ahlstrom-Munksjö, Helsinki, Finland) used in Afghanistan, two GenCollect specimen collection cards (Ahlstrom-Munksjö, Helsinki, Finland) used in Afghanistan, four OC-Auto Sampling tubes (Eiken Chemical, Tokyo, Japan) used in France, two Hemotrust tubes (Biosynex, Illkirch-Graffenstaden, France) used in Morocco, two One-Step FOB tubes (Padyabteb, Tehran, Iran) used in Iran, and two Specimen Collection Container A tubes (Alfresa Pharma, Osaka, Japan) used in Turkey. For specimen collection cards, a disposable wooden spatula was used to smear a small portion of the homogenised feces on each window of the specimen collection cards; the flaps on the cards were closed and then each card was placed in a separate biohazard bag with desiccant bags. The FIT tubes were filled following the instructions provided by the different colorectal cancer screening programs. Specifically, the FIT probes were dipped into the homogenised fecal specimen, and returned to the FIT tubes and were then shaken to mix as instructed. All of these steps were performed by the same laboratory technician.

Fecal sample storage - The study samples are outlined in Table 1. After sample collection and processing at IARC, the two cryotubes and one of each sample type

were immediately stored at -80°C. One of the OC-Auto Sampling tubes was placed in a blue mailer for colorectal cancer screening and mailed to the Reference Centre for Epidemiology and Cancer Prevention in Piemonte (CPO Piemonte) laboratory within 24 hours after processing. To mimic mailing during seasonal temperatures in France, one of the OC-Auto Sampling tubes was stored at a winter temperature (4°C) for 7 days in a refrigerator, returned to room temperature for at least 4 hours, and then frozen at -80°C. The remaining OC-Auto Sampling tubes was stored at a summer temperature (30°C) for 7 days in a water bath, returned to room temperature for at least 4 hours, and then frozen at -80°C. The remaining FIT tubes (i.e. Hemotrust, two One-Step FOB and Specimen Collection Container A tubes) were stored for 7 days at room temperature and then frozen at -80°C. The remaining half of the specimen collection cards (i.e. GenCollect and GenSaver cards) remained at room temperature in a closed cupboard from the date of collection until the date of DNA extraction (average 70.2 days).

On arrival at the CPO Piemonte laboratory, the FIT tubes were processed using standard colorectal cancer screening procedures (248). After testing for occult blood, the FIT tubes were immediately closed with parafilm and shipped to the Micalis Institute (INRAE/AgroParisTech) in Jouy-en-Josas, France for 16S rRNA gene profiling. Upon receipt of all the specimens at the same time, the samples were removed from the tube using pliers and a sterile pipette, transferred to a 1.5 mL Eppendorf tube, shaken, and then two 250 µL aliquots were transferred to a sterile tube and frozen at -80°C.

DNA extraction, real-time quantitative PCR and 16SrRNA sequencing - The samples remaining at IARC were shipped on dry ice, except for the specimen collection cards stored at room temperatures, to the Micalis Institute (INRAE/AgroParisTech) in Jouy-en-Josas, France. For each collection method, sampling volumes were collected with different approaches, for optimal recovery, sterility and safety, as described below. For the specimen collection cards, a rectangular section of the filter was cut in the centre of the filter using a sterile scalpel (average concentration per slice 41.5 ng/µL).

The FIT tubes were thawed for 20 minutes on ice, then manually shaken for 15 seconds and finally, 500 µL of fecal solution were collected as follows: Hemotrust tubes

were manually opened, using the bottom red plastic screw, and gently pressed to transfer all of the liquid into a sterile 2 mL screw cap tube; One-Step FOB tubes were manually opened using the top stopper and the liquid poured into a 2 mL sterile screw cap tube; OC-Auto Sampling tubes and Specimen Collection Container A tubes were opened from the top and a long pipette tip was used to collect 4x125 µL while shaking in between each sampling.

Total DNA was extracted from 500 µL aliquots or one sliced section of the cards using the PowerFecal DNA Isolation kit (MoBio Laboratories, Carlsbad, CA, United States) according to the manufacturer's protocol with minor modifications consisting of i) decreasing the volume of buffer based on the 500 µL aliquots taken from the FIT kits or ii) for the GenCollect and GenSaver keeping the paper filters during the first centrifugations and cell lysis. DNA purity and concentrations of FIT tubes and specimen collection cards were estimated using NanoDrop (Thermo Fisher Scientific, Waltham, Massachusetts, USA) measurements. To account for lack of specificity of NanoDrop, qPCR was used to compare biomass and DNA yield differences in all samples. DNA extracts were stored at -20°C before being used for real-time qPCR and 16S rRNA gene V3-V4 PCR amplification.

DNA extracts were used for bacterial biomass quantification by qRT-PCR analysis of the 16S rRNA genes as previously described (249). Total bacterial biomass was quantified by real-time qPCR using specific primers (HAD-1: 5'-TGGCTCAGGACGAACGCTGGCGGC -3' and HAD-2: 5'-CCTACTGCTGCCTCCCGTAGGAGT-3', annealing at 59 ºC), Fast SYBR Green MasterMix (Applied Biosystems) and the StepOne Real-Time PCR system (Applied Biosystems, Life Technologies). A standard curve was generated from serial dilutions of a known copy number of the target gene cloned into a plasmid vector. Final dilution of DNA samples for qPCR were 10^{-3} or 10^{-4} -fold depending on the sample. Bacterial biomass was obtained from two distinct qPCR runs for each participant (Tables S2a, S2b, S2c and S2d in supplementary material). A sample of 25ng of DNA underwent PCR amplification of the V3-V4 region of the 16S rRNA gene; DNA from the specimen collection cards and cryotubes were diluted to 1/10 if their concentration exceeded 30 ng/µL. Sequencing was performed at the GenoToul INRAE platform (Castanet-Tolosan, France) using Illumina technology with MiSeq kit V2 2 x 250 bp.

Quality control - A 384-well plate was used for sequencing all the samples in one run. The plate contained quality control samples including negative controls after V3-V4 PCR and positive controls from human gut microbiota that had already been sequenced and analyzed. Most samples from each collection method presented sufficient bacterial biomass for the V3-V4 PCR preceding 16S rRNA sequencing. However, differences in bacterial amounts were observed between the collection methods, with lower quantity for Specimen Collection Container A tubes and One-Step FOB tubes (Table S2c in supplementary material). In detail, one subject had a lower amount of material for two of the One-Step FOB tubes, and two subjects for one of the One-Step FOB tubes, preventing to continue with processing for these samples. Three GenCollect and one GenSaver specimen collection cards were first negative for PCR amplification and then positive after a second PCR with 50ng of DNA instead of 25ng. Conversely, for some fit kits samples (stored at room temperature), PCR which were first negative, were positive when 10ng of DNA extract was used, suggesting the presence of PCR inhibitors. Additionally, technical variability was also quantified based on the distance between the two replicate samples without solution frozen immediately. Technical reproducibility was high, with ICCs ≥89% between replicates of the gold standards for all seven tested metrics.

Bioinformatics - Data were stored on secured servers at INRAE MAIAGE (Jouy-en-Josas, France) and IARC (Lyon, France). Sequencing data were analyzed with Find, Rapidly, Otus with Galaxy Solution (FROGS) v3.1.0 (250). Briefly, this pipeline included a pre-processing step where reads were merged with Paired-end read merger (PEAR) (251), dereplicated, and filtered according to their length, mismatches in primers with cutadapt (252), and N content. This step was followed by Swarm clustering (253) with an agglomeration distance of d = 3. Chimera detection was then performed using VSEARCH (254) before applying an OTU abundance filter (OTUs <0.005% of the total abundance are discarded). The most abundant sequence of each OTU was then affiliated with 100% similarity with blastn against the Silva v138 database (255). Two samples were discarded because of low sequencing depth and all remaining samples were rarefied to 7,144 reads per sample using R package phyloseq. Diversity metrics were then computed to represent the diversity of OTUs in each sample (alpha diversities: Shannon index, number of observed OTUs and Inverse

Simpson index) and the differences between samples (beta diversities: Jaccard, Bray-Curtis, weighted and unweighted UniFrac distances).

Statistical analysis – Statistical analyses were conducted using R, version 3.6.2 and the packages DESeq2, icc, phyloseq and vegan (202,256–258). Descriptive characteristics of study participants were based on the questionnaire provided by the participants. We performed visualization using multidimensional scaling (MDS) plots and all samples from a given participant tended to group together. To assess the impact of the protocol on the alpha diversity, comparison of measures of alpha diversity between each fecal collection method was performed and a linear mixed-effects models with the collection method as fixed effect and the participant as random effect was fitted to the data and used to calculate least-squares means of the alpha diversity metrics. Mean diversities for the collection methods were then compared using Tukey's HSD tests. To estimate a distance-based coefficient of determination (R^2) explained by participant and collection method from unweighted UniFrac, weighted UniFrac, Jaccard and Bray-Curtis distance matrices, permutational multivariate analysis of variance was performed (adonis() function, vegan package, R) (202). We calculated intraclass correlation coefficients (ICCs) using the variance components from a one-way ANOVA to evaluate the stability and accuracy of the different fecal collection methods. The ICCs were calculated based on (i) the square root of the relative abundances of the three most dominant phyla (Actinobacteria, Bacteroidetes, and Firmicutes) ant the most abundant genera (*Faecalibacterium*, *Bacteroides*, *UCG-002*, *Subdoligranulum*, *Roseburia*, *Eubacterium eligens* group, *Blautia*, *Christensenellaceae R-7* group, *Ruminococcus*), which were present in at least 50% of fecal samples with relative abundance of ≥0.1%, (ii) our three alpha diversity metrics, and (iii) the first multidimensional scaling axis, also called first principal coordinate (PC1), of our four beta diversity metrics. The first axis explained 8.2%, 15.8%, 17.6%, and 49.7% of the variability for Jaccard, Bray-Curtis, unweighted UniFrac and weighted UniFrac distances, respectively. To calculate accuracy ICCs, we compared one replicate of samples without solution frozen immediately (considered as the gold standard), selected randomly, to one sample from each of the other collection methods for each participant. To calculate stability ICCs at different temperatures and procedures for each fecal collection method, we compared one sample frozen immediately to one stored at different conditions for each participant. The 95%

confidence interval was estimated using the ICCest() function from the R ICC package with default option confidence interval = “Smith”.

Acknowledgments

Disclaimer - Where authors are identified as personnel of the International Agency for Research on Cancer / World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer / World Health Organization.

Funding - This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors; this research was jointly funded by IARC and INRAE.

Recognition of personal assistance – The authors of this paper would like to thank Christophe Lallemand from IARC and Laurence Brazeau from INRAE for their help with laboratory work.

Author contributions - Conceptualization, S.Z., M.G., G.S., M-C.B-R., R.S. and E.V.; laboratory methodology, M.L., and Z.K.; screening procedures, G.M., C.S. and A.N.; formal analysis, O.R., M.M., S.Z. and M.L.; writing—original draft preparation, S.Z.; writing—review and editing, M.G., M.L., M.J., I.H., and E.V.; All authors have read and agreed to the published version of the manuscript.

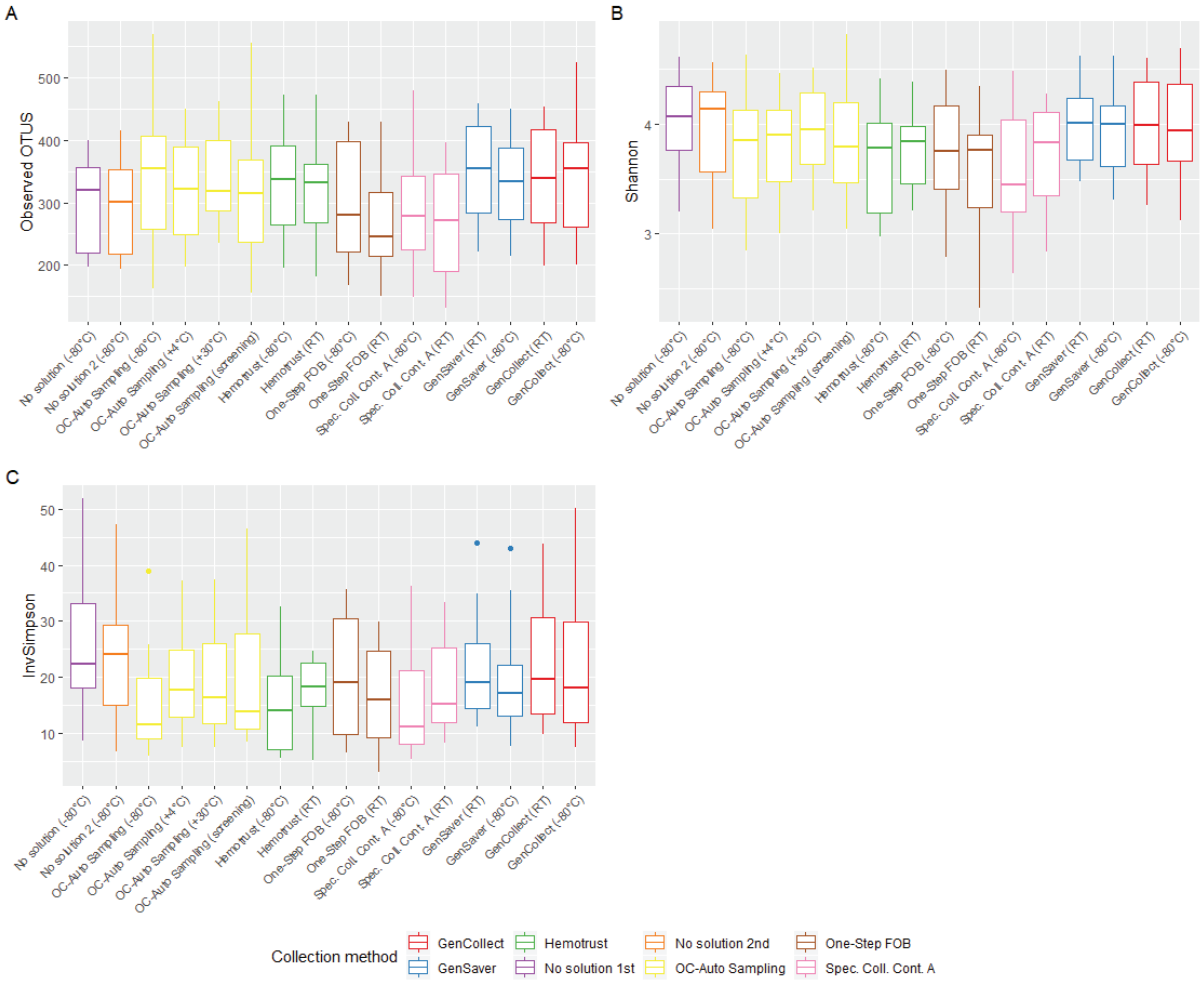
Tables

Table 1 - Collection methods for fecal samples and number of aliquots used for microbiome analyses per participants, IARC, Lyon, France, 2019.

Collection Method	Frozen immediately	4°C for 7 days	30°C for 7 days	Room temperature for 7 days	Room temperature for study duration	Mailed to CPO Piemonte, Italy
No solution	2	0	0	0	0	0
GenSaver cards	1	0	0	0	1	0
GenCollect cards	1	0	0	0	1	0
OC-Auto Sampling tubes	1	1	1	0	0	1
Hemotrust tubes	1	0	0	1	0	0
One-Step FOB tubes	1	0	0	1	0	0
Specimen Collection Container A tubes	1	0	0	1	0	0

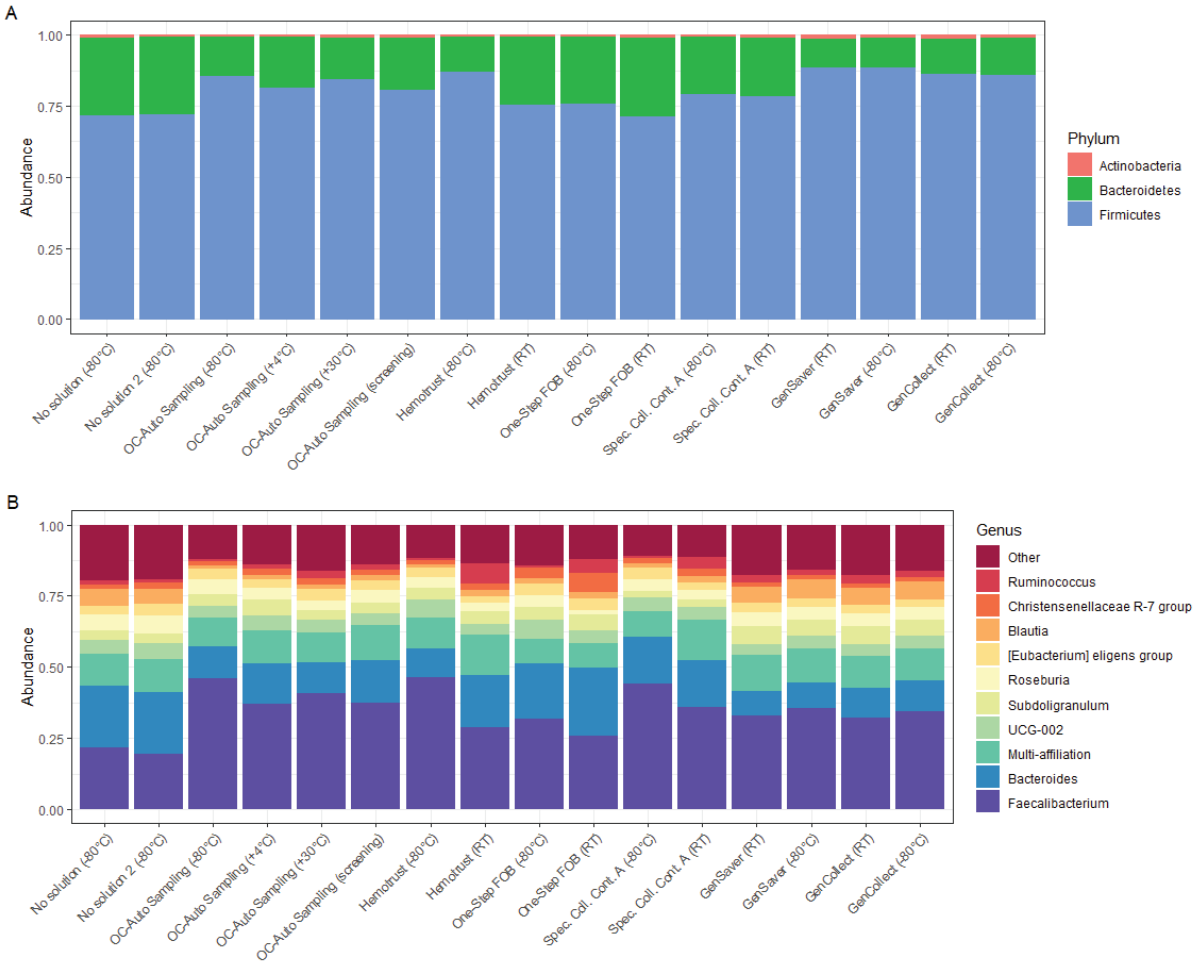
Figures

Figure 1 Impact of specimen collection methods on microbial alpha diversity indexes.



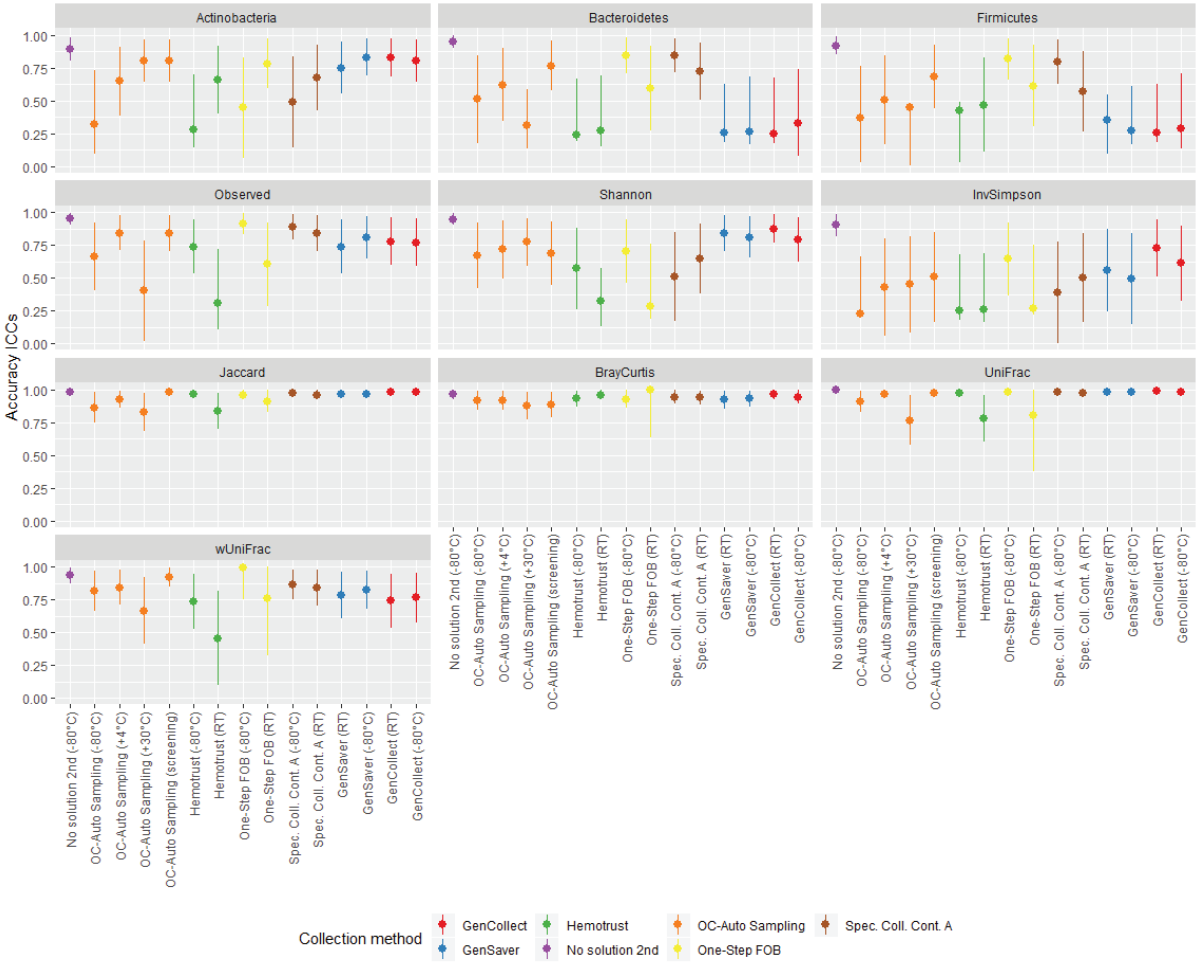
Box represent the median and interquartile range for observed OTUs (A), Shannon index (B) and inverse Simpson index (C) by specimen collection method.

Figure 2 - Variability of microbial phyla and prevalent genera obtained from the collection methods.



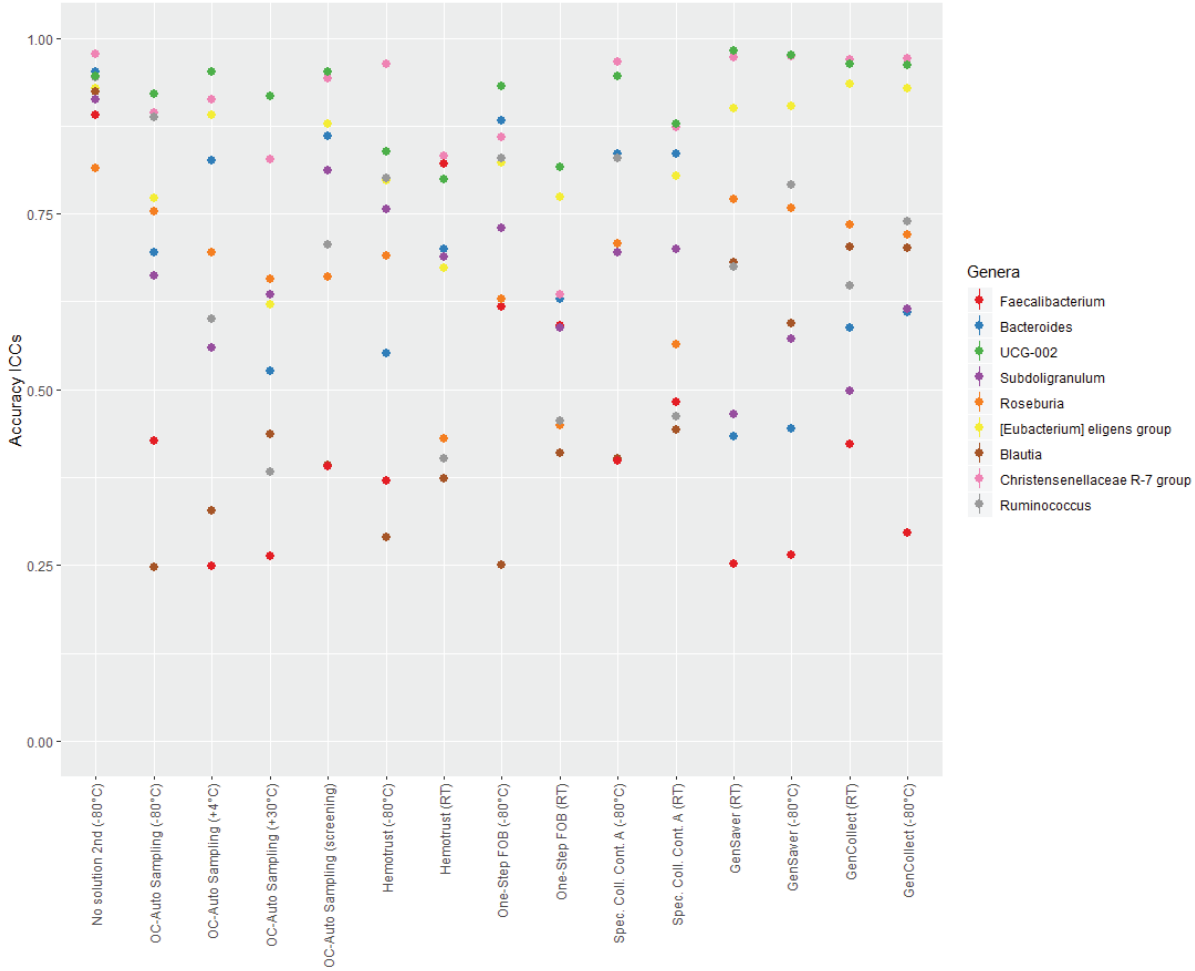
Bars represent the relative abundance of bacterial phylum (A) and genus (B) present in at least 50% of fecal samples with relative abundance of ≥0.1% in each collection methods.

Figure 3 - The accuracy of phylum distribution and diversity metrics accuracy differ upon the collection method.



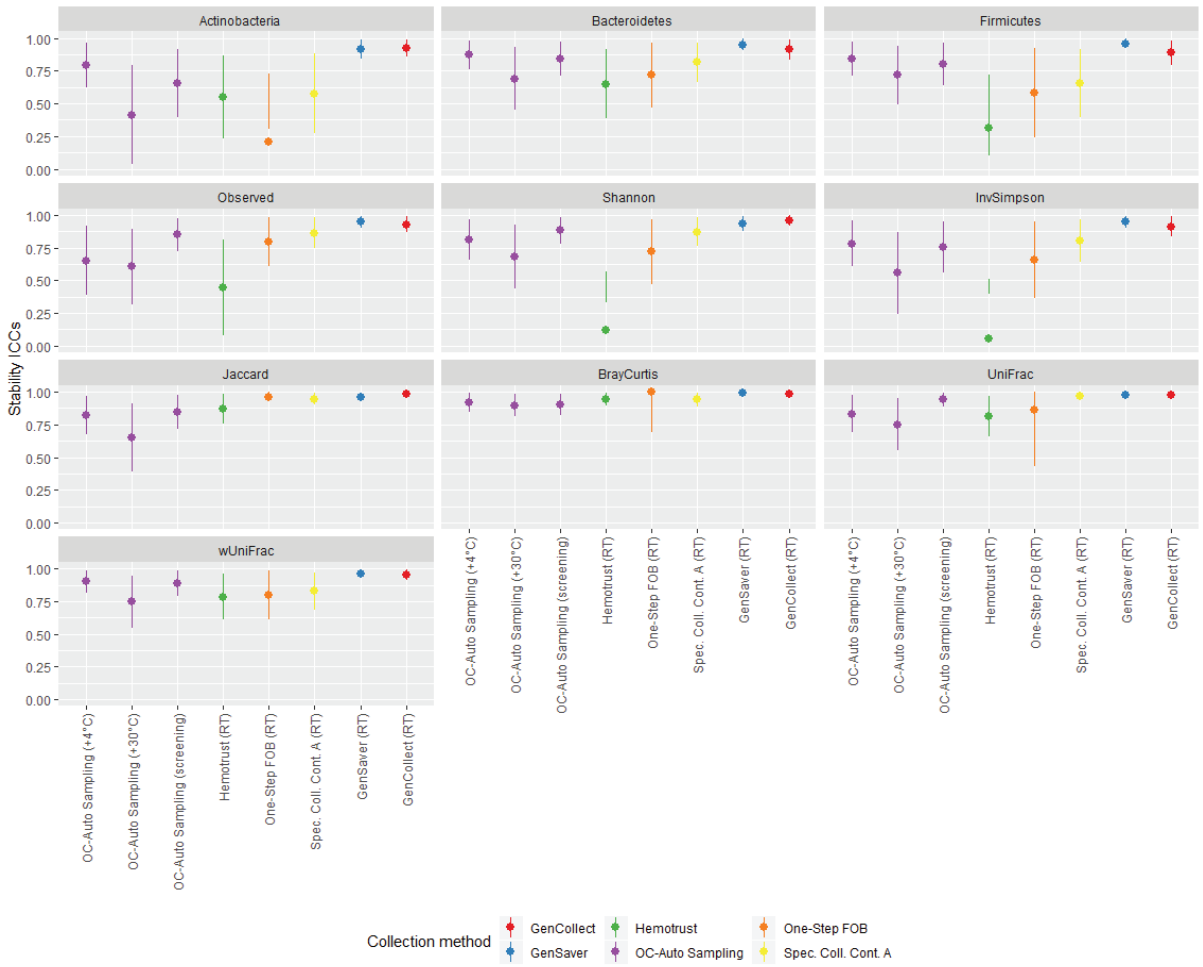
ICCs for accuracy of microbiome diversity metrics are represented of each fecal sample collection method compared to the gold-standard (samples with no solution, frozen immediately).

Figure 4 - The accuracy of genera distribution differ upon the collection method.



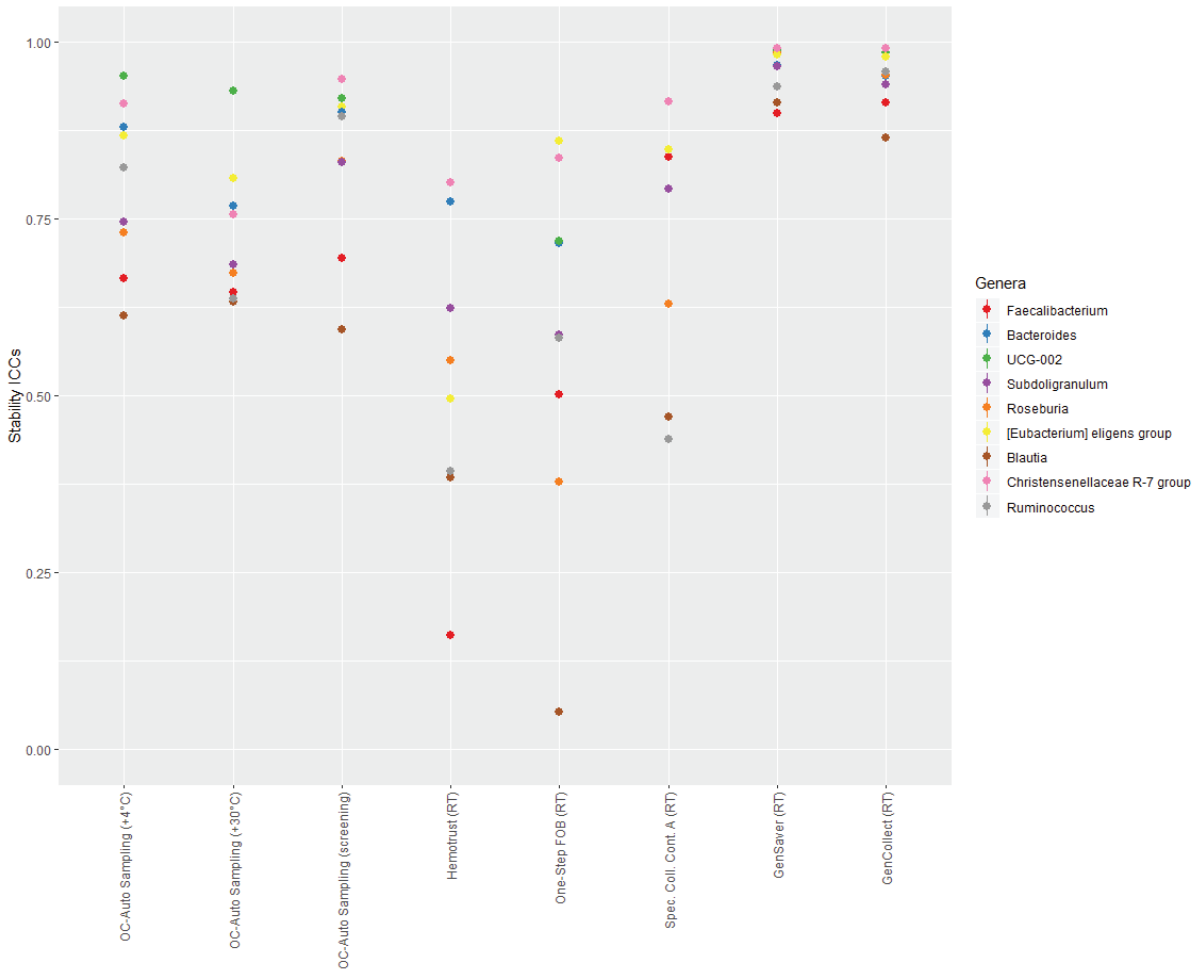
ICCs for accuracy of the relative abundance of bacteria at the genus level are represented of each fecal sample collection method compared to the gold-standard (samples with no solution, frozen immediately).

Figure 5 – Impact of storage conditions on the microbiota composition and diversity indexes.



ICCs for stability of microbiome diversity metrics of each fecal sample collection method stored in different conditions compared to those directly frozen.

Figure 6 – Impact of storage conditions on the most abundant genera.



ICCs for stability of the most abundant bacteria at the genus level of each fecal sample collection method stored in different conditions compared to those directly frozen.

Supplementary material

Figures

Figure S1 - Pictures of the different collection methods: (A) OC-Auto Sampling tubes, (B) Hemotrust tubes, (C) Specimen Collection Container A tubes, (D) One-Step FOB tubes, (E) GenSaver and GenCollect specimen collection cards.

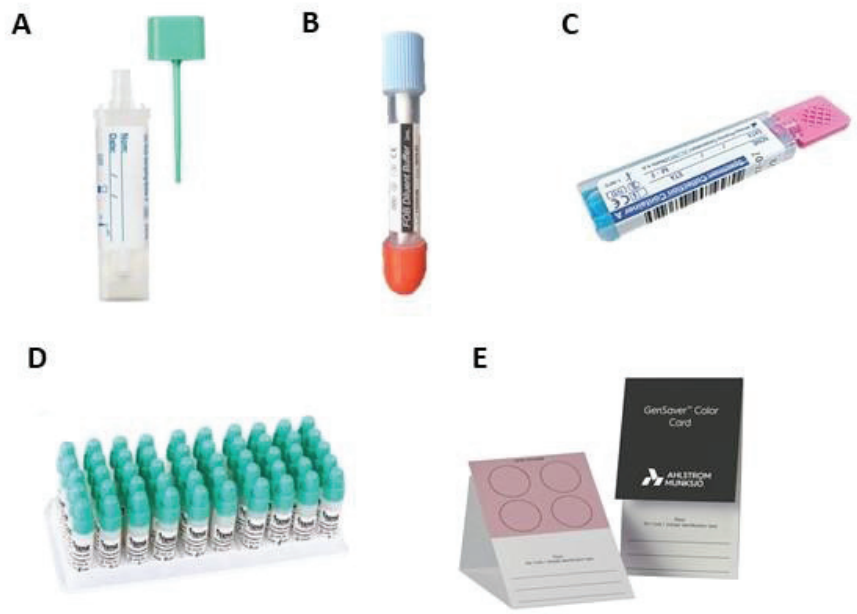


Figure S2 – Multidimensional scaling (MDS) plots for microbiome feature data types. In multidimensional scaling (MDS) plots, each point represents one microbiome sample for 19 participants and seven collection methods. Four distance matrices were considered: Jaccard (A), Bray-Curtis (B), unweighted UniFrac (C) and weighted UniFrac (D). All samples from a given participant tended to group together (as shown by ellipses, one per participant). The “gold standard”, or reference methods (two samples without solution, directly frozen), were highlighted by bigger points and most points that were far from their reference correspond to storages at room temperature or high degree (+30°C).

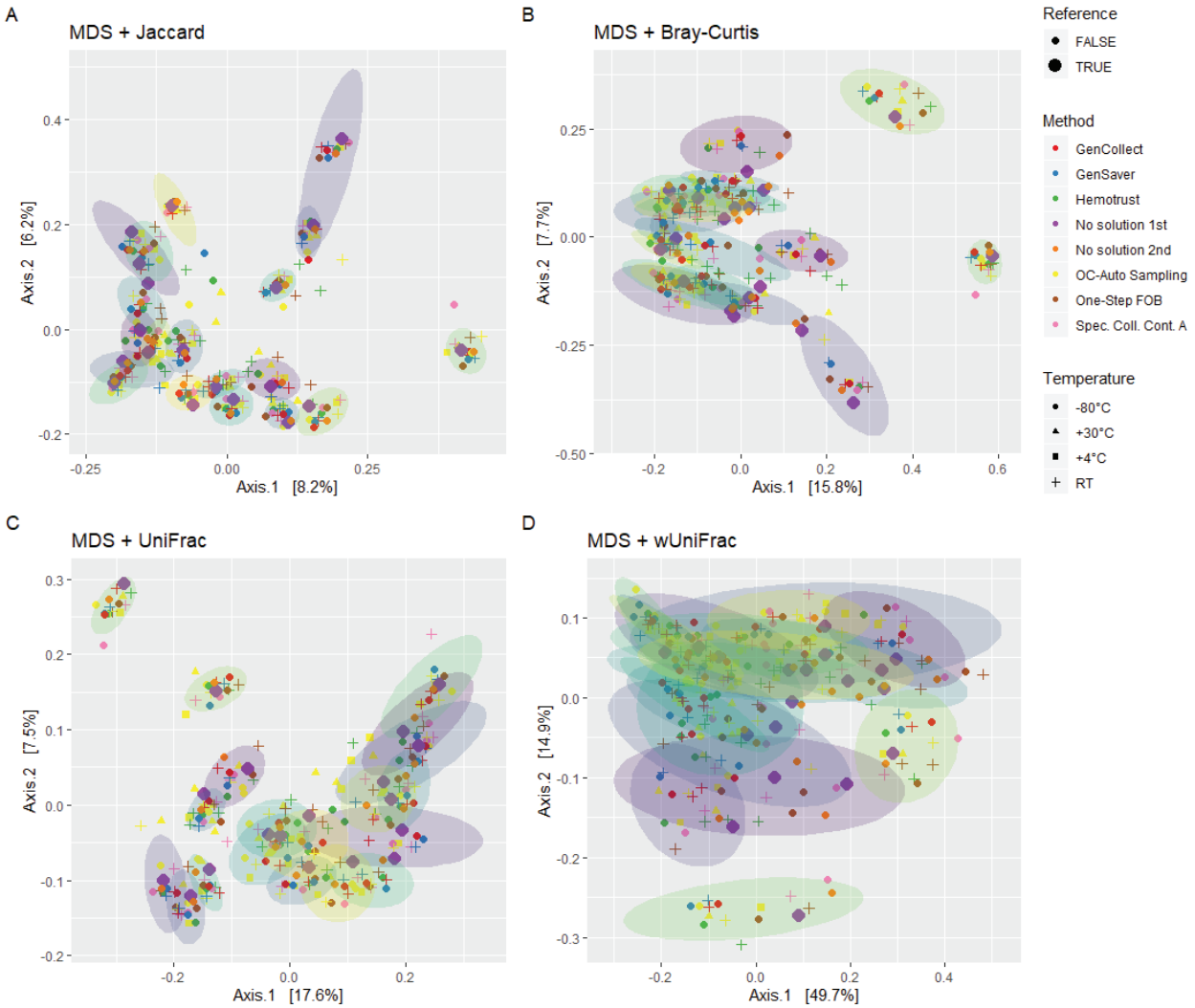


Figure S3 – Percent variability explained by individuals and collection methods using a distance-based coefficient of determination (R^2) for four beta diversity estimates. Based on Bray-Curtis, Jaccard, weighted and unweighted UniFrac distance matrices, inter-participant variability (individual, grey) explained 79%, 54.7%, 64.1% and 71.3% of variability, respectively; whereas the collection method (protocol, orange) explained 6.9%, 4.8%, 4.7% and 14.8% of variability, respectively.

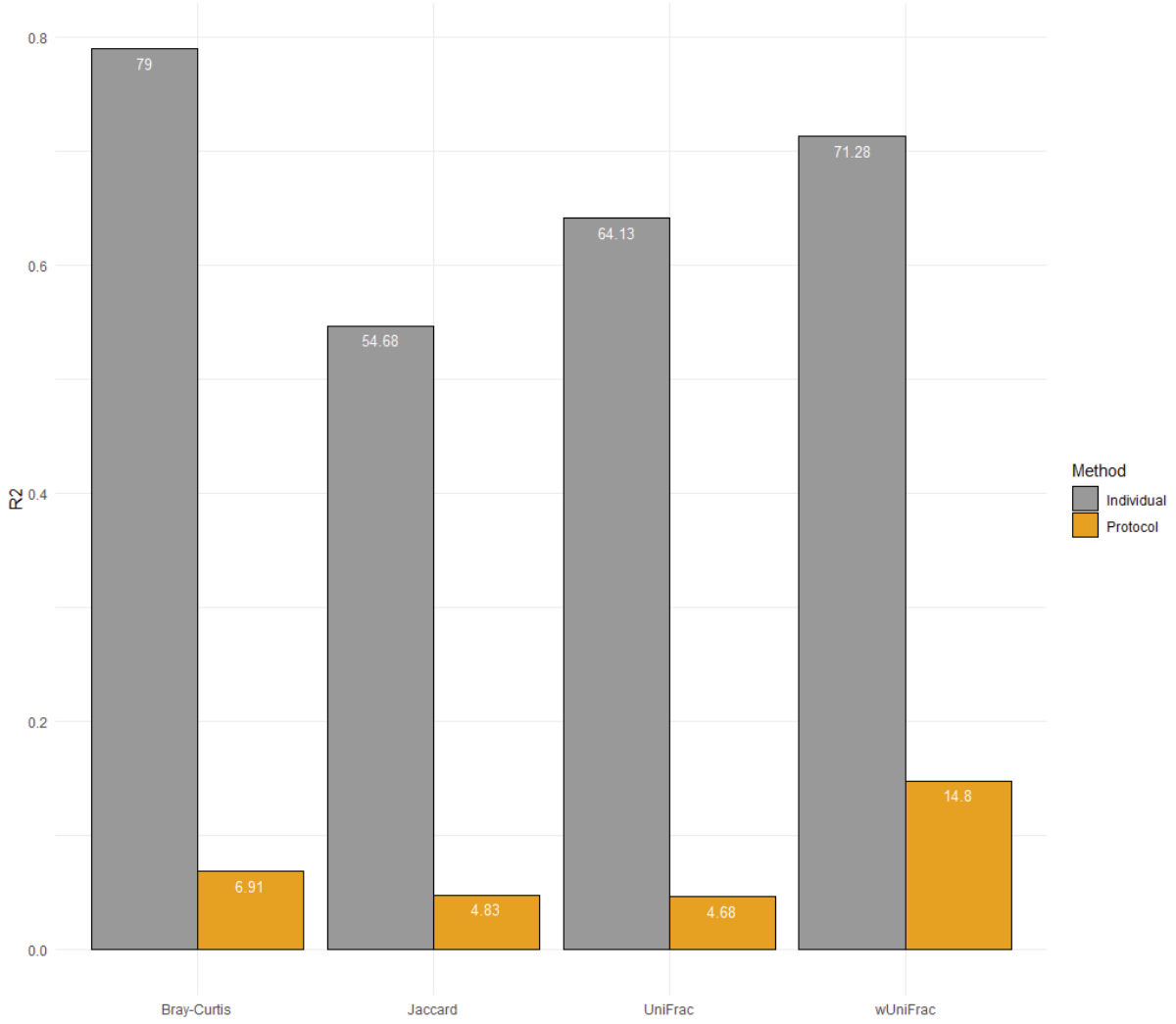


Figure S4 - Relative abundance of bacterial phylum (A) and genus (B) present in at least 50% of fecal samples with relative abundance of ≥0.1% in each collection methods, by participant.

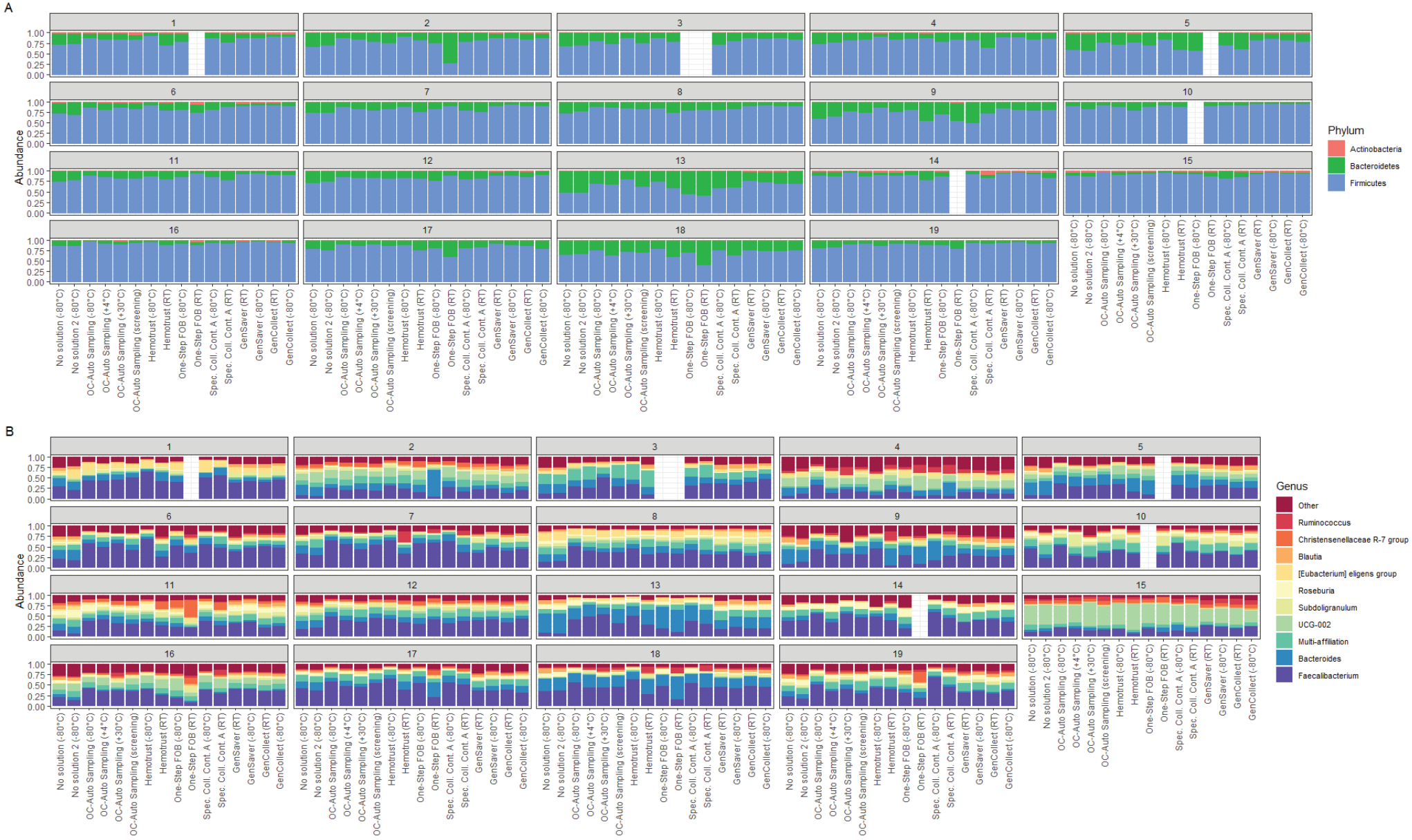
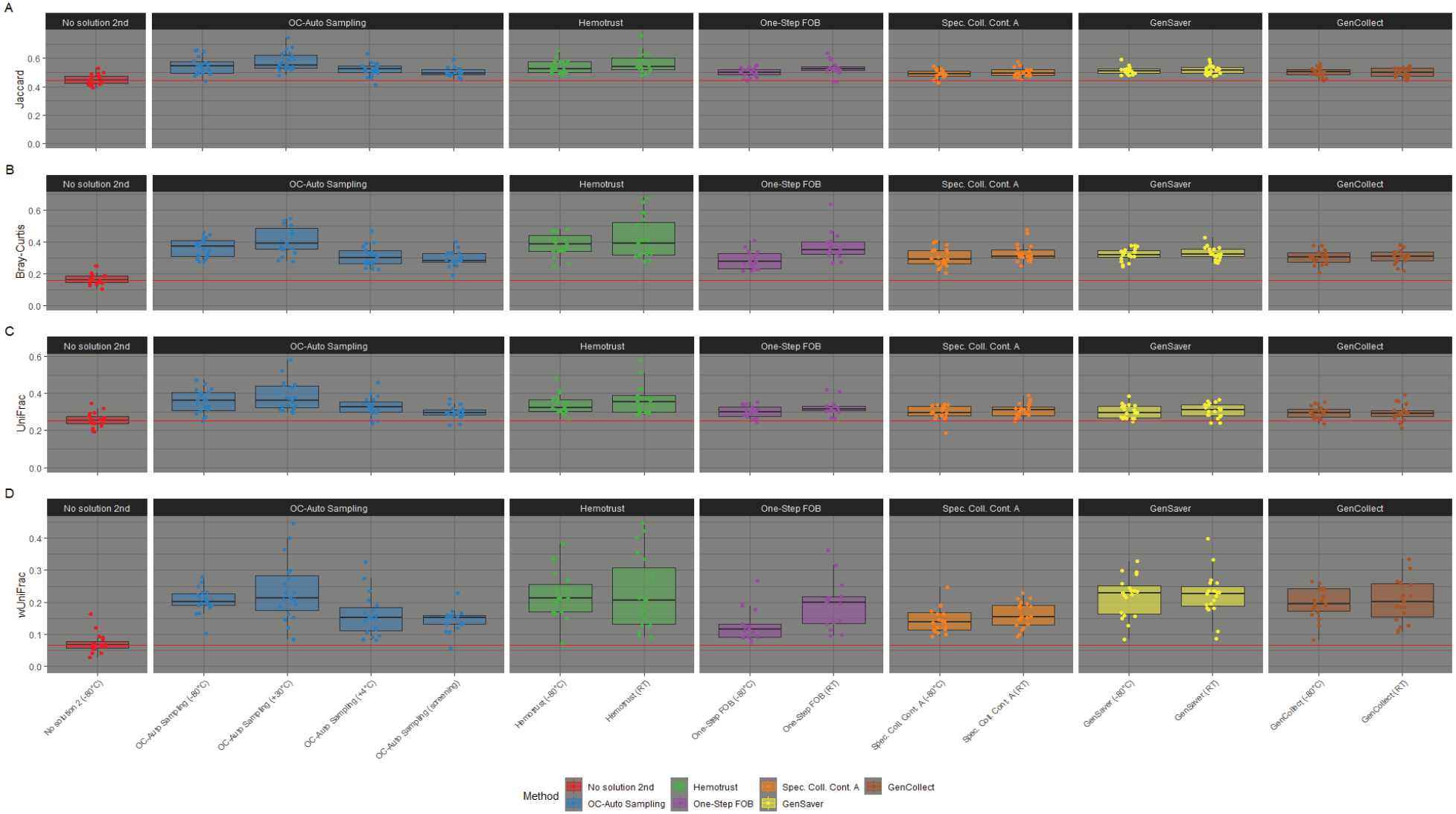


Figure S5 – Distance between the gold standard and each collection method based on Jaccard (A), Bray-Curtis (B), unweighted UniFrac (C) and weighted UniFrac (D) distance matrices. Technical variability was quantified based on the distance between the two replicates of the gold standard and is highlighted by the red line.



Tables

Table S1 – Characteristics of study participants (n=19).

Characteristic	N (%)	Mean (± SD)
Female	15 (78.9)	
Age, years		40.2 ± 8.43
Body Mass Index, kg/m^2		22.7 ± 3.14
No weight variation in the past 6 months	14 (73.7)	
Bowel movement at least once/day	18 (94.8)	

Table S2d – Results from qPCR, means for all collection methods.

Sample name	Bacterial DNA concentration (ng/µl)	Sample mass (mg)	Bacterial quantity/g	log(Bacterial quantity/g)	Bacterial quantity/slices	log(Bacterial quantity/slices)	Bacterial quantity/mL	log(Bacterial quantity/mL)
Mean cryotube1 (-80°C)	87.95	284.47						
Mean cryotube2 (-80°C)	62.78	131.79	9.99E+10	10.90				
Mean GenSaver (-80°C)	43.12				4.39E+10	10.58		
Mean GenSaver (RT)	36.62				1.92E+10	10.21		
Mean GenCollect (-80°C)	48.24				4.03E+10	10.52		
Mean GenCollect (RT)	38.03				3.56E+10	10.36		
Mean OC-Auto Sampling (-80°C)	4.34						2.34E+09	9.21
Mean OC-Auto Sampling (+4°C)	3.71						1.77E+09	9.06
Mean OC-Auto Sampling (+30°C)	3.40						1.56E+09	9.11
Mean OC-Auto Sampling (screening)	3.13						7.28E+08	8.75
Mean Hemotrust (-80°C)	3.30						5.37E+09	9.66
Mean Hemotrust 19 (RT)	3.59						3.48E+09	9.37
Mean Specimen Collection Container A (-80°C)	1.75						9.09E+08	8.83
Mean Specimen Collection Container A (RT)	1.71						3.64E+08	8.45
Mean One-Step FOB (-80°C)	1.63						4.20E+09	9.54
Mean One-Step FOB (RT)	-0.23						2.75E+09	9.26

Table S3 – Adjusted means for observed OTUs, the Shannon index and the inverse Simpson index by fecal sample collection method.

Collection method	Observed OTUs			Shannon index			Inverse Simpson index		
	Adj. Mean	Lower 95% CI	Upper 95% CI	Adj. Mean	Lower 95% CI	Upper 95% CI	Adj. Mean	Lower 95% CI	Upper 95% CI
No solution (-80°C)	296	257	335	4.03	3.81	4.24	26	21.5	30.5
No solution 2 (-80°C)	294	255	333	3.97	3.75	4.18	24.3	19.8	28.8
GenCollect (-80°C)	340	300	379	3.95	3.73	4.17	21.9	17.4	26.4
GenCollect (RT)	339	299	378	3.99	3.78	4.21	23	18.5	27.5
GenSaver (-80°C)	337	297	376	3.93	3.71	4.14	19.4	14.9	23.9
GenSaver (RT)	347	308	387	4	3.78	4.21	21.4	16.9	26
Hemotrust (-80°C)	331	292	371	3.64	3.42	3.85	14.3	9.8	18.8
Hemotrust (RT)	319	280	358	3.76	3.54	3.97	17.2	12.7	21.7
OC-Auto Sampling (-80°C)	344	305	383	3.73	3.51	3.94	14.8	10.3	19.3
OC-Auto Sampling (+30°C)	338	298	377	3.92	3.71	4.14	19.2	14.7	23.8
OC-Auto Sampling (+4°C)	324	284	363	3.81	3.59	4.03	18.9	14.4	23.4
OC-Auto Sampling (screening)	312	273	351	3.8	3.58	4.01	19.5	15	24
One-Step FOB (-80°C)	300	260	339	3.76	3.54	3.98	19	14.5	23.6
One-Step FOB (RT)	253	213	294	3.49	3.27	3.72	15.4	10.8	20.1
Spec. Coll. Cont. A (-80°C)	285	246	324	3.6	3.38	3.82	15.2	10.7	19.7
Spec. Coll. Cont. A (RT)	266	227	306	3.7	3.48	3.91	18.6	14.1	23.1

Table S4 – ICCs for accuracy of microbiome diversity metrics of each fecal sample collection method compared to the "gold-standard" (samples with no solution, -80°C frozen immediately).

Collection method	Actinobacteria			Bacteroidetes			Firmicutes			Observed OTUs			Shannon index		
	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI
No solution 2nd (-80°C)	0.89	0.80	0.99	0.95	0.90	0.99	0.92	0.85	0.99	0.95	0.90	0.99	0.95	0.90	0.99
OC-Auto Sampling (-80°C)	0.32	0.09	0.73	0.51	0.17	0.85	0.37	0.03	0.76	0.66	0.40	0.92	0.67	0.42	0.92
OC-Auto Sampling (+4°C)	0.65	0.38	0.91	0.62	0.34	0.90	0.51	0.17	0.84	0.84	0.71	0.97	0.72	0.50	0.94
OC-Auto Sampling (+30°C)	0.81	0.65	0.97	0.31	0.14	0.58	0.45	0.01	0.46	0.40	0.02	0.78	0.77	0.59	0.96
OC-Auto Sampling (screening)	0.80	0.64	0.96	0.77	0.58	0.95	0.68	0.44	0.93	0.84	0.70	0.97	0.68	0.44	0.93
Hemotrust (-80°C)	0.28	0.14	0.70	0.24	0.19	0.67	0.42	0.03	0.49	0.74	0.53	0.95	0.57	0.26	0.88
Hemotrust (RT)	0.66	0.40	0.92	0.27	0.15	0.70	0.47	0.11	0.82	0.31	0.11	0.72	0.32	0.13	0.58
One-Step FOB (-80°C)	0.45	0.06	0.83	0.84	0.70	0.98	0.82	0.66	0.97	0.91	0.83	0.99	0.70	0.46	0.94
One-Step FOB (RT)	0.78	0.59	0.97	0.60	0.27	0.92	0.61	0.30	0.93	0.60	0.28	0.92	0.29	0.19	0.76
Spec. Coll. Cont. A (-80°C)	0.49	0.14	0.84	0.84	0.71	0.98	0.79	0.63	0.96	0.89	0.79	0.98	0.51	0.17	0.85
Spec. Coll. Cont. A (RT)	0.67	0.43	0.92	0.72	0.50	0.94	0.57	0.26	0.88	0.84	0.70	0.97	0.65	0.38	0.91
GenSaver (RT)	0.75	0.55	0.95	0.26	0.18	0.62	0.36	0.10	0.55	0.74	0.53	0.95	0.84	0.70	0.97
GenSaver (-80°C)	0.83	0.69	0.97	0.26	0.16	0.69	0.27	0.17	0.61	0.81	0.64	0.97	0.81	0.65	0.97
GenCollect (RT)	0.83	0.68	0.97	0.25	0.18	0.68	0.26	0.18	0.62	0.78	0.60	0.96	0.88	0.77	0.98
GenCollect (-80°C)	0.80	0.64	0.96	0.33	0.08	0.74	0.29	0.13	0.71	0.77	0.59	0.96	0.79	0.62	0.96

Collection method	Inverse Simpson			Jaccard PC1			Bray-Curtis PC1			UniFrac PC1			wUniFrac PC1		
	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI
No solution 2nd (-80°C)	0.90	0.81	0.99	0.98	0.97	1.00	0.96	0.93	1.00	0.99	0.99	1.00	0.93	0.87	0.99
OC-Auto Sampling (-80°C)	0.23	0.20	0.66	0.86	0.75	0.98	0.92	0.84	0.99	0.91	0.83	0.99	0.81	0.66	0.97
OC-Auto Sampling (+4°C)	0.43	0.06	0.80	0.93	0.86	0.99	0.92	0.84	0.99	0.97	0.94	1.00	0.84	0.71	0.97
OC-Auto Sampling (+30°C)	0.45	0.09	0.81	0.83	0.68	0.97	0.88	0.77	0.98	0.77	0.58	0.95	0.66	0.41	0.92
OC-Auto Sampling (screening)	0.51	0.17	0.84	0.98	0.96	1.00	0.89	0.79	0.98	0.97	0.95	1.00	0.92	0.84	0.99
Hemotrust (-80°C)	0.25	0.18	0.68	0.96	0.93	1.00	0.93	0.87	0.99	0.97	0.94	1.00	0.74	0.53	0.94
Hemotrust (RT)	0.26	0.16	0.69	0.84	0.70	0.97	0.96	0.92	1.00	0.78	0.60	0.96	0.46	0.09	0.82
One-Step FOB (-80°C)	0.64	0.36	0.92	0.96	0.92	1.00	0.93	0.86	0.99	0.98	0.97	1.00	0.99	0.75	1.00
One-Step FOB (RT)	0.27	0.22	0.75	0.91	0.83	0.99	1.00	0.64	1.00	0.80	0.38	1.00	0.76	0.32	1.00
Spec. Coll. Cont. A (-80°C)	0.39	0.00	0.78	0.97	0.95	1.00	0.94	0.89	0.99	0.98	0.97	1.00	0.87	0.75	0.98
Spec. Coll. Cont. A (RT)	0.50	0.16	0.84	0.96	0.92	1.00	0.94	0.88	0.99	0.97	0.94	1.00	0.84	0.71	0.97
GenSaver (RT)	0.56	0.24	0.87	0.97	0.94	1.00	0.92	0.85	0.99	0.98	0.96	1.00	0.78	0.60	0.96
GenSaver (-80°C)	0.49	0.15	0.84	0.97	0.94	1.00	0.93	0.87	0.99	0.98	0.96	1.00	0.83	0.68	0.97
GenCollect (RT)	0.73	0.51	0.94	0.98	0.96	1.00	0.96	0.93	1.00	0.99	0.97	1.00	0.74	0.54	0.95
GenCollect (-80°C)	0.61	0.33	0.90	0.98	0.97	1.00	0.94	0.89	0.99	0.98	0.97	1.00	0.76	0.57	0.95

Table S5 – ICCs for stability of microbiome diversity metrics of each fecal sample collection method stored in different conditions compared to those directly frozen

Collection method	Actinobacteria			Bacteroidetes			Firmicutes			Observed OTUs			Shannon index		
	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI
OC-Auto Sampling (+4°C)	0.79	0.62	0.96	0.87	0.76	0.98	0.84	0.71	0.98	0.65	0.39	0.91	0.81	0.66	0.97
OC-Auto Sampling (+30°C)	0.41	0.04	0.79	0.69	0.45	0.93	0.72	0.50	0.94	0.61	0.32	0.89	0.68	0.44	0.93
OC-Auto Sampling (screening)	0.65	0.39	0.91	0.85	0.72	0.98	0.80	0.64	0.96	0.85	0.73	0.98	0.88	0.78	0.98
Hemotrust (RT)	0.55	0.23	0.87	0.65	0.39	0.91	0.31	0.10	0.72	0.45	0.08	0.81	0.12	0.33	0.57
One-Step FOB (RT)	0.21	0.31	0.73	0.72	0.47	0.97	0.58	0.24	0.93	0.80	0.61	0.98	0.72	0.47	0.97
Spec. Coll. Cont. A (RT)	0.58	0.27	0.88	0.82	0.67	0.97	0.66	0.40	0.92	0.86	0.74	0.98	0.87	0.76	0.98
GenSaver (RT)	0.91	0.84	0.99	0.95	0.91	0.99	0.96	0.92	1.00	0.95	0.90	0.99	0.93	0.87	0.99
GenCollect (RT)	0.93	0.86	0.99	0.91	0.84	0.99	0.89	0.79	0.98	0.93	0.86	0.99	0.96	0.92	1.00

Collection method	Inverse Simpson			Jaccard PC1			Bray-Curtis PC1			UniFrac PC1			wUniFrac PC1		
	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI	ICC	Lower 95%CI	Upper 95%CI
OC-Auto Sampling (+4°C)	0.78	0.60	0.96	0.82	0.67	0.97	0.92	0.84	0.99	0.83	0.69	0.97	0.90	0.82	0.99
OC-Auto Sampling (+30°C)	0.56	0.24	0.87	0.65	0.39	0.91	0.90	0.81	0.99	0.75	0.55	0.95	0.75	0.54	0.95
OC-Auto Sampling (screening)	0.76	0.56	0.95	0.85	0.72	0.98	0.90	0.82	0.99	0.94	0.89	0.99	0.89	0.79	0.98
Hemotrust (RT)	0.06	0.40	0.51	0.87	0.76	0.98	0.94	0.89	0.99	0.81	0.66	0.97	0.79	0.61	0.96
One-Step FOB (RT)	0.66	0.37	0.95	0.96	0.92	1.00	1.00	0.69	1.00	0.87	0.43	1.00	0.80	0.61	0.98
Spec. Coll. Cont. A (RT)	0.80	0.64	0.96	0.95	0.90	0.99	0.94	0.89	0.99	0.97	0.95	1.00	0.83	0.68	0.97
GenSaver (RT)	0.95	0.90	0.99	0.96	0.93	1.00	0.99	0.99	1.00	0.97	0.95	1.00	0.96	0.93	1.00
GenCollect (RT)	0.91	0.84	0.99	0.99	0.97	1.00	0.98	0.97	1.00	0.98	0.96	1.00	0.95	0.91	1.00

Chapter V: Discussion and perspectives

In this chapter, the main findings reported in the three result chapters (II to IV) are summarised with regard to the objectives and compared with the broader evidence. In addition, research perspectives are discussed and future recommendations formulated.

1. Discussion

The first part of the thesis aimed to evaluate the strengths and weaknesses of the current evidence on the relationship between the human microbiome and cancer in epidemiological studies (objective I). Among all the anatomical sites evaluated, the gut microbiome followed by the oral microbiome were most frequently studied in human observational studies. The majority of articles reported specific differences in microbiome diversity and/or composition between cases and control groups (including not only non-cancer patients but also tissues). Fifty studies reported significant differences in gut microbiome composition between colorectal cancer patients and various controls. Overall, there was consistent evidence that the abundance of pro-inflammatory opportunistic taxa, such as *Fusobacterium*, *Parvimonas* and *Porphyromonas*, was increased in colorectal cancer patients compared to non-cancer cases while a depletion of butyrate-producing bacteria was observed. Sixteen studies described differences in oral microbiome composition between oral cancer patients and controls. Enrichment of *Fusobacterium* and depletion of *Streptococcus* was reported in patients with oral cancer compared with controls. Interestingly, many bacteria from the oral cavity were not only associated with oral cancer but also with other cancer sites.

With respect to objective I, the discovery of consistent associations for certain taxa of the gut microbiome and colorectal cancer risk, and for the oral microbiome and oral cancer risk, represented a significant advancement in the current evidence on the relationship between the human microbiome and cancer in epidemiological studies. However, this strength was somewhat outweighed by certain methodological barriers and the overall evidence was still too weak to draw firm conclusions. For instance, other differences in the abundance of certain bacteria were found between cancer

cases and controls, but were less consistent between studies. As stated in the introduction of the thesis, these inconsistent findings are likely due to the fact that measurement and analysis of microbiome data are still very limited in the field of cancer epidemiology. Most studies on the relation between the human microbiome and cancer have employed case-control and other cross-sectional designs without repeated measurements. Thus, it remains difficult to establish the causality of the cancer–microbe associations. It is not clear whether the carcinogenic process changes the local micro-environment and creates new niches for bacteria, or if an alteration of microbiome composition and functions contribute to carcinogenesis (129,187). In addition, poor reproducibility between studies and more specifically, variations in study design and methods used to characterise the microbiome can impact the comparability between these different studies (196). Moreover, much is still unknown on microbiome stability and therefore, different sample storage strategies (259,260), laboratory methods, sterilisation procedures, sequencing strategies and various processing pipelines for the raw microbiome sequence data (261) can all impact the comparability of different studies (196,262). Standardised methods for the collection of samples, preparation and handling of samples, and bioinformatics processing of data are now needed (196). Therefore, repeated, prospectively collected samples from population-based cohort studies, using standardised protocols, are necessary for better understanding of the temporal nature of microbial associations with cancer.

The second part of the thesis aimed to investigate the association between insulin resistance and inflammation, established colorectal cancer risk factors, and the gut microbiome using population-based epidemiological data (objective II). To this end, we examined gut bacteria 16S rRNA gene sequence data with serologic levels of homeostatic model for insulin resistance, HbA1c and CRP in two populations of individuals from the Northern Finland Birth Cohort and TwinsUK. Globally, we found that higher levels of these metabolic parameters were associated with decreasing richness or diversity of microbiome composition in the two cohorts. Most importantly, this relationship between high levels of metabolic parameters and lower microbiome diversity remained statistically significant, even after control for body mass index and other health indicators. In addition, in both cohorts, insulin resistance was associated with unweighted UniFrac but not with weighted UniFrac. As seen in table 2 of the introduction, UniFrac is a measure based on phylogenetic measures of community

beta diversity. The unweighted UniFrac measure allows the detection of differences in the presence or absence of lineages of bacteria in different communities (263). On the other hand, weighted UniFrac is a quantitative variant of UniFrac and helps detecting differences in abundance even when the overall groups of organisms that are present in each sample remain the same. In other words, when considering the full microbial community, we found that insulin resistance was associated with differences in the presence or absence of bacteria in different communities but not with differences in abundance. We have also investigated the association between insulin resistance and specific genera using the zero-inflated beta regression model in which, the presence/absence of the taxon in the samples is modelled with a logistic component and the non-zero abundance of the taxon is modelled with a beta component (264). Using these indicators, we found that insulin resistance was associated with the relative abundance of specific genera rather than their presence or absence.

In regard to objective II, the Northern Finland Birth Cohort and TwinsUK data sets are currently quite unique in that fecal samples had been systematically collected and despite their cross-sectional design, by utilising microbiome data obtained from these samples we were able to provide evidence that support previous experimental observations on the link between metabolic health and the gut microbiome. Our findings suggest that there is a relationship between the gut microbiome and metabolic health that is independent of obesity and other measured factors. However, due to the cross-sectional nature of these two studies, it is impossible to conclude from our findings if these differences in microbial composition and taxa associations are a cause or a consequence of metabolic dysfunction. This relationship is likely complicated by the fact that the interactions between the gut microbiome, the host metabolism, the host immune system and potential associated chronic diseases, such as obesity, could be multidirectional or even cyclical. Conclusions from both objective I and objective II indicate that epidemiological findings from cross-sectional population-based studies should encourage the development of prospective studies with longitudinal collections of fecal and blood samples to further interrogate the causal nature of the association between colorectal cancer risk factors and the gut microbiome.

The third part of the thesis aimed to evaluate the impact of fecal sample collection methods used in ongoing colorectal cancer screening programs on the gut microbiota

(objective III). We investigated seven methods of fecal sample collection gathered from nineteen volunteers. These methods included no solution, two different specimen collection cards (recently used by an IARC-based study in Afghanistan), and four types of FIT used in ongoing colorectal cancer screening programs in France (and most other European countries), Morocco, Iran and Turkey. Stability of fecal microbial profiles after storage for 7 days at room temperature, 4°C, 30°C and after screening for the presence of occult blood in the stool, and accuracy compared to samples frozen immediately with no solution were estimated. Reassuringly, the major source of variation of fecal microbial profiles using 16S rRNA gene sequencing was the between individual variability, followed by between collection methods. Overall, accuracy of beta diversity measures was generally very high when compared to samples frozen immediately with no solution. However, these collection methods differed in the relative abundance of various phyla and genera, and alpha diversity measures. In addition, microbial profile stability was very high for specimen collection cards and seemed generally acceptable for FIT tubes, except for Hemotrust tubes stored at room temperature and OC-Auto Sampling tubes stored at 30°C. Most importantly, our findings indicated that passage through colorectal cancer screening tests did not impact microbiome stability in FIT tubes.

With respect to objective III, the evaluation of these fecal sample collection methods used in ongoing colorectal cancer screening programs indicates that opportunistic collection of fecal samples in FIT tubes after colorectal cancer screening is feasible. This supports the hypothesis stated in the introduction that colorectal cancer screening registries in which millions of FIT samples are being collected each year, might provide rich opportunities to establish cohorts of individuals with pre-diagnostic collected fecal samples. Our study was informative on several levels since it showed that most, but not all, fecal sample collection methods were viable for microbiome analysis, and therefore provided information that researchers may wish to consider in planning and executing future prospective studies. The variations in accuracy and mostly in stability between the different types of FIT tubes for the relative abundance of various bacteria highlighted the importance of using at least one consistent method for study comparisons. As seen in objective I, conducting pooled analyses or meta-analyses is not possible if individual studies collect fecal samples using different methods. However, to attain adequate power to detect disease associations with microbiome

profiles, microbiome data will likely have to be pooled across multiple studies. In addition, variations in accuracy and mostly in stability were observed between the different types of FIT tubes. A large variety of FIT tubes from different manufacturers are available on the market, and each FIT buffer might have different DNA-stabilizing and anti-microbial properties. The most commonly employed OC-Auto Sampling FIT seems to be reliable for microbiome analysis, however other types, such as Hemotrust tubes, revealed limitations in terms of stability at room temperature. Moreover, the low microbial profile stability of OC-Auto Sampling tubes stored at 30°C indicated that seasonal variation should be taken into account when collecting fecal samples from colorectal cancer screening programmes. Although opportunistic collection of fecal samples from populations using colorectal cancer screening methodologies is a cost-effective option that might facilitate the inclusion of participants and the collection of repeated specimens, effort should be granted to coordinate protocols across studies for the use of similar types of collection methods.

2. Research perspectives

With respect to the findings of this thesis, future research aiming at improving the understanding of the role of the intestinal microbiota in colorectal cancer development should consider the following challenges and opportunities.

One major limitation identified from the first part of this thesis was that it was not possible to provide quantitative results from meta-analysis due incomparable results from individual studies. As previously indicated, comparing published data is complicated by lack of standard processing and analysis method. Prior meta-analyses of case-control gut microbiome studies focusing on obesity, inflammatory bowel disease or colorectal cancer have been conducted, however, results were not always consistent (265–268). For example, a meta-analysis from multiple microbiome studies to identify microbial markers associated with colorectal cancer revealed that the ability to detect certain microbiome-colorectal cancer associations was limited due to differences across studies in terms of sample collection, DNA extraction methods and 16S rRNA gene sequencing region as characterised by samples clustering primarily by their original studies rather than colorectal cancer case-control status (268). In addition, these meta-analyses have been limited by focusing on one or two diseases,

and therefore, focusing on a broader landscape of diseases, such as different cancer sites, is even more complicated. Pending the implementation of collaborative microbiome cohorts with global standardisation, the sharing of microbiome data along with metadata should be encouraged to allow members of the scientific community to replicate, confirm and build on individual existing findings (261,262). However, if collecting raw amplicon sequencing data sets could permit systematic re-processing and re-analyse of the data in a way that allows meta-analysis, many other possible sources of variation should still be considered, such as experimental and sequencing artefacts or host confounders (54,56).

Findings from the second part of this thesis revealed an interesting statistical consideration in the analysis and interpretation of microbiome data. As mentioned above, in both the Northern Finland Birth Cohort and TwinsUK, we found that insulin resistance was associated with differences in the presence or absence of bacteria in different communities but not with differences in abundance when using weighted and unweighted UniFrac beta diversity measures. On the other hand, when using a zero-inflated beta regression model, we found that insulin resistance was associated with the relative abundance of specific genera rather than their presence or absence. One explanation could be that the association between insulin resistance and microbiome diversity captured by beta diversity measures was likely driven by a consortium of bacteria present at low relative abundance, as beta diversity matrices take into account the entire microbial community while the zero-inflated beta regression model directly addresses the distribution of single bacteria (269). Another explanation might be that the specific lineages that are driving the association with insulin resistance were excluded from the zero-inflated beta regression model due to very low relative abundance or prevalence since this model directly addresses the distribution of genera relative abundance that resemble a zero-inflated beta distribution (264). Finally, one other explanation could be that beta diversity matrices were computed at the ASVs level and not at the genus level. Microbiome data are high-dimensional structured multivariate sparse data and because of their compositional nature, microbiome data analysis can be challenging (270,271). The Northern Finland Birth Cohort and TwinsUK data could provide opportunities to evaluate current methodologies employed in microbiome statistical analysis and could help to develop new standardised methods for future studies.

In the Northern Finland Birth Cohort analysis, a potentially important finding was that the relationship between higher levels of insulin resistance and lower microbiome diversity was independent of body mass index suggesting that the observed relationship between the body mass index and the gut microbiome diversity might also be influenced by measures of insulin resistance. In addition, although this finding were only observed in the Northern Finland Birth Cohort, the association of body mass index and microbiome diversity was attenuated when adjusting the model for insulin resistance, supporting that some of the body mass index-microbiome relation might be driven through insulin resistance. However, this hypothesis needs to be further explored. One possibility would be to stratify the data by body mass index and verify if the association between insulin resistance and gut microbiome diversity is still observed in lean individuals. However, this analysis would require a larger sample size as only few participants in our study had normal weight with high insulin resistance levels or were obese with low insulin resistance levels. A recent study from Korea evaluated differences in microbiota composition between 317 metabolically healthy and 430 metabolically unhealthy obese individuals (272). Significant differences in gut microbiome composition was observed between the two groups and metabolically unhealthy obese individuals had lower microbial diversity compared to metabolically healthy obese controls. Although these results should be interpreted carefully due to the cross-sectional design of the study and the selection of obese individuals only, these findings support the hypothesis that there is a link between the gut microbiota and metabolic health independent of obesity. Another study, using data from the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort, has shown that metabolic health (characterised by insulin levels) might be a risk factor for colorectal cancer independent of obesity (203). However, the etiology of poor metabolic heath in lean phenotype is still unknown and therefore, makes the prevention of associated diseases such as colorectal cancer difficult. One potential explanation could be that alterations of the gut microbiome could interfere with intestinal permeability, increasing the absorption of LPS which initiates activation of Toll-like receptors, inducing increased inflammation and impairment of insulin signalling pathways (273). Understanding the relationship between the gut microbiome and insulin resistance in lean persons could open up new research lines for the prevention of colorectal cancer by targeting the gut microbiome through diet and drugs.

The Northern Finland Birth Cohort represents a valuable platform for future microbiome analysis as many different types of data have been collected. For instance, global biochemical profiles were also measured using a non-targeted metabolomics approach in a subset of fecal samples and have been explored in relation to obesity (204). The availability of both microbiome and metabolomics data will allow us to explore connections between the gut microbiome and fecal metabolome. Such integrated analysis will provide further insight into the likely complex relationship that we observed between gut microbiome profiles and markers of metabolic health. As mentioned in the introduction, 16S rRNA gene sequencing data does not provide sufficient resolution for species-level analyses and therefore, insights into the functional relationships between the gut microbiome and the host remain challenging. The use of metabolomics in the Northern Finland Birth Cohort might help to infer the functional status of host-microbial relationships in fecal specimens (204–206). Indeed, while 16S rRNA gene profiling provides information on gut microbiome composition and diversity, metabolomics could additionally describe microbial activity. We plan to run such microbiome-metabolomics analyses in the Northern Finland Birth Cohort, to better understand the link between metabolic health parameters and the gut microbiome.

In addition to fecal samples, serum samples are also available in the Northern Finland Birth Cohort and as such, potential associations between metabolic profiles in fecal and serum samples could be analysed and compared with gut microbiome composition. Previous studies using mouse models have shown that the accumulated levels of gut microbial metabolites in the intestinal tract does not always correspond with that in the host circulation (207,208). Another extension of our work could consist of parallel analyses of fecal and serum metabolites to help understand the biosynthetic pathways of gut microbial metabolites under different pathophysiological states of the host or identify gut microbiome-derived metabolites.

The observed association of microbiome diversity measures with serologic levels of metabolic health biomarkers in the Northern Finland Birth Cohort may have also been inflated or attenuated by unmeasured or poorly measured confounders, such as dietary intake. Given the diversity of foods in diets, accuracy of dietary exposure is always challenging in free-living populations, and some degree of measurement error is inevitable due to the inaccurate identification of different food components (209,210).

In the Northern Finland Birth Cohort, dietary questionnaire data and fecal metabolomics, along with serum metabolomics, could be analysed to identify metabolites related to specific food components or food patterns and assess associations of these markers with gut microbiome composition and diversity, and metabolic health parameters (211,212). We plan to run this analysis in the next phase of our research.

It is important to place the results from the third part of this thesis in context with cost and feasibility for implementation of fecal sample collection in large-scale studies and use for other –omics analyses. For example, immediately freezing samples with no solution at -80°C is likely to be infeasible in larger population-based studies and, in absence of DNA stabilising agents, their stability when left at room temperature for longer periods of time is questionable (189,213–216). In addition, although considered as the “gold standard”, it is difficult to ascertain whether the results seen with these samples is representative of the actual microbiome composition in the intestinal tract of the host, especially since bacterial growth is possible during freezing/thawing processes or collection conditions (e.g. self-collection at home) (217–219).

On the other hand, antimicrobial agents in the buffer allows DNA stabilisation which confer acceptable microbial stability for FIT tubes stored at room temperature (194,234,242,243), but not at high temperatures. However, FIT tubes also include stabilising solutions which are proprietary and stability may differ by manufacturers, as described in this thesis. In addition, FIT tubes have previously been found to be suboptimal for metabolomics studies, especially when stored at ambient temperature (224). Laboratory practices should also be considered as the implementation in large-scale studies might require automatization of certain processes. For example, we found that due to the rigid plastic frame of certain type of FIT tubes, manual buffer extraction was complicated. Based on qPCR data, our findings indicated that it is unnecessary to extract the complete volume of the solution to obtain sufficient DNA which might be logistically easier in population-based studies. In addition, depending on the manufacturer, FIT tubes are of different sizes and some can be quite large, which could be another challenge to address if thousands of samples need to be stored at -80°C in future population-based studies.

In our study, the specimen collection cards stored at room temperatures for ten weeks showed excellent stability when compared to the immediately frozen cards, as previously described (230,233,235–240). In addition, they also have the advantage to be easily transportable and storable, and cheap, which are important considerations in the context of setting up large-scale population-based studies. Specimen collection cards were also found to be optimal for metabolomics studies, as they were reproducible, stable, and accurate in different studies (233,288).

Further analyses could be conducted to evaluate these collection methods since additional samples frozen immediately at -80°C with no solution are stored at IARC. For example, effects of long-term storage without buffer can be tested as microbiome composition could be investigated over long periods in many future population-based prospective cohorts, and new analyses may be required after several years. The effect of long-term freezing on fecal microbiome stability has previously been investigated for storage at -80°C for more than two years (282) and for storage at -20°C for more than 14 years (289). In these studies, long-term storage seemed to induce few significant changes in microbial community, with few variations in some relative abundances of specific taxa but similar microbiome diversity measurements, confirming stability of the fecal microbiome during long periods of freezing, even for freezing conditions at -20°C. The replication and validation of these results using our samples would be interesting, even if cryopreservation protocols, such as the time and storage conditions until freezing as well as the freezing process itself, should not be ignored (247). In addition, technical reproducibility of our results could also be performed by re-processing the samples frozen immediately at -80°C with no solution in another laboratory, taking account of the potential effect of longer storage conditions. Finally, the demultiplexed 16S rRNA gene sequencing data could also be used to compare the impact of different bioinformatics pipelines and processes on the results. For example, comparison of Swarm (253) with some of the most widely-used denoising methods, such as DADA2 (23) and Deblur (290), could be run on an artificial community to estimate alpha diversity measures and microbial composition and then replicated in our dataset on host-associated communities.

Previous methodological work, in addition to our findings indicating that detection of occult blood in stool samples does not impact microbiome accuracy and stability in FIT

tubes, now serve as a proof-of-concept for the establishment of an international microbiome network that has been initiated by investigators from IARC and the National Cancer Institute (USA), along with researchers with expertise in the microbiome and representatives of colorectal cancer screening programs from different countries. One project from this international consortium that is already ongoing is a study of FIT-collected fecal samples in the colorectal cancer screening program of the Piedmont region in Italy. Gut microbiome profiles will be characterised and compared using shallow shotgun sequencing in screened-individuals who were FIT-positive and diagnosed with colorectal neoplasms, or FIT-positive but colorectal neoplasm-free after colonoscopy, or FIT-negative. Another aspect of this network comprises the implementation of feasibility studies nested within various colorectal cancer screening programs to evaluate the achievability of establishing a large multi-country prospective cohort of individuals with repeated fecal microbiome samples as well as epidemiologic and other clinical data.

3. Conclusion

This thesis has contributed to the investigation of the role of the gut microbiome in colorectal cancer using epidemiological approaches. The systematic review of the current evidence linking the human microbiome with cancer stressed that prospective studies with standardised methods are now needed to quantify the potential effect of the gut microbiome on cancer risk. Although, studying the direct relationship between the gut microbiome and colorectal cancer risk is currently not possible due to the lack of prospective studies with prediagnostic fecal samples, this thesis provided informative insights on the association between the gut microbiome and some parameters of metabolic health that are recognised as colorectal cancer risk factors. Finally, the evaluation of microbiome stability and accuracy in fecal samples collected using different methods employed in ongoing colorectal cancer screening programs indicated that commonly used fecal sample collections such as fecal immunotests and paper-based collection cards are, in general, suitable media for microbiome measurements though storage factors such as ambient temperature can impact on stability for some methods. In addition, the opportunistic collection of fecal samples in fecal immunochemical test tubes after colorectal cancer screening is likely a viable method for establishing cohorts with prediagnostic fecal specimens. In conclusion, this

thesis presents the state-of-the-art on epidemiological evidence for the role of the microbiome in tumorigenesis, provides novel insights on the association of metabolic risk factors for colorectal cancer with the gut microbiome in population-based studies, and finally has generated important methodological data on the impact of fecal sample collection tools on microbial measurements that is needed for future epidemiological research on the microbiome and colorectal cancer, as well as other chronic diseases.

References

1. Sender R, Fuchs S, Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell. 2016 Jan 28;164(3):337–40.

2. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature. 2007 Oct;449(7164):804–10.

3. Locey KJ, Lennon JT. Scaling laws predict global microbial diversity. Proc Natl Acad Sci U S A. 2016 May 24;113(21):5970–5.

4. Hooper LV, Gordon JI. Commensal host-bacterial relationships in the gut. Science. 2001 May 11;292(5519):1115–8.

5. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature. 2010 Mar 4;464(7285):59–65.

6. Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe. 2014 Sep 10;16(3):276–89.

7. Marchesi JR, Ravel J. The vocabulary of microbiome research: a proposal. Microbiome [Internet]. 2015 [cited 2020 Nov 21];3. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4520061/

8. Brenner DJ, Staley JT, Krieg NR. Classification of Procaryotic Organisms and the Concept of Bacterial Speciation. In: Bergey's Manual of Systematics of Archaea and Bacteria. American Cancer Society; 2015. p. 1–9.

9. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol Rev. 1995 Mar;59(1):143–69.

10. Hugenholtz P, Goebel BM, Pace NR. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. J Bacteriol. 1998 Sep;180(18):4765–74.

11. Suenaga H. Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. Environ Microbiol. 2012 Jan;14(1):13–22.

12. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017 Sep;35(9):833–44.

13. Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol. 1991 Jan;173(2):697–703.

14. Clarridge JE. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clin Microbiol Rev. 2004 Oct;17(4):840–62.

15. Coenye T, Vandamme P. Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. FEMS Microbiol Lett. 2003 Nov 1;228(1):45–9.

16. Pereira F, Carneiro J, Matthiesen R, van Asch B, Pinto N, Gusmão L, et al. Identification of species by multiplex analysis of variable-length sequences. Nucleic Acids Res. 2010 Dec;38(22):e203.

17. Fukuda K, Ogawa M, Taniguchi H, Saito M. Molecular Approaches to Studying Microbial Communities: Targeting the 16S Ribosomal RNA Gene. J UOEH. 2016 Sep;38(3):223–32.

18. Postollec F, Falentin H, Pavan S, Combrisson J, Sohier D. Recent advances in quantitative PCR (qPCR) applications in food microbiology. Food Microbiol. 2011 Aug;28(5):848–61.

19. Janda JM, Abbott SL. 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J Clin Microbiol. 2007 Sep;45(9):2761–4.

20. Woo PCY, Lau SKP, Teng JLL, Tse H, Yuen K-Y. Then and now: use of 16S rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2008 Oct;14(10):908–34.

21. Mignard S, Flandrois JP. 16S rRNA sequencing in routine bacterial identification: a 30-month experiment. J Microbiol Methods. 2006 Dec;67(3):574–81.

22. Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, et al. Defining operational taxonomic units using DNA barcode data. Philos Trans R Soc B Biol Sci. 2005 Oct 29;360(1462):1935–43.

23. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13(7):581–3.

24. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017 Dec;11(12):2639–43.

25. Stirling G, Wilsey B. Empirical Relationships between Species Richness, Evenness, and Proportional Diversity. Am Nat. 2001 Sep;158(3):286–99.

26. Whittaker RH. Evolution and Measurement of Species Diversity. Taxon. 1972;21(2/3):213–51.

27. Lemos LN, Fulthorpe RR, Triplett EW, Roesch LFW. Rethinking microbial diversity analysis in the high throughput sequencing era. J Microbiol Methods. 2011 Jul;86(1):42–51.

28. Simpson EH. Measurement of Diversity. Nature. 1949 Apr;163(4148):688–688.

29. Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJM. Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. Appl Environ Microbiol. 2001 Oct 1;67(10):4399–406.

30. Gotelli NJ, Colwell RK. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol Lett. 2001;4(4):379–91.

31. Bray J, Curtis J. An Ordination of the Upland Forest Communities of Southern Wisconsin. 1957;

32. Levandowsky M, Winter D. Distance between Sets. Nature. 1971 Nov;234(5323):34–5.

33. Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. ISME J. 2011 Feb;5(2):169–72.

34. Irrazábal T, Belcheva A, Girardin SE, Martin A, Philpott DJ. The multifaceted role of the intestinal microbiota in colon cancer. Mol Cell. 2014 Apr 24;54(2):309–20.

35. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. Nat Rev Microbiol. 2011 May;9(5):356–68.

36. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and Temporal Diversity of the Human Skin Microbiome. Science. 2009 May 29;324(5931):1190–2.

37. Lloyd-Price J, Mahurkar A, Rahnavard G, Crabtree J, Orvis J, Hall AB, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. Nature. 2017 Oct;550(7674):61–6.

38. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012 May 9;486(7402):222–7.

39. O'Keefe SJD, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, et al. Fat, fibre and cancer risk in African Americans and rural Africans. Nat Commun. 2015 Apr 28;6:6342.

40. Rothschild D, Weissbrod O, Barkan E, Kurilshikov A, Korem T, Zeevi D, et al. Environment dominates over host genetics in shaping human gut microbiota. Nature. 2018 Mar;555(7695):210–5.

41. Vich Vila A, Collij V, Sanna S, Sinha T, Imhann F, Bourgonje AR, et al. Impact of commonly used drugs on the composition and metabolic function of the gut microbiota. Nat Commun. 2020 Jan 17;11(1):362.

42. Collado MC, Cernada M, Neu J, Pérez-Martínez G, Gormaz M, Vento M. Factors influencing gastrointestinal tract and microbiota immune interaction in preterm infants. Pediatr Res. 2015 Jun;77(6):726–31.

43. Dong TS, Gupta A. Influence of Early Life, Diet, and the Environment on the Microbiome. Clin Gastroenterol Hepatol. 2019 Jan 1;17(2):231–42.

44. Cresci GA, Bawden E. Gut Microbiome: What We Do and Don't Know. Nutr Clin Pract Off Publ Am Soc Parenter Enter Nutr. 2015 Dec;30(6):734–46.

45. Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. J Transl Med. 2017 Apr 8;15(1):73.

46. Desai MS, Seekatz AM, Koropatkin NM, Kamada N, Hickey CA, Wolter M, et al. A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. Cell. 2016 17;167(5):1339-1353.e21.

47. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Sci Transl Med. 2009 Nov 11;1(6):6ra14-6ra14.

48. Makki K, Deehan EC, Walter J, Bäckhed F. The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. Cell Host Microbe. 2018 13;23(6):705–15.

49. Sonnenburg ED, Sonnenburg JL. Starving our microbial self: the deleterious consequences of a diet deficient in microbiota-accessible carbohydrates. Cell Metab. 2014 Nov 4;20(5):779–86.

50. Tamburini S, Shen N, Wu HC, Clemente JC. The microbiome in early life: implications for health outcomes. Nat Med. 2016 07;22(7):713–22.

51. Dethlefsen L, Relman DA. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A. 2011 Mar 15;108 Suppl 1:4554–61.

52. Francino MP. Antibiotics and the Human Gut Microbiome: Dysbioses and Accumulation of Resistances. Front Microbiol. 2015;6:1543.

53. Biedermann L, Brülisauer K, Zeitz J, Frei P, Scharl M, Vavricka SR, et al. Smoking cessation alters intestinal microbiota: insights from quantitative investigations on human fecal samples using FISH. Inflamm Bowel Dis. 2014 Sep;20(9):1496–501.

54. Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, et al. Population-level analysis of gut microbiome variation. Science. 2016 Apr 29;352(6285):560–4.

55. Le Roy CI, Wells PM, Si J, Raes J, Bell JT, Spector TD. Red Wine Consumption Associated With Increased Gut Microbiota α-Diversity in 3 Independent Cohorts. Gastroenterology. 2020;158(1):270-272.e2.

56. Vujkovic-Cvijin I, Sklar J, Jiang L, Natarajan L, Knight R, Belkaid Y. Host variables confound gut microbiota studies of human disease. Nature. 2020 Nov;587(7834):448–54.

57. Rowland I, Gibson G, Heinken A, Scott K, Swann J, Thiele I, et al. Gut microbiota functions: metabolism of nutrients and other food components. Eur J Nutr. 2018;57(1):1–24.

58. Zimmermann M, Zimmermann-Kogadeeva M, Wegmann R, Goodman AL. Mapping human microbiome drug metabolism by gut bacteria and their genes. Nature. 2019 Jun;570(7762):462–7.

59. Brestoff JR, Artis D. Commensal bacteria at the interface of host metabolism and the immune system. Nat Immunol. 2013 Jul;14(7):676–84.

60. Huttenhower C, Gevers D, Knight R, Abubucker S, Badger JH, Chinwalla AT, et al. Structure, function and diversity of the healthy human microbiome. Nature. 2012 Jun;486(7402):207–14.

61. Flint HJ, Scott KP, Louis P, Duncan SH. The role of the gut microbiota in nutrition and health. Nat Rev Gastroenterol Hepatol. 2012 Oct;9(10):577–89.

62. Shafquat A, Joice R, Simmons SL, Huttenhower C. Functional and phylogenetic assembly of microbial communities in the human microbiome. Trends Microbiol. 2014 May;22(5):261–6.

63. Hinsberger A, Sandhu BK. Digestion and absorption. Curr Paediatr. 2004 Dec 1;14(7):605–11.

64. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-Bacterial Mutualism in the Human Intestine. Science. 2005 Mar 25;307(5717):1915–20.

65. Koh A, De Vadder F, Kovatcheva-Datchary P, Bäckhed F. From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. Cell. 2016 Jun 2;165(6):1332–45.

66. Ríos-Covián D, Ruas-Madiedo P, Margolles A, Gueimonde M, de Los Reyes-Gavilán CG, Salazar N. Intestinal Short Chain Fatty Acids and their Link with Diet and Human Health. Front Microbiol. 2016;7:185.

67. A Sense of Community. In: Understanding the Gut Microbiota [Internet]. John Wiley & Sons, Ltd; 2016 [cited 2020 Oct 29]. p. 19–38. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118801413.ch3

68. Cummings JH, Macfarlane GT. Role of intestinal bacteria in nutrient metabolism. JPEN J Parenter Enteral Nutr. 1997 Dec;21(6):357–65.

69. Ulven T. Short-chain free fatty acid receptors FFA2/GPR43 and FFA3/GPR41 as new potential therapeutic targets. Front Endocrinol. 2012;3:111.

70. Kimura I, Inoue D, Hirano K, Tsujimoto G. The SCFA Receptor GPR43 and Energy Metabolism. Front Endocrinol. 2014;5:85.

71. Roberfroid M, Gibson GR, Hoyles L, McCartney AL, Rastall R, Rowland I, et al. Prebiotic effects: metabolic and health benefits. Br J Nutr. 2010 Aug;104(S2):S1–63.

72. Staels B, Fonseca VA. Bile Acids and Metabolic Regulation. Diabetes Care. 2009 Nov;32(Suppl 2):S237–45.

73. Lefebvre P, Cariou B, Lien F, Kuipers F, Staels B. Role of bile acids and bile acid receptors in metabolic regulation. Physiol Rev. 2009 Jan;89(1):147–91.

74. Swann JR, Want EJ, Geier FM, Spagou K, Wilson ID, Sidaway JE, et al. Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. Proc Natl Acad Sci. 2011 Mar 15;108(Supplement 1):4523–30.

75. Watanabe M, Houten SM, Wang L, Moschetta A, Mangelsdorf DJ, Heyman RA, et al. Bile acids lower triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c. J Clin Invest. 2004 May;113(10):1408–18.

76. Thomas C, Gioiello A, Noriega L, Strehle A, Oury J, Rizzo G, et al. TGR5-mediated bile acid sensing controls glucose homeostasis. Cell Metab. 2009 Sep;10(3):167–77.

77. Zhang G, Meredith TC, Kahne D. On the essentiality of lipopolysaccharide to Gram-negative bacteria. Curr Opin Microbiol. 2013 Dec;16(6):779–85.

78. Nikaido H. Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev MMBR. 2003 Dec;67(4):593–656.

79. Vatanen T, Kostic AD, d'Hennezel E, Siljander H, Franzosa EA, Yassour M, et al. Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. Cell. 2016 May 5;165(4):842–53.

80. d'Hennezel E, Abubucker S, Murphy LO, Cullen TW. Total Lipopolysaccharide from the Human Gut Microbiome Silences Toll-Like Receptor Signaling. mSystems [Internet]. 2017 Nov 14 [cited 2020 Oct 30];2(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5686520/

81. Yoshida N, Yamashita T, Kishino S, Watanabe H, Sasaki K, Sasaki D, et al. A possible beneficial effect of Bacteroides on faecal lipopolysaccharide activity and cardiovascular diseases. Sci Rep. 2020 Aug 3;10(1):13009.

82. Bergen WG, Wu G. Intestinal nitrogen recycling and utilization in health and disease. J Nutr. 2009 May;139(5):821–5.

83. Metges CC. Contribution of microbial amino acids to amino acid homeostasis of the host. J Nutr. 2000 Jul;130(7):1857S-64S.

84. Ma N, Ma X. Dietary Amino Acids and the Gut-Microbiome-Immune Axis: Physiological Metabolism and Therapeutic Prospects. Compr Rev Food Sci Food Saf. 2019;18(1):221–42.

85. Macfarlane GT, Allison C, Gibson SA, Cummings JH. Contribution of the microflora to proteolysis in the human large intestine. J Appl Bacteriol. 1988 Jan;64(1):37–46.

86. Dai Z-L, Wu G, Zhu W-Y. Amino acid metabolism in intestinal bacteria: links between gut ecology and host health. Front Biosci Landmark Ed. 2011 Jan 1;16:1768–86.

87. Blachier F, Mariotti F, Huneau JF, Tomé D. Effects of amino acid-derived luminal metabolites on the colonic epithelium and physiopathological consequences. Amino Acids. 2007 Nov;33(4):547–62.

88. Libao-Mercado AJO, Zhu CL, Cant JP, Lapierre H, Thibault J-N, Sève B, et al. Dietary and endogenous amino acids are the main contributors to microbial protein in the upper gut of normally nourished pigs. J Nutr. 2009 Jun;139(6):1088–94.

89. Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. Science. 2006 Jun 2;312(5778):1355–9.

90. Rossi O, van Baarlen P, Wells JM. Host-recognition of pathogens and commensals in the mammalian intestine. Curr Top Microbiol Immunol. 2013;358:291–321.

91. Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature. 2006 Dec 21;444(7122):1027–31.

92. Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, et al. A core gut microbiome in obese and lean twins. Nature. 2009 Jan 22;457(7228):480–4.

93. Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature. 2012 Oct 4;490(7418):55–60.

94. Larsen N, Vogensen FK, van den Berg FWJ, Nielsen DS, Andreasen AS, Pedersen BK, et al. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. PloS One. 2010 Feb 5;5(2):e9085.

95. Ahn J, Sinha R, Pei Z, Dominianni C, Wu J, Shi J, et al. Human gut microbiome and risk for colorectal cancer. J Natl Cancer Inst. 2013 Dec 18;105(24):1907–11.

96. Fernández MF, Reina-Pérez I, Astorga JM, Rodríguez-Carrillo A, Plaza-Díaz J, Fontana L. Breast Cancer and Its Relationship with the Microbiota. Int J Environ Res Public Health. 2018 Aug;15(8).

97. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018 Feb;14(2):88–98.

98. Taylor R. Insulin Resistance and Type 2 Diabetes. Diabetes. 2012 Apr;61(4):778–9.

99. Kahn SE, Hull RL, Utzschneider KM. Mechanisms linking obesity to insulin resistance and type 2 diabetes. Nature. 2006 Dec;444(7121):840–6.

100. Owei I, Umekwe N, Provo C, Wan J, Dagogo-Jack S. Insulin-sensitive and insulin-resistant obese and non-obese phenotypes: role in prediction of incident pre-diabetes in a longitudinal biracial cohort. BMJ Open Diabetes Res Care. 2017 Jul 1;5(1):e000415.

101. Eckel RH, Grundy SM, Zimmet PZ. The metabolic syndrome. The Lancet. 2005 Apr 16;365(9468):1415–28.

102. Saklayen MG. The Global Epidemic of the Metabolic Syndrome. Curr Hypertens Rep. 2018 Feb 26;20(2):12.

103. Cameron AJ, Shaw JE, Zimmet PZ. The metabolic syndrome: prevalence in worldwide populations. Endocrinol Metab Clin North Am. 2004 Jun;33(2):351–75, table of contents.

104. Loucks EB, Rehkopf DH, Thurston RC, Kawachi I. Socioeconomic disparities in metabolic syndrome differ by gender: evidence from NHANES III. Ann Epidemiol. 2007 Jan;17(1):19–26.

105. Boucher J, Kleinridders A, Kahn CR. Insulin Receptor Signaling in Normal and Insulin-Resistant States. Cold Spring Harb Perspect Biol [Internet]. 2014 Jan [cited 2020 Nov 22];6(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3941218/

106. Taniguchi CM, Emanuelli B, Kahn CR. Critical nodes in signalling pathways: insights into insulin action. Nat Rev Mol Cell Biol. 2006 Feb;7(2):85–96.

107. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, et al. Richness of human gut microbiome correlates with metabolic markers. Nature. 2013 Aug 29;500(7464):541–6.

108. Karlsson FH, Tremaroli V, Nookaew I, Bergström G, Behre CJ, Fagerberg B, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. Nature. 2013 Jun 6;498(7452):99–103.

109. Lee CJ, Sears CL, Maruthur N. Gut microbiome and its role in obesity and insulin resistance. Ann N Y Acad Sci. 2020 Feb;1461(1):37–52.

110. Yassour M, Lim MY, Yun HS, Tickle TL, Sung J, Song Y-M, et al. Sub-clinical detection of gut microbial biomarkers of obesity and type 2 diabetes. Genome Med. 2016 Feb 17;8(1):17.

111. Lee CC, Watkins SM, Lorenzo C, Wagenknecht LE, Il'yasova D, Chen Y-DI, et al. Branched-Chain Amino Acids and Insulin Metabolism: The Insulin Resistance Atherosclerosis Study (IRAS). Diabetes Care. 2016 Apr 1;39(4):582–8.

112. Pedersen HK, Gudmundsdottir V, Nielsen HB, Hyotylainen T, Nielsen T, Jensen BAH, et al. Human gut microbes impact host serum metabolome and insulin sensitivity. Nature. 2016 Jul;535(7612):376–81.

113. Tilg H, Moschen AR. Microbiota and diabetes: an evolving relationship. Gut. 2014 Sep;63(9):1513–21.

114. Cotillard A, Kennedy SP, Kong LC, Prifti E, Pons N, Le Chatelier E, et al. Dietary intervention impact on gut microbial gene richness. Nature. 2013 Aug 29;500(7464):585–8.

115. Gesta S, Blüher M, Yamamoto Y, Norris AW, Berndt J, Kralisch S, et al. Evidence for a role of developmental genes in the origin of obesity and body fat distribution. Proc Natl Acad Sci U S A. 2006 Apr 25;103(17):6676–81.

116. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011 May;473(7346):174–80.

117. Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, et al. Diabetes and cancer: a consensus report. Diabetes Care. 2010 Jul;33(7):1674–85.

118. Giovannucci E. Insulin, Insulin-Like Growth Factors and Colon Cancer: A Review of the Evidence. J Nutr. 2001 Nov 1;131(11):3109S-3120S.

119. Limburg PJ, Stolzenberg-Solomon RZ, Vierkant RA, Roberts K, Sellers TA, Taylor PR, et al. Insulin, Glucose, Insulin Resistance, and Incident Colorectal Cancer in Male Smokers. Clin Gastroenterol Hepatol. 2006 Dec 1;4(12):1514–21.

120. Ma Y, Yang Y, Wang F, Zhang P, Shi C, Zou Y, et al. Obesity and Risk of Colorectal Cancer: A Systematic Review of Prospective Studies. PLOS ONE. 2013 Jan 17;8(1):e53916.

121. Murphy N, Jenab M, Gunter MJ. Adiposity and gastrointestinal cancers: epidemiology, mechanisms and future directions. Nat Rev Gastroenterol Hepatol. 2018 Nov;15(11):659–70.

122. Shanmugalingam T, Bosco C, Ridley AJ, Van Hemelrijck M. Is there a role for IGF-1 in the development of second primary cancers? Cancer Med. 2016;5(11):3353–67.

123. Lu C-C, Chu P-Y, Hsia S-M, Wu C-H, Tung Y-T, Yen G-C. Insulin induction instigates cell proliferation and metastasis in human colorectal cancer cells. Int J Oncol. 2017 Feb;50(2):736–44.

124. Vella V, Milluzzo A, Scalisi NM, Vigneri P, Sciacca L. Insulin Receptor Isoforms in Cancer. Int J Mol Sci. 2018 Nov 16;19(11).

125. Bultman SJ. The microbiome and its potential as a cancer preventive intervention. Semin Oncol. 2016 Feb;43(1):97–106.

126. Schwabe RF, Jobin C. The microbiome and cancer. Nat Rev Cancer. 2013 Nov;13(11):800–12.

127. Gopalakrishnan V, Helmink BA, Spencer CN, Reuben A, Wargo JA. The Influence of the Gut Microbiome on Cancer, Immunity, and Cancer Immunotherapy. Cancer Cell. 2018 Apr 9;33(4):570–80.

128. Fulbright LE, Ellermann M, Arthur JC. The microbiome and the hallmarks of cancer. PLOS Pathog. 2017 Sep 21;13(9):e1006480.

129. Garrett WS. Cancer and the microbiota. Science. 2015 Apr 3;348(6230):80–6.

130. Francescone R, Hou V, Grivennikov SI. Microbiome, inflammation, and cancer. Cancer J Sudbury Mass. 2014 Jun;20(3):181–9.

131. Fouad YA, Aanei C. Revisiting the hallmarks of cancer. Am J Cancer Res. 2017 May 1;7(5):1016–36.

132. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. Cell. 2011 Mar 4;144(5):646–74.

133. Goodwin AC, Destefano Shields CE, Wu S, Huso DL, Wu X, Murray-Stewart TR, et al. Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. Proc Natl Acad Sci U S A. 2011 Sep 13;108(37):15354–9.

134. Nougayrède J-P, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G, et al. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science. 2006 Aug 11;313(5788):848–51.

135. Xavier JB, Young VB, Skufca J, Ginty F, Testerman T, Pearson AT, et al. The Cancer Microbiome: Distinguishing Direct and Indirect Effects Requires a Systemic View. Trends Cancer. 2020 Mar 1;6(3):192–204.

136. Goodman B, Gardner H. The microbiome and cancer. J Pathol. 2018;244(5):667–76.

137. Pitot HC. The molecular biology of carcinogenesis. Cancer. 1993;72(S3):962–70.

138. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. Gastroenterology. 2015 Oct;149(5):1177-1190.e3.

139. GRADY WM, CARETHERS JM. Genomic and Epigenetic Instability in Colorectal Cancer Pathogenesis. Gastroenterology. 2008 Oct;135(4):1079–99.

140. Bakhoum SF, Silkworth WT, Nardi IK, Nicholson JM, Compton DA, Cimini D. The mitotic origin of chromosomal instability. Curr Biol CB. 2014 Feb 17;24(4):R148–9.

141. Leggett B, Whitehall V. Role of the serrated pathway in colorectal cancer pathogenesis. Gastroenterology. 2010 Jun;138(6):2088–100.

142. Mojarad EN, Kuppen PJ, Aghdaei HA, Zali MR. The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterol Hepatol Bed Bench. 2013;6(3):120–8.

143. 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 1;10(6):639–45.

144. Itzkowitz SH, Yio X. Inflammation and cancer IV. Colorectal cancer in inflammatory bowel disease: the role of inflammation. Am J Physiol Gastrointest Liver Physiol. 2004 Jul;287(1):G7-17.

145. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer. 2015 Mar 1;136(5):E359-386.

146. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov;68(6):394–424.

147. Araghi M, Soerjomataram I, Bardot A, Ferlay J, Cabasag CJ, Morrison DS, et al. Changes in colorectal cancer incidence in seven high-income countries: a population-based study. Lancet Gastroenterol Hepatol. 2019;4(7):511–8.

148. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. Gut. 2017;66(4):683–91.

149. Jasperson KW, Tuohy TM, Neklason DW, Burt RW. Hereditary and familial colon cancer. Gastroenterology. 2010 Jun;138(6):2044–58.

150. Kune GA. The Melbourne Colorectal Cancer Study: reflections on a 30-year experience. Med J Aust. 2010 Dec 6;193(11):648–52.

151. Mousavi SM, Fallah M, Sundquist K, Hemminki K. Age- and time-dependent changes in cancer incidence among immigrants to Sweden: colorectal, lung, breast and prostate cancers. Int J Cancer. 2012 Jul 15;131(2):E122-128.

152. Colorectal cancer [Internet]. World Cancer Research Fund. 2018 [cited 2020 Nov 2]. Available from: https://www.wcrf.org/dietandcancer/colorectal-cancer

153. Gunter MJ, Alhomoud S, Arnold M, Brenner H, Burn J, Casey G, et al. Meeting report from the joint IARC–NCI international cancer seminar series: a focus on colorectal cancer. Ann Oncol. 2019 Apr 1;30(4):510–9.

154. Johnson CM, Wei C, Ensor JE, Smolenski DJ, Amos CI, Levin B, et al. Meta-analyses of colorectal cancer risk factors. Cancer Causes Control. 2013 Jun 1;24(6):1207–22.

155. Tabung FK, Brown LS, Fung TT. Dietary Patterns and Colorectal Cancer Risk: A Review of 17 Years of Evidence (2000-2016). Curr Colorectal Cancer Rep. 2017 Dec;13(6):440–54.

156. Ruiz-Casado A, Martín-Ruiz A, Pérez LM, Provencio M, Fiuza-Luces C, Lucia A. Exercise and the Hallmarks of Cancer. Trends Cancer. 2017;3(6):423–41.

157. Lynch BM. Sedentary behavior and cancer: a systematic review of the literature and proposed biological mechanisms. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2010 Nov;19(11):2691–709.

158. Ma P, Yao Y, Sun W, Dai S, Zhou C. Daily sedentary time and its association with risk for colorectal cancer in adults: A dose-response meta-analysis of prospective cohort studies. Medicine (Baltimore). 2017 Jun;96(22):e7049.

159. Tsilidis KK, Kasimis JC, Lopez DS, Ntzani EE, Ioannidis JPA. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. BMJ [Internet]. 2015 Jan 2 [cited 2020 Nov 2];350. Available from: https://www.bmj.com/content/350/bmj.g7607

160. Thrift AP, Gong J, Peters U, Chang-Claude J, Rudolph A, Slattery ML, et al. Mendelian Randomization Study of Body Mass Index and Colorectal Cancer Risk. Cancer Epidemiol Prev Biomark. 2015 Jul 1;24(7):1024–31.

161. Ouchi N, Parker JL, Lugus JJ, Walsh K. Adipokines in inflammation and metabolic disease. Nat Rev Immunol. 2011 Feb;11(2):85–97.

162. Park J, Morley TS, Kim M, Clegg DJ, Scherer PE. Obesity and cancer--mechanisms underlying tumour progression and recurrence. Nat Rev Endocrinol. 2014 Aug;10(8):455–65.

163. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004 Aug;4(8):579–91.

164. Lu Y, Chen J, Zheng J, Hu G, Wang J, Huang C, et al. Mucosal adherent bacterial dysbiosis in patients with colorectal adenomas. Sci Rep [Internet]. 2016 May 19 [cited 2018 Mar 29];6. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4872055/

165. O'Keefe SJD. Diet, microorganisms and their metabolites, and colon cancer. Nat Rev Gastroenterol Hepatol. 2016 Dec;13(12):691–706.

166. Louis P, Hold GL, Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol. 2014 Oct;12(10):661–72.

167. Shang F-M, Liu H-L. Fusobacterium nucleatum and colorectal cancer: A review. World J Gastrointest Oncol. 2018 Mar 15;10(3):71–81.

168. Han YW, Ikegami A, Rajanna C, Kawsar HI, Zhou Y, Li M, et al. Identification and Characterization of a Novel Adhesin Unique to Oral Fusobacteria. J Bacteriol. 2005 Aug 1;187(15):5330–40.

169. Abed J, Emgård JEM, Zamir G, Faroja M, Almogy G, Grenov A, et al. Fap2 Mediates Fusobacterium nucleatum Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. Cell Host Microbe. 2016 Aug 10;20(2):215–25.

170. Gur C, Ibrahim Y, Isaacson B, Yamin R, Abed J, Gamliel M, et al. Binding of the Fap2 protein of Fusobacterium nucleatum to human inhibitory receptor TIGIT protects tumors from immune cell attack. Immunity. 2015 Feb 17;42(2):344–55.

171. Yu T, Guo F, Yu Y, Sun T, Ma D, Han J, et al. Fusobacterium nucleatum Promotes Chemoresistance to Colorectal Cancer by Modulating Autophagy. Cell. 2017 Jul 27;170(3):548-563.e16.

172. Mirzaei R, Mirzaei H, Alikhani MY, Sholeh M, Arabestani MR, Saidijam M, et al. Bacterial biofilm in colorectal cancer: What is the real mechanism of action? Microb Pathog. 2020 May 1;142:104052.

173. Cougnoux A, Delmas J, Gibold L, Faïs T, Romagnoli C, Robin F, et al. Small-molecule inhibitors prevent the genotoxic and protumoural effects induced by colibactin-producing bacteria. Gut. 2016 Feb;65(2):278–85.

174. Lauby-Secretan B, Vilahur N, Bianchini F, Guha N, Straif K. The IARC Perspective on Colorectal Cancer Screening. N Engl J Med. 2018 May 3;378(18):1734–40.

175. Navarro M, Nicolas A, Ferrandez A, Lanas A. Colorectal cancer population screening programs worldwide in 2016: An update. World J Gastroenterol. 2017 May 28;23(20):3632–42.

176. Fecal immunochemical test | Products | EIKEN CHEMICAL CO.,LTD, [Internet]. [cited 2020 Nov 22]. Available from: https://www.eiken.co.jp/en/products/fit/

177. Shaukat A, Mongin SJ, Geisser MS, Lederle FA, Bond JH, Mandel JS, et al. Long-Term Mortality after Screening for Colorectal Cancer. N Engl J Med. 2013 Sep 19;369(12):1106–14.

178. Scholefield JH, Moss SM, Mangham CM, Whynes DK, Hardcastle JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. Gut. 2012 Jul;61(7):1036–40.

179. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med. 2014 Feb 4;160(3):171.

180. Brenner H, Tao S. Superior diagnostic performance of faecal immunochemical tests for haemoglobin in a head-to-head comparison with guaiac based faecal occult blood test among 2235 participants of screening colonoscopy. Eur J Cancer. 2013 Sep 1;49(14):3049–54.

181. Chiu H-M, Chen SL-S, Yen AM-F, Chiu SY-H, Fann JC-Y, Lee Y-C, et al. Effectiveness of fecal immunochemical testing in reducing colorectal cancer mortality from the One Million Taiwanese Screening Program. Cancer. 2015 Sep 15;121(18):3221–9.

182. Giorgi Rossi P, Vicentini M, Sacchettini C, Di Felice E, Caroli S, Ferrari F, et al. Impact of Screening Program on Incidence of Colorectal Cancer: A Cohort Study in Italy. Am J Gastroenterol. 2015 Sep;110(9):1359–66.

183. Ventura L, Mantellini P, Grazzini G, Castiglione G, Buzzoni C, Rubeca T, et al. The impact of immunochemical faecal occult blood testing on colorectal cancer incidence. Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver. 2014 Jan;46(1):82–6.

184. Tinmouth J, Lansdorp-Vogelaar I, Allison JE. Faecal immunochemical tests versus guaiac faecal occult blood tests: what clinicians and colorectal cancer screening programme organisers need to know. Gut. 2015 Aug;64(8):1327–37.

185. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJY, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015 Oct;64(10):1637–49.

186. Taux de participation au programme de dépistage organisé du cancer colorectal 2018-2019 [Internet]. [cited 2020 Nov 22]. Available from: /maladies-et-traumatismes/cancers/cancer-du-colon-rectum/articles/taux-de-participation-au-programme-de-depistage-organise-du-cancer-colorectal-2018-2019

187. Amitay EL, Werner S, Vital M, Pieper DH, Höfler D, Gierse I-J, et al. Fusobacterium and colorectal cancer: causal factor or passenger? Results from a large colorectal cancer screening study. Carcinogenesis. 2017 Aug 1;38(8):781–8.

188. Ottawa Hospital Research Institute [Internet]. Available from: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp

189. Jaacks LM, Vandevijvere S, Pan A, McGowan CJ, Wallace C, Imamura F, et al. The obesity transition: stages of the global epidemic. Lancet Diabetes Endocrinol. 2019 Mar 1;7(3):231–40.

190. Rantakallio P. The longitudinal study of the northern Finland birth cohort of 1966. Paediatr Perinat Epidemiol. 1988 Jan;2(1):59–88.

191. Loftfield E, Karl-Heinz H, J Gregory C, Andriy D, Yunhu W, Dorotha B, et al. Association of Body Mass Index with Fecal Microbial Diversity and Metabolites in the Northern Finland Birth Cohort. [in press]. 2020;

192. Perkiömäki N, Auvinen J, Tulppo MP, Hautala AJ, Perkiömäki J, Karhunen V, et al. Association between Birth Characteristics and Cardiovascular Autonomic Function at Mid-Life. PLOS ONE. 2016 Aug 23;11(8):e0161604.

193. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985 Jul;28(7):412–9.

194. Vogtmann E, Chen J, Amir A, Shi J, Abnet CC, Nelson H, et al. Comparison of Collection Methods for Fecal Samples in Microbiome Studies. Am J Epidemiol. 2017 Jan 15;185(2):115–23.

195. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 2012 Aug;6(8):1621–4.

196. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. Nat Biotechnol. 2017 Nov;35(11):1077–86.

197. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019;37(8):852–7.

198. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 2018 17;6(1):90.

199. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol. 2006 Jul;72(7):5069–72.

200. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013 Apr;30(4):772–80.

201. Price MN, Dehal PS, Arkin AP. FastTree 2--approximately maximum-likelihood trees for large alignments. PloS One. 2010 Mar 10;5(3):e9490.

202. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package [Internet]. 2019 [cited 2020 Jan 31]. Available from: https://CRAN.R-project.org/package=vegan

203. Verdi S, Abbasian G, Bowyer RCE, Lachance G, Yarand D, Christofidou P, et al. TwinsUK: The UK Adult Twin Registry Update. Twin Res Hum Genet Off J Int Soc Twin Stud. 2019 Dec;22(6):523–9.

204. Jamshidi Y, Snieder H, Wang X, Spector TD, Carter ND, O'Dell SD. Common polymorphisms in SOCS3 are not associated with body weight, insulin sensitivity or lipid profile in normal female twins. Diabetologia. 2006 Feb;49(2):306–10.

205. Goodrich JK, Davenport ER, Beaumont M, Jackson MA, Knight R, Ober C, et al. Genetic Determinants of the Gut Microbiome in UK Twins. Cell Host Microbe. 2016 May 11;19(5):731–43.

206. Wells PM, Adebayo AS, Freidin MB, Finckh A, Strowig T, Lesker TR, et al. A polygenic risk score for rheumatoid arthritis sheds light on the Prevotella association. medRxiv. 2019 Dec 14;2019.12.09.19014183.

207. Zhao N, Chen J, Carroll IM, Ringel-Kulka T, Epstein MP, Zhou H, et al. Testing in Microbiome-Profiling Studies with MiRKAT, the Microbiome Regression-Based Kernel Association Test. Am J Hum Genet. 2015 May 7;96(5):797–807.

208. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat. 2005;54(3):507–54.

209. Vrieze A, Van Nood E, Holleman F, Salojärvi J, Kootte RS, Bartelsman JFWM, et al. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology. 2012 Oct;143(4):913-916.e7.

210. Li G, Xie C, Lu S, Nichols RG, Tian Y, Li L, et al. Intermittent Fasting Promotes White Adipose Browning and Decreases Obesity by Shaping the Gut Microbiota. Cell Metab. 2017 Oct 3;26(4):672-685.e4.

211. Wang L, Li P, Tang Z, Yan X, Feng B. Structural modulation of the gut microbiota and the relationship with body weight: compared evaluation of liraglutide and saxagliptin treatment. Sci Rep. 2016 Sep 16;6(1):1–10.

212. Meijnikman AS, Gerdes VE, Nieuwdorp M, Herrema H. Evaluating Causality of Gut Microbiota in Obesity and Diabetes in Humans. Endocr Rev. 2018 01;39(2):133–53.

213. Tremaroli V, Bäckhed F. Functional interactions between the gut microbiota and host metabolism. Nature. 2012 Sep;489(7415):242–9.

214. Saad MJA, Santos A, Prada PO. Linking Gut Microbiota and Inflammation to Obesity and Insulin Resistance. Physiology. 2016 Jun 1;31(4):283–93.

215. Rivière A, Selak M, Lantin D, Leroy F, De Vuyst L. Bifidobacteria and Butyrate-Producing Colon Bacteria: Importance and Strategies for Their Stimulation in the Human Gut. Front Microbiol. 2016;7:979.

216. Tolhurst G, Heffron H, Lam YS, Parker HE, Habib AM, Diakogiannaki E, et al. Short-Chain Fatty Acids Stimulate Glucagon-Like Peptide-1 Secretion via the G-Protein–Coupled Receptor FFAR2. Diabetes. 2012 Feb 1;61(2):364–71.

217. Kimura I, Ozawa K, Inoue D, Imamura T, Kimura K, Maeda T, et al. The gut microbiota suppresses insulin-mediated fat accumulation via the short-chain fatty acid receptor GPR43. Nat Commun. 2013;4:1829.

218. Uemura H, Katsuura-Kamano S, Yamaguchi M, Bahari T, Ishizu M, Fujioka M, et al. Relationships of serum high-sensitivity C-reactive protein and body size with insulin resistance in a Japanese cohort. PLOS ONE. 2017 Jun 2;12(6):e0178672.

219. Verdam FJ, Fuentes S, Jonge C de, Zoetendal EG, Erbil R, Greve JW, et al. Human intestinal microbiota composition is associated with local and systemic inflammation in obesity. Obesity. 2013;21(12):E607–15.

220. Leiva-Gea I, Sánchez-Alcoholado L, Martín-Tejedor B, Castellano-Castillo D, Moreno-Indias I, Urda-Cardona A, et al. Gut Microbiota Differs in Composition and Functionality Between Children With Type 1 Diabetes and MODY2 and Healthy Control Subjects: A Case-Control Study. Diabetes Care. 2018;41(11):2385–95.

221. Lambeth SM, Carson T, Lowe J, Ramaraj T, Leff JW, Luo L, et al. Composition, Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. J Diabetes Obes. 2015 Dec 26;2(3):1–7.

222. Saravia G, Civeira F, Hurtado-Roca Y, Andres E, Leon M, Pocovi M, et al. Glycated Hemoglobin, Fasting Insulin and the Metabolic Syndrome in Males. Cross-Sectional Analyses of the Aragon Workers' Health Study Baseline. PLOS ONE. 2015 Aug 4;10(8):e0132244.

223. Schott EM, Farnsworth CW, Grier A, Lillis JA, Soniwala S, Dadourian GH, et al. Targeting the gut microbiome to treat the osteoarthritis of obesity. JCI Insight [Internet]. 2018 Apr 24 [cited 2020 Feb 11];3(8). Available from: https://insight.jci.org/articles/view/95997

224. Naderpoor N, Mousa A, Gomez-Arango LF, Barrett HL, Dekker Nitert M, de Courten B. Faecal Microbiota Are Related to Insulin Sensitivity and Secretion in Overweight or Obese Adults. J Clin Med [Internet]. 2019 Apr 4 [cited 2020 Feb 11];8(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6518043/

225. Serena C, Ceperuelo-Mallafré V, Keiran N, Queipo-Ortuño MI, Bernal R, Gomez-Huelgas R, et al. Elevated circulating levels of succinate in human obesity are linked to specific gut microbiota. ISME J. 2018 Jul;12(7):1642–57.

226. He Y, Wu W, Zheng H-M, Li P, McDonald D, Sheng H-F, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. Nat Med. 2018 Oct 1;24(10):1532–5.

227. Ley RE. Obesity and the human microbiome. Curr Opin Gastroenterol. 2010 Jan;26(1):5–11.

228. Vogtmann E, Goedert JJ. Epidemiologic studies of the human microbiome and cancer. Br J Cancer. 2016 Feb 2;114(3):237–42.

229. Huybrechts I, Zouiouich S, Loobuyck A, Vandenbulcke Z, Vogtmann E, Pisanu S, et al. The Human Microbiome in Relation to Cancer Risk: A Systematic Review of Epidemiologic Studies. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2020 Jul 29;

230. Sinha R, Chen J, Amir A, Vogtmann E, Shi J, Inman KS, et al. Collecting Fecal Samples for Microbiome Analyses in Epidemiology Studies. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2016 Feb;25(2):407–16.

231. Sinha R, Abnet CC, White O, Knight R, Huttenhower C. The microbiome quality control project: baseline study design and future directions. Genome Biol [Internet]. 2015

[cited 2018 Sep 28];16. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4674991/

232. Vogtmann E, Chen J, Kibriya MG, Chen Y, Islam T, Eunes M, et al. Comparison of Fecal Collection Methods for Microbiota Studies in Bangladesh. Appl Environ Microbiol. 2017 15;83(10).

233. Wang Z, Zolnik CP, Qiu Y, Usyk M, Wang T, Strickler HD, et al. Comparison of Fecal Collection Methods for Microbiome and Metabolomics Studies. Front Cell Infect Microbiol. 2018;8:301.

234. Byrd DA, Chen J, Vogtmann E, Hullings A, Song SJ, Amir A, et al. Reproducibility, stability, and accuracy of microbial profiles by fecal sample collection method in three distinct populations. PLOS ONE. 2019 Nov 18;14(11):e0224757.

235. Song SJ, Amir A, Metcalf JL, Amato KR, Xu ZZ, Humphrey G, et al. Preservation Methods Differ in Fecal Microbiome Stability, Affecting Suitability for Field Studies. mSystems. 2016 Jun;1(3).

236. Dominianni C, Wu J, Hayes RB, Ahn J. Comparison of methods for fecal microbiome biospecimen collection. BMC Microbiol. 2014 Apr 23;14:103.

237. Rounge TB, Meisal R, Nordby JI, Ambur OH, de Lange T, Hoff G. Evaluating gut microbiota profiles from archived fecal samples. BMC Gastroenterol. 2018 Nov 8;18(1):171.

238. Nechvatal JM, Ram JL, Basson MD, Namprachan P, Niec SR, Badsha KZ, et al. Fecal collection, ambient preservation, and DNA extraction for PCR amplification of bacterial and human markers from human feces. J Microbiol Methods. 2008 Feb;72(2):124–32.

239. Taylor M, Wood HM, Halloran SP, Quirke P. Examining the potential use and long-term stability of guaiac faecal occult blood test cards for microbial DNA 16S rRNA sequencing. J Clin Pathol. 2017 Jul;70(7):600–6.

240. Tap J, Cools-Portier S, Pavan S, Druesne A, Öhman L, Törnblom H, et al. Effects of the long-term storage of human fecal microbiota samples collected in RNAlater. Sci Rep. 2019 24;9(1):601.

241. Hale VL, Tan CL, Knight R, Amato KR. Effect of preservation method on spider monkey (Ateles geoffroyi) fecal microbiota over 8weeks. J Microbiol Methods. 2015 Jun 1;113:16–26.

242. Gudra D, Shoaie S, Fridmanis D, Klovins J, Wefer H, Silamikelis I, et al. A widely used sampling device in colorectal cancer screening programmes allows for large-scale microbiome studies. Gut. 2019 Sep 1;68(9):1723–5.

243. Baxter NT, Koumpouras CC, Rogers MAM, Ruffin MT, Schloss PD. DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. Microbiome. 2016 Nov 14;4(1):59.

244. Byrd DA, Sinha R, Hoffman KL, Chen J, Hua X, Shi J, et al. Comparison of Methods To Collect Fecal Samples for Microbiome Studies Using Whole-Genome Shotgun Metagenomic Sequencing. mSphere. 2020 26;5(1).

245. Löwenmark T, Löfgren-Burström A, Zingmark C, Eklöf V, Dahlberg M, Wai SN, et al. Parvimonas micra as a putative non-invasive faecal biomarker for colorectal cancer. Sci Rep. 2020 Sep 17;10(1):15250.

246. Bahl MI, Bergström A, Licht TR. Freezing fecal samples prior to DNA extraction affects the Firmicutes to Bacteroidetes ratio determined by downstream quantitative PCR analysis. FEMS Microbiol Lett. 2012 Apr;329(2):193–7.

247. Fouhy F, Deane J, Rea MC, O'Sullivan Ó, Ross RP, O'Callaghan G, et al. The effects of freezing on faecal microbiota as determined using MiSeq sequencing and culture-based investigations. PloS One. 2015;10(3):e0119355.

248. Segnan N, Senore C, Andreoni B, Azzoni A, Bisanti L, Cardelli A, et al. Comparing Attendance and Detection Rate of Colonoscopy With Sigmoidoscopy and FIT for Colorectal Cancer Screening. Gastroenterology. 2007 Jun 1;132(7):2304–12.

249. Vidal-Lletjós S, Andriamihaja M, Blais A, Grauso M, Lepage P, Davila A-M, et al. Mucosal healing progression after acute colitis in mice. World J Gastroenterol. 2019 Jul 21;25(27):3572–89.

250. Escudié F, Auer L, Bernard M, Mariadassou M, Cauquil L, Vidal K, et al. FROGS: Find, Rapidly, OTUs with Galaxy Solution. Bioinformatics. 2018 Apr 15;34(8):1287–94.

251. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014 Mar 1;30(5):614–20.

252. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011 May 2;17(1):10–2.

253. Mahé F, Rognes T, Quince C, Vargas C de, Dunthorn M. Swarm v2: highly-scalable and high-resolution amplicon clustering. PeerJ. 2015 Dec 10;3:e1420.

254. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016 Oct 18;4:e2584.

255. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013 Jan 1;41(D1):D590–6.

256. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.

257. McMurdie PJ, Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS One. 2013;8(4):e61217.

258. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.

259. Choo JM, Leong LE, Rogers GB. Sample storage conditions significantly influence faecal microbiome profiles. Sci Rep. 2015 Nov 17;5(1):16350.

260. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, et al. Towards standards for human fecal sample processing in metagenomic studies. Nat Biotechnol. 2017 Nov;35(11):1069–76.

261. Overcoming hurdles in sharing microbiome data. Nat Microbiol. 2017 Dec;2(12):1573–1573.

262. Sinha R, Ahsan H, Blaser M, Caporaso JG, Carmical JR, Chan AT, et al. Next steps in studying the human microbiome and health in prospective studies, Bethesda, MD, May 16-17, 2017. Microbiome. 2018 26;6(1):210.

263. Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and Qualitative β Diversity Measures Lead to Different Insights into Factors That Structure Microbial Communities. Appl Environ Microbiol. 2007 Mar 1;73(5):1576–85.

264. Ho NT, Li F, Wang S, Kuhn L. metamicrobiomeR: an R package for analysis of microbiome relative abundance data using zero-inflated beta GAMLSS and meta-analysis across studies using random effects models. BMC Bioinformatics. 2019 Apr 16;20(1):188.

265. Walters WA, Xu Z, Knight R. Meta-analyses of human gut microbes associated with obesity and IBD. FEBS Lett. 2014 Nov 17;588(22):4223–33.

266. Finucane MM, Sharpton TJ, Laurent TJ, Pollard KS. A Taxonomic Signature of Obesity in the Microbiome? Getting to the Guts of the Matter. PLOS ONE. 2014 Jan 8;9(1):e84689.

267. Sze MA, Schloss PD. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. mBio. 2016 Sep 7;7(4).

268. Shah MS, DeSantis TZ, Weinmaier T, McMurdie PJ, Cope JL, Altrichter A, et al. Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer. Gut. 2018 May;67(5):882–91.

269. Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics. 2012 Aug 15;28(16):2106–13.

270. Gloor GB, Wu JR, Pawlowsky-Glahn V, Egozcue JJ. It's all relative: analyzing microbiome data as compositions. Ann Epidemiol. 2016 May 1;26(5):322–9.

271. Calle ML. Statistical Analysis of Metagenomics Data. Genomics Inform [Internet]. 2019 Mar 31 [cited 2020 Nov 24];17(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6459172/

272. Kim M-H, Yun KE, Kim J, Park E, Chang Y, Ryu S, et al. Gut microbiota and metabolic health among overweight and obese individuals. Sci Rep. 2020 Nov 10;10(1):19417.

273. Caricilli AM, Saad MJA. The Role of Gut Microbiota on Insulin Resistance. Nutrients. 2013 Mar;5(3):829–51.

274. Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. Nat Genet. 2018 Jun;50(6):790–5.

275. Chen MX, Wang S-Y, Kuo C-H, Tsai I-L. Metabolome analysis for investigating host-gut microbiota interactions. J Formos Med Assoc Taiwan Yi Zhi. 2019 Mar;118 Suppl 1:S10–22.

276. Kieffer DA, Piccolo BD, Vaziri ND, Liu S, Lau WL, Khazaeli M, et al. Resistant starch alters gut microbiome and metabolomic profiles concurrent with amelioration of chronic kidney disease in rats. Am J Physiol-Ren Physiol. 2016 Feb 3;310(9):F857–71.

277. Lin H, An Y, Hao F, Wang Y, Tang H. Correlations of Fecal Metabonomic and Microbiomic Changes Induced by High-fat Diet in the Pre-Obesity State. Sci Rep. 2016 Feb 26;6(1):21618.

278. Westerterp KR, Goris AHC. Validity of the assessment of dietary intake: problems of misreporting. Curr Opin Clin Nutr Metab Care. 2002 Sep;5(5):489–93.

279. Freedman LS, Schatzkin A, Midthune D, Kipnis V. Dealing with dietary measurement error in nutritional cohort studies. J Natl Cancer Inst. 2011 Jul 20;103(14):1086–92.

280. Playdon MC, Ziegler RG, Sampson JN, Stolzenberg-Solomon R, Thompson HJ, Irwin ML, et al. Nutritional metabolomics and breast cancer risk in a prospective study. Am J Clin Nutr. 2017 Aug;106(2):637–49.

281. Guertin KA, Loftfield E, Boca SM, Sampson JN, Moore SC, Xiao Q, et al. Serum biomarkers of habitual coffee consumption may provide insight into the mechanism underlying the association between coffee consumption and colorectal cancer. Am J Clin Nutr. 2015 May;101(5):1000–11.

282. Shaw AG, Sim K, Powell E, Cornwell E, Cramer T, McClure ZE, et al. Latitude in sample handling and storage for infant faecal microbiota studies: the elephant in the room? Microbiome. 2016 30;4(1):40.

283. Stearns JC, Lynch MDJ, Senadheera DB, Tenenbaum HC, Goldberg MB, Cvitkovitch DG, et al. Bacterial biogeography of the human digestive tract. Sci Rep. 2011;1:170.

284. Ott SJ, Musfeldt M, Timmis KN, Hampe J, Wenderoth DF, Schreiber S. In vitro alterations of intestinal bacterial microbiota in fecal samples during storage. Diagn Microbiol Infect Dis. 2004 Dec;50(4):237–45.

285. Roesch LFW, Casella G, Simell O, Krischer J, Wasserfall CH, Schatz D, et al. Influence of fecal sample storage on bacterial community diversity. Open Microbiol J. 2009;3:40–6.

286. Amir A, McDonald D, Navas-Molina JA, Debelius J, Morton JT, Hyde E, et al. Correcting for Microbial Blooms in Fecal Samples during Room-Temperature Shipping. mSystems [Internet]. 2017 Apr 21 [cited 2020 Nov 25];2(2). Available from: https://msystems.asm.org/content/2/2/e00199-16

287. Aagaard K, Petrosino J, Keitel W, Watson M, Katancik J, Garcia N, et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. FASEB J Off Publ Fed Am Soc Exp Biol. 2013 Mar;27(3):1012–22.

288. Loftfield E, Vogtmann E, Sampson JN, Moore SC, Nelson H, Knight R, et al. Comparison of Collection Methods for Fecal Samples for Discovery Metabolomics in Epidemiologic Studies. Cancer Epidemiol Prev Biomark. 2016 Nov 1;25(11):1483–90.

289. Kia E, Wagner Mackenzie B, Middleton D, Lau A, Waite DW, Lewis G, et al. Integrity of the Human Faecal Microbiota following Long-Term Sample Storage. PloS One. 2016;11(10):e0163666.

290. Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Zech Xu Z, et al. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. mSystems. 2017 Mar 7;2(2).

Appendices

Supplementary material: The human microbiome in relation to cancer risk: a systematic review of epidemiological studies

Supplementary Table S1: Search Strings used in literature searches*

Pubmed:	
#1	Neoplasms[MeSH Terms]
#2	cancer
#3	microbiome[MeSH Terms]
#4	microflora
#5	review[Publication Type]
#6	infection[Title/Abstract]
#7	infections[Title/Abstract]
#8	(#1 or #2) and (#3 or #4) not #5 not #6 not #7
Embase:	
#1	Neoplasm
#2	Cancer
#3	Microbiome
#4	#1 or #2 and #3
The Cochrane Library search strategy:	
#1	Intestinal bacteria
#2	cancer
#3	intestinal microbiome
#4	#2 and #3
#5	neoplasm
#6	neoplasia
#7	microbiome
#8	microflora
#9	microbiota
#10	(#2 or #5 or #6) and (#7 or #8 or#9)

*Search strings containing keywords and database-specific terms (MeSH, Emtree terms and exploded terms) were decided after testing different searches using neoplasms, cancer, microbiome, microflora and several synonyms as main concepts. The Cochrane library was consulted to assure completeness of the literature search.

Supplementary Table S2: Criteria used for the quality assessment - Newcastle Ottawa Scale

Selection		**Proper selection of study population**
	1	Case definition adequate
	2	Representativeness of the cases
	3	Selection of controls
	4	Definition of controls
Comparability		**Comparability of the study groups**
	5	Studies that controlled for age, sex and traditional confounders typical for the cancer site under study (e.g. smoking for most cancers)
	6	Studies that controlled for other relevant/important confounders (considering the cancer site; e.g. breast feeding for breast cancer)
Exposure		**Ascertainment of the exposure or outcome of interest**
	7	Ascertainment of exposure
	8	Same method of ascertainment for cases and controls
	9	Non-response rate

Supplementary Table S3: The design of the studies used in the articles, presenting sample size, age and origin of participants and adjustments available for analyses. The term "case control samples" is used for studies using adjacent normal tissue from the same subject as controls.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Liu, 2018 Case control	Airway microbiome	Lung cancer	24 lung cancer patients with unilateral lobar masses (paired samples from cancerous site and the contralateral noncancerous site) and 18 healthy controls undergoing bronchoscopies	Lung cancer: 60.58 Controls: 56.61	China	chemotherapy, radiation therapy, undergoing surgery or other treatments for lung cancer before bronchoscopy, antibiotics in the prior 3 months	Gender, Age, BMI, Smoking, Tumor characteristics
Tsay, 2018 Case control	Airway microbiome	Lung cancer	39 subjects with final lung cancer diagnoses and 36 subjects with non-cancer diagnoses, 10 healthy control subjects	61.2 years	USA	prior history of cancer or recent (1 mo) antibiotic use	age, gender, race, smoking status, pack-years, diagnosis
Lee, 2016 Cohort	Lung microbiome	Lung cancer	28 patients were included in this study; 20 patients were diagnosed with lung cancer, and 8 were diagnosed with a benign mass-like lesion	64 ± 11 years	South Korea	less than 20 years of age, pregnant, or had undergone any procedure other than bronchoscopy to evaluate the lung mass.	Demographic and clinical data, including age, gender, body mass index (BMI), pulmonary function, smoking status, smoking amount, comorbidities, and final diagnosis
Carpagnano, 2014 Case control	Lung microbiome	Lung cancer cases	43 pat. with non-small cell lung cancer (NSCLC) and 21 healthy controls (H).	(NSCLC) 68.4 years [±SD 9.2], (H) 64.1 years [±SD 13.1].	Italy		Adjustments for sex, age, histotype, stage, smoking habit, pack years, time since quitting smoking in subjects with fungal colonization.
Hosgood, 2014 Case control	Lung microbiome	Lung cancer cases	8 never smoking female lung cancer pat. (LC) and 8 never	The age range of (LC) and (H) was 45-72 years.	China	°controls previously diagnosed with lung cancer °smokers	Case-control matching by age and hospital

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			smoking female healthy controls (H).				
Cameron, 2017 Case control	Sputum microbiome	Lung cancer	10 patients referred with possible LC, of which four were eventually diagnosed with LC), and six had no LC after one year.	Cases: 58.8 (14.8) Controls: 73.3 (7.2)	UK	Not available	Smoking status, Antibiotic use, CO level (ppm), FEV1% of Predicted
Aviles-Jimenez, 2016 Case control	Bile duct microbiome	Cholangiocarcinoma cases	200 subjects: 100 pat. with extrahepatic cholangiocarcinoma (ECCA) and 100 pat. with benign pathology of the common bile duct (BBP).	(ECCA) 66.2 years [range 50-82 years], (BBP) 53.1 years [range 23-83].	Mexico		Case-control matching for sex, age (± 5 years) and place of residence + clinical variables incl. the time of evolution of lithiasis, H. bilis/ H. hepaticus infection, sex, age or body mass index.
Chng, 2016 Case control	Bile duct microbiome	Cholangiocarcinoma cases	60 pat. with cholangiocarcinoma (CCA) and adjacent matched normal samples, 5 non-cancer hepatic samples, 2 bile fluid samples and 4 non-cancer gastric mucosa samples (non-CCA). Of the CCA, 28 pat. are Opisthorchis viverriniviverrini associated (OVa), and 32 pat. are not (non-OVa).	(CCA OVa) 57.9 years [range 38-71], (CCA non-OVa) 56.6 years [range 33-69]. For the other samples, mean age is 59.6 years [range 46-82].	Thailand, Singapore or Romania		Case-control matching for age, gender and anatomical subtype.
Hieken 2016 Case control	Breast tissue microbiome	Breast cancer cases	28 pat. of whom 13 benign non-atypia breast disease and	Benign patients median [range] age: 49 years	USA		Adjustments for contamination from the skin, batch effects

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			15 invasive breast cancer	[33–70] Breast cancer pat.: 75 years [44–84]			
Urbaniak 2016 Case control	Breast tissue microbiome	Breast cancer cases	81 pat. of whom 13 benign breast tumors, 45 invasive breast cancer and 23 disease free controls	Age range [19–90 years] Benign patients median age: 36 years IBC patients median age: 62 years Healthy controls median age: 53 years	Canada		Adjustments for contamination from the skin, batch effects
Audirac-Chalifour, 2016 Case control	Cervical microbiome	Cervical cancer cases	32 subjects: 8 with cervical cancer and HPV+ (CC), 4 with squamous intraepithelial lesions and HPV+ (SIL) and 20 healthy controls with no cervical lesions (H): 10 HPV- and 10 HPV+.	(CC) 43 years [±SD 11],(SIL) 40 years [±SD 14], (H) 34 years [±SD 8].	Mexico	°insufficient reads °patient´s recruitment on the same day of menstrual period ° the non-use of douches and no sexual activity in previous days of the sampling °records of used medication in the last 30 days previous to sampling °molecular HPV+ diagnosis.	Adjustments for age, parity, contraceptive method and HPV-genotype.
Oh, 2015 Case control	Cervical microbiome	Cervical cancer cases (CIN)	120 women: 70 with cervical intraepithelial neoplasia (CIN) 55 CIN1, 15 CIN2 or CIN3, and 50 healthy controls (H).	42 subjects younger than 39 years, 37 between 40 and 49 years, 30 subjects between 50 and 59 years and 12 subjects 60 years or older.	Korea	°history of gynecological cancers, °insufficient data on the questionnaire °inadequate blood for evaluation, chronic disease, drug dependency, °psychological problems.	Adjustments for age, marital status, menopause, smoking status, oral contraceptive use and histological grade but not significant.
Seo, 2016 Case control	Cervical microbiome	Cervical cancer cases (CIN)	137 subjects: 65 pat. with cervical intraepithelial	43.6 years [±SD 11.2].	South Korea	°histories of gynecologic cancers, chronic diseases, drug	Adjustments for daily intake of nutrients, age, BMI, marital status, menopausal status,

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			neoplasia (CIN), from who 50 pat. with CIN1 and 15 with CIN2 or CIN3, and 72 healthy controls (H).			dependencies, or psychological problems °insufficient data on the questionnaire or in specimens.	smoking status, alcohol drinking status, oncogenic HPV infection, monthly family income, parity and oral contraceptive use.
Mannell, 1983 Case control	Esophageal microbiome	Esophageal cancer cases	101 subjects: 50 pat. with esophagus carcinoma (EC) and 51 healthy controls (H).	Not mentioned.	South Africa	/	
Wang, 2016 Case control	Gastric microbiome	Gastric cancer	315 patients, including 212 patients with chronic gastritis and 103 patients with gastric cancer	55.8±13.5 years	China	endoscopic findings of peptic ulcer, polyps, or any other local lesions ; Patients who showed histological evidence of atrophy or intestinal metaplasia	Not available
Castaño-Rodríguez, 2017 Case control	Gastric microbiome	gastric cancer	gastric cancer (GC, n = 12) and controls (functional dyspepsia (FD), n = 20)	Not available	Singapore/Malaysia	Subjects who had been prescribed antibiotics, non-steroidal anti-inflammatory drugs (NSAIDs) or acid suppressants in the two-month period prior to recruitment as well as subjects known to be infected with the Human Immunodeficiency Virus (HIV	age, gender or country of origin
Li, 2017 Case control	Gastric microbiome	Gastric cancer	33 individuals including 9 subjects with HP-associated chronic gastritis, 9 gastric intestinal metaplasia, 7 gastric adenocarcinoma	Not available	China	active gastroduodenal ulcers on endoscopy, acid suppressive therapy including PPIs and histamine (H2)-receptor blocker, NSAIDs, anti-	age, gender, HP state, sampling location, and HP-eradication

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			and 8 HP-negative controls			coagulants, anti-platelet medications, antibiotics within recent four weeks, upper gastrointestinal endoscopy, prior gastric surgery or HP eradication treatment	
Yu, 2017 Case control	Gastric microbiome	Gastric cancer	80 paired samples (non-malignant and matched tumor tissues) collected from GC patients	median 62 (58–67)	China	Not available	Alcohol and pickled vegetable consumption, history of cancer, tumor grade and stage, metastasis, survival (days)
Yu, 2017 Case control	Gastric microbiome	Gastric cancer	160 gastric cancer patients with 80 from China and 80 from Mexico	Chinese cases:60.8 years old Mexican cases: 64.5 years old	China/Mexico	Chinese cases: <18 years old, with cancer other than GC or with previous treatment for GC Mexican cases: <18 years old, with any autoimmune disease, diabetes, or cancer other than GC, and with a previous treatment for GC	Age, gender, tumoral regions of the stomach, country
Coker, 2018 Case control	Gastric microbiome	Gastric cancer	81 cases; validated in 126 additional cases	Not available	China	Not available	age, gender, H. pylori status and tissue positions
Hu, 2018 Case control	Gastric microbiome	gastric cancer	6 patients with GC and 5 patients with SG	superficial gastritis: 60.5 ± 6.5 years gastric adenocarcinoma: 55.2 ± 5.6 years	China	use ofantibiotics within 6months, receiving chemotherapy or radiotherapy prior to the collection of specimens, the use of proton pump inhibitors or other digestive system drugs within 4 weeks.	Gender, Age, Diabetes, BMI

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Liu, 2019 Cohort	Gastric microbiome	Gastric cancer	276 GC patients without preoperative chemotherapy, 230 normal, 247 peritumoral and 229 tumoral tissues	GC: 61.11 ± 11.82	China	body mass index (BMI =weight in kilograms divided by the height in meters squared) > 30; use of antibiotics, probiotics, prebiotics, or synbiotics in the previous month; preoperative chemotherapy, radiotherapy, or other biological treatment before gastrectomy	age, gender, weight, height, BMI, complications, tumor localization, tumor differentiation, lauren typing, tumor stage, HP infection, antibiotic use, PPI use, Pre-operative chemotherapy, sample collection
Aviles-Jimenez, 2014 Case control	Gastric microbiome	Gastric cancer cases	15 pat.: 5 pat. with non-atrophic gastritis (NAG), 5 pat. with intestinal metaplasia (IM) and 5 pat. with an intestinal-type of gastric cancer (GC).	(NAG) 44 years [range 32-76], (IM) 67 years [range 60-71], (GC) 70.6 years [range 52-81].	Mexico	°pat. with immunodeficiency °diabetes or other chronic diseases °pat. who received certain medication during the last three months °pat. who previously received therapy for H. pylori eradication.	Adjustments for sex and age.
Seo, 2014 Case control	Gastric microbiome	Gastric cancer cases	16 pat. with gastric cancer from who two samples each were taken: one sample of the gastric cancer (GC), one of the adjacent normal gastric mucosa (H).	67.18 years [range 37-75] for the 11 pat whose samples were used for statistical analyses.	South Korea	/	
Nasrollahzadeh, 2015 Case control	Gastric microbiome	Esophageal cancer cases	91 subjects: 19 pat. with esophageal squamous cell carcinoma (ESCC), 18 pat with esophageal squamous dysplasia	(ESCC) and (ESD) 64.5 years [±SD 11.8], (DC) 63.6 years [±SD 14.0] and (HC) 62.1 year [±SD 16.3].	Iran	°samples < 1000 reads.	Case-control matching for age and sex.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			(ED), 17 pat with mid-esophageal esophagitis as the diseased controls (DC) and 37 healthy controls (HC).				
Baxter, 2016 Case control	Gut microbiome	colorectal cancer	490 patients: 120 CRC, 198 adenomas, and 172 without colonic lesions	29–89 years with a median of 60 years	Canada and USA	Excluded if they had undergone surgery, radiation, or chemotherapy for current CRC prior to baseline samples or had inflammatory bowel disease, known hereditary non-polyposis CRC, or familial adenomatous polyposis	No adjustment
Sinha, 2016 Case control	Gut microbiome	Colorectal cancer	42 CRC cases and 89 matched controls	Cases: 63.4 (13.1) Controls: 58.4 (13.0)	USA	Not available	Controls were frequency matched to cases by gender and body mass index (BMI) Adjustment on age, sex, race, hospital and BMI
Thomas, 2016 Case control	Gut microbiome	Colorectal cancer	18 rectal-cancer subjects, 18 non-cancer controls	Cases: 59.3 ± 8.8 Controls: 55.2 ± 15.7	Brazil	No subjects had personal or familial history of colorectal cancer or colitis (either ulcerative, Crohn's, radiation or infectious colitis, chronic inflammatory illnesses), previous colonic or small bowel resection, nor previous colon adenomas or familial polyposis syndrome. Only individuals with	Chi-Square tests were performed on subject's categorical data such as gender, alcohol and tobacco use and vital status. Student t-tests were performed to compare differences in the means between both groups for age, height, weight, BMI.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
						complete colonoscopies that allowed the full visualization of the entire colon and showed no significant clinical alterations were included.	
Amitay, 2017 Case control	Gut microbiome	Colorectal cancer	500 participants, including 46, 113, 110 and 231 individuals with colorectal cancer, advanced adenomas, non-advanced adenomas and without any neoplasms	63.2 years	Germany	inflammatory bowel disease or had a history of CRC	sociodemographic, nutritional and lifestyle factors
Flemer, 2017 Case control	Gut microbiome	Colorectal cancer	59 patients undergoing surgery for CRC, 21 individuals with polyps and 56 healthy	Not available	Ireland	personal history of CRC, IBD or IBS	age-matched controls
Gao, 2017 Case control	Gut microbiome	Colorectal cancer	65 patients with CRC	63.89 years for males and 63.08 years for females	China	antibiotics and probiotics within 4 weeks, acute diarrhea, confirmed simple adenoma or polys, inflammatory bowel disease (IBD), or irritable bowel syndrome	body mass index (BMI), age, gender, and disease history
Xu, 2017 Case control	Gut microbiome	Colorectal cancer	61 healthy controls, 47 adenoma patients, and 52 carcinoma patients	Not available	China	Not available	Not available

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Yu, 2017 Case control	Gut microbiome	Colorectal cancer	74 patients with CRC and 54 controls from China, and validation in 16 patients and 24 controls from Denmark.	Chinese cases: median age 67 years Mexican cases: median age 62 years	China/Denmark	all patients with adenoma	validation of the results with data from Denmark
Allali, 2018 Case control	Gut microbiome	Colorectal cancer	11 CRC patients and 12 healthy individuals	Cases: 52.8 ± 54 Controls: 49.3 ± 46	Morocco	no gastrointestinal disorders, no antibiotic use during the last 3 months and those who had been recently diagnosed and had not yet started treatment	age, gender, tumor location, BMI, dietary data, family history of cancer
Burns, 2018 Case control	Gut microbiome	Colorectal cancer	88 tissue samples from 44 individuals	Not available	USA	Not available	one tumor and one normal sample from each individual
Dai, 2018 Case control	Gut microbiome	Colorectal cancer	271 controls and 255 CRC cases	Cohort C1 (American): 60.96(13.56) Cohort C2 (Austrian): 67.1(10.91) Cohort C3 (Chinese): 65.90(10.61) Cohort C4 (German and French): 68.44(12.22)	China, Austria, USA, Germany, France	Cohort C1: Prior to surgery and treatment Cohort C2: Not available Cohort C3: No antibiotics and no invasive medical intervention for 3 months; no vegetarian diet; no history of cancer or inflammatory disease of intestine Cohort C4: No previous colon or rectal surgery, colorectal cancer, inflammatory, or infectious injuries of the intestine; no need for need for emergency colonoscopy	Sample size, Age, Gender, BMI
Hale, 2018 Case control	Gut microbiome	Colorectal cancer	83 individuals	(dMMR): 74 (18) (pMMR): 63 (13)	USA	chemotherapy or radiation in the 2	sex, age, BMI, smoking status, tumor location, stage

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
						weeks leading up to enrollment	
Loke, 2018 Case control	Gut microbiome	Colorectal cancer	17 Malaysian patients	left-sided CRC: 60.9 (95% CI: 54.2–67.6) right-sided CRC: 62.8 (95% CI: 47.6–78.1)	Malaysia	pre-operative radiation, chemotherapy treatment or had a past history ofCRC or inflammatory bowel disease	gender, race, age, type of tumor, location, stage, biofilm
Mori, 2018 Case control	Gut microbiome	Colorectal cancer	Normal biopsies: 18 subjects 21 subjects were diagnosed with ADK, 21 with HRA, 18 with LRA, 14 with HP.	Not available	Italy	Subjects with family history of cancer	Demographic information, clinic data, comobordities data
Richard, 2018 Case control	Gut microbiome	Colorectal cancer	7 CAC patients, 10 SC patients and 10 HS	colitis associated cancer (CAC): 50.7 (10) years sporadic cancer (SC): 68.8 (12.1) years healthy subjects (HS): 48.3 (13.4) years	France	infectious colitis, coagulation disorders or concomitant treatment with anticoagulant therapy, and treatment with antibiotics or antifungal treatments in the two months before inclusion	gender, age, Montreal classification, Type of cancer, Previous abdominal surgery, Current treatment, previous treatment
Shah, 2018 Case control	Gut microbiome	Colorectal cancer	tumor-adjacent biopsies (n = 294 pairs, n = 588 biospecimens) and matched tumor biopsy:fecal pairs (n = 42 pairs, n = 84 biospecimens)	Not available	USA	Not available	clinical and demographic factors of the participants, location of the tumor in the colon, and stage and grade of tumor were not available for all of the participants
Zhang, 2018 Case control	Gut microbiome	Colorectal cancer	30 initially diagnosed CRC patients, 88 advanced colorectal adenoma patients,	CRC: 60.5 (9.8) A-CRA: 59.6 (10.3) Polyps: 56.5 (8.9)	China	(a) patients younger than 40 years of age, (b) persons not Han people, (c) patients with prior diagnoses of	gender, age, BMI, alcohol, smoking, lesion location, hypertension, heart disease, diabetes

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			62 patients with benign intestinal polyps and 130 controls	Controls: 58.6 (8.9)		colorectal cancer, colorectal adenoma, inflammatory bowel disease (IBD) or other cancers, (d) patients had a family history of colorectal cancer in first- and second-degree relatives and no family history of neoplastic polyps or hereditary syndromes in first degree relatives under 60 years of age, and (e) patients had used antibiotics in last 6 months before colonoscopy.	
Saito, 2019 Case control	Gut microbiome	Colorectal cancer	81 Japanese patients, including 47 CRA and 24 intramucosal CRC patients, and 10 healthy subjects	CRC: 67±9 CRA: 66±8 Controls: 58±15	Japan	Not available	Gender, age, BMI, diabetes, hypercholesteronlemia, hypertension, antibiotic treatment, reason for colonoscopy
Yachida, 2019 Cohort	Gut microbiome	Colorectal cancer	616 participants who underwent colonoscopy	Not available	Japan	hereditary or suspected hereditary disease (for example, familial adenomatous polyposis, hereditary non-polyposis colorectal cancer, microsatellite instability-high), an inflammatory bowel disease, an abdominal surgical history or for which stool samples	body-mass index (BMI), alcohol consumption (grams per day) and smoking habits (Brinkman index), dietary habits

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
						were insufficient for data collection	
Ahn, 2013 Case control	Gut microbiome	Colorectal cancer cases	141: 47colorecal cancer (CRC), 94 healthy controls (HC)	CRC: 62.9 years [±SD: 12.9] HC: 58.5 years [±SD: 12.9]	CRC: 72.3%white, 25.5 % black, 2.1% others ; HC: 86.2% white, 11.7% black, 2.1% others (not significant different between CRC and HC); Washington DC	°less than 100mg of lyophilized feces available	CRC and HC were matched for age, sex, BMI and race before the test.
Amiot, 2015 Case control	Gut microbiome	Colorectal cancer cases	55: 33 advanced colorectal neoplasia (20CRC and 13 with advanced adenoma) and 22 HCs 13with normal colonoscopy and 9 with small early adenomas)	advanced neoplasia: 59.4 years [± SD: 6.9]; HCs: 52 years [±(SD: 12)]	France	°previous history of colorectal surgery, inflammatory or infectious bowel diseases and patients with need for emergency colonoscopy °use of antibiotics for at least 2 months prior to enrolment.	No statistical differences in gender, BMI, history of polyps, familial history of polyps/CRC/cancer, diabetes mellitus, dyslipidemia, specific diet and medication of the two groups only the mean age of patients were significantly higher. No adjustments made.
Chen, 2012 Case control	Gut microbiome	Colorectal cancer cases	43: 21 CRC, 22 HCs	CRC: 64 years (37-78) HCs: 64 years (37-84)	China	°diabetes, infectious diseases or particular diets. °taking medications at the time of sample collection °use of antibiotics within at least one month of sample collection °watery stool, stool too thin to collect	CRCs and HCs were matched on the basis of the design for age and gender and the BMI
Feng, 2015 Case control	Gut microbiome	Colorectal cancer cases	103: 46CRC, 57 HCs	CRC: 67.1 years ; HC: 63.8 years	Austria	°use of probiotics or antibiotics within the last 3 months	CRCs and HCs were matched on the basis of the design for age, gender and BMI

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Fukugaiti, 2015 Case control	Gut microbiome	Colorectal cancer cases	17: 7 CRC, 10 HCs	CRC: 65.4 years [±SD: 1.1] HCs: 54.8 years [±SD: 1.3]	Sao Paulo, Brazil	°patients who had taken antibiotics °any systemic infection	Patients with cancer were significantly older than patients without cancer. No adjustments made.
Kanazawa, 1996 Case control	Gut microbiome	Colorectal cancer cases	27: High risk group: 13 male patients (who had previously undergone surgery for sigmoid colon cancer and who developed new second or third colonic epithelial neoplasia.), 14 male HCs	High risk group: 60 years [±SD:7.3] ; HCs: 62.2 years [±SD: 7.4]	Tochigi, Japan	°patients with lesions that might have been overlooked at the initial preoperative examination °neither the high risk or control groups had been receiving anticancer drugs, antibiotics, medicines containing living Lactobacillus or Bifidobacterium and antiphlogistic medicines.	CRCs and HCs were matched on the basis of the design for age and gender
Kasai, 2016 Case control	Gut microbiome	Colorectal cancer cases	58: 9 CRC (3invasive cancer, 6 carcinoma in adenoma), 49 HCs	CRC: 54.3 years [±SD: 7.9] HCs: 48.8 years (SD: 8.2)	Yokkaichi, Japan	°current use of antibiotics °history of or current chronic bowel or liver disease °history of chemotherapy or radiation therapy and regular use of immunomodulators (steroids, interferons, etc.) or probiotics °persons who did not provide consent °age >65years	Total cholesterol and high-density-lipoprotein cholesterol levels were significantly lower in the cancer subjects. The average age and BMI of the cancer subjects were significant higher than those of the control subjects. Smoking and alcohol intake results are the same in both groups. No adjustments were made.
Mastromarino, 1978 Case control	Gut microbiome	Colorectal cancer cases	20: 13 CRC, 7 HCs	CRC: 58 years ; HCs: 46 years	New York, USA	°antibiotic treatment or other therapy for at least 4 weeks prior to collection of the fecal samples. °No laxatives	Not mentioned

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
						were used within 1 week of stool collection	
Mira-Pascual, 2014 Case control	Gut microbiome	Colorectal cancer cases	20: 15 case (7tumor, 8 polyp), 5 HCs	CRC: 71.1 years [SD: 10.1] Polyp: 63.3 [±SD: 13.1] HCs: 52.6 [±SD: 15.2]	Oviedo, Asturias (Spain)	°patients with hereditary syndromes (familial adenomatous polyposis (FAP) and hereditary non polypoid colorectal cancer (HNPCC)	Cases and HCs were matched on the basis of the design for age and gender.
Ohigashi, 2013 Case control	Gut microbiome	Colorectal cancer cases	152: 93 CRC, 49 HCs (22diagnosed with adenoma in the past 3 years, and 27 did not have adenomas)	CRC: 68.9 years [±SD: 12.1] adenoma: 66.6 [±SD: 9.2] ; non-adenoma: 65.6 [±SD: 13.5]	Tokyo, Japan	°history of colectomy or proctectomy °obstructive CRC °patients treated with antibiotics at the point of hospitalization °feces from patients taken after starting pre-operative preparation with bowel cleansing and with oral antibiotics (CRC)	Univariate analysis (the 3 groups did not differ significantly in age, gender, BMI or in past histories of cholecystectomy, appendectomy or breast cancer).
Vogtmann, 2016 Case control (same samples as Ahn 2013)	Gut microbiome	Colorectal cancer cases	104: 52 CRC, 52 HCs	CRC:61.8 years; HCs: 61.2 years	CRC: 75,0% white, 23,1% black, 1,9% others ; HC: 90,4% white, 5,8% black, 3,8% others (USA)	Not mentioned	CRCs and HCs were matched on the basis of the design for gender and BMI
Weir, 2013 Case control	Gut microbiome	Colorectal cancer cases	21: 10 CRC, 11 HCs	CRC: 63.7 years; HCs: 40.7 years	Colorado, USA	°use of antibiotics within two months of study participation °regular use of NSAIDs, statins or probiotics °chronic bowel disorders or food allergies/dietary restrictions °chemotherapy or radiation treatments prior to surgery	Univariate analysis (the majority of taxa that significantly differed in stool samples between healthy and CRC cohorts were a result of disease status and not of differences in age or BMI)

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Wu, 2013 Case control	Gut microbiome	Colorectal cancer cases	39: 19 CRC, 20 HCs	CRC: 58.3 years [±SD: 8.7] HCs: 53.2 years [±SD: 5.4]	Beijing, China	°complicating diseases such as diabetes or hypertension °family history of CRC °antibiotic use within 3 months prior to the study participation	The HCs were selected based on matched sex, age, and BMI
Yu, 2015 Case control	Gut microbiome	Colorectal cancer cases	128: 74 CRC, 54 HCs	CRC: 67 years; HCs: 63 years	China	°use of antibiotics within the past 3 months °on a vegetarian diet °invasive medical intervention within the past 3 months °past history of any cancer °inflammatory disease of the intestine.	colonoscopy (when correcting for colonoscopy, the differences in richness and alpha diversity exhibited p>0,5)
Zackular, 2014 Case control	Gut microbiome	Colorectal cancer cases	60: 30 CRC, 30 HCs	CRC: 59.4 years [±SD: 11] HCs: 55.3 years [±SD: 9.2]	CRC: non-hispanic white: 28, other:2; HCs: non-hispanic white: 21, other:9 (USA)	°known HIV or chronic viral hepatitis °known NHPCC or FAP °inflammatory bowel disease °any surgery, radiation or chemotherapy for their current CRC °colonic adenoma	There were no significant differences in age or current medication use among both groups. However, among our samples, men, whites and those with greater BMI were more likely to have colorectal cancer. No adjustments.
Zeller, 2014 Case control	Gut microbiome	Colorectal cancer cases	114: 53 CRC, 61 HCs	CRC: 68.5 years; HCs: 63 years	France	°previous colon or rectal surgery, °inflammatory or infectious injuries of the intestine. °Patients with need for emergency	BMI and gender did not differ significantly when adjusted in analysis, age does. No adjustments, sensitivity analysis.
Alexander, 2016 Case control	Gut microbiome	Colorectal cancer cases	18 subjects; 54 samples: 18 samples on tumor, 18 samples 5cm off tumor and 18	76 years [55-85]	London, UK	°patients undergoing emergency surgery °patients treated with neoadjuvant chemotherapy and/or	Not mentioned

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			samples 10cm off tumor			radiotherapy. °patients who had been on antibiotics or probiotic therapies within the previous six weeks °history of Familial Adenomatous Polyposis (FAP) °patients with inflammatory bowel disease °undergone previous colorectal surgery	
Allali, 2015 Prospective nested case control samples	Gut microbiome	Colorectal cancer cases	22 subjects; 44 samples: 22 samples on tumor, 22 samples with distance off tumor unknown	63.6 years [42-88]	US	Not mentioned	No significant confounding by region of the colon, age, gender.
Allali, 2015 Case control (Spain)	Gut microbiome	Colorectal cancer cases	23 subjects; 46 samples: 23 samples on tumor, 23 samples 5cm off tumor	69.8 years [49-85]	Spain	Not mentioned	No significant confounding by region of the colon, age, gender.
Burns, 2015 Case control samples	Gut microbiome	Colorectal cancer cases	44 subjects; 88 samples: 44 samples on tumor, 44 matched adjacent tissue samples	64.9 years [±SD:16.7]	Minnesota	°no availability of patient-matched normal and tumor tissue specimens	Age, gender, tumor stage and tumor site (but none of these factors were found to have a significant impact).
Chen, 2012 Case control	Gut microbiome	Colorectal cancer cases	27 subjects; 91 samples: 27 samples on tumor, 27 samples 2-5cm from the tumor, 27 samples 10-20cm from the tumor	61 years [37-81]	China	°diabetes, infectious diseases or particular diets °taking medications at the time of sample collection, or used antibiotics within at least one month of sample collection	Matched on the basis of the design for age, gender and BMI.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Chen, 2012 Case control	Gut microbiome	Colorectal cancer cases	66 subjects with gut swab samples: CRC: 32; HC: 34	CRC: 65 years (37-86) HCs: 56 years (42-77)	China	°diabetes, infectious diseases or particular diets °taking medications at the time of sample collection, or used antibiotics within at least one month of sample collection	CRCs and HCs were matched on the basis of the design for age, gender and BMI.
Flemer, 2016 Case control	Gut microbiome	Colorectal cancer cases	115: CRC: 59,HCs: 56	did not mentioned	Cork	°personal history of CRC, IBD or IBS. °treated with antibiotics in the month prior to surgery	Age matched study design
Gao, 2015 Case control	Gut microbiome	Colorectal cancer cases	22:11 CRC; 11 HCs	CRC: 68 years {±SD: 7.3] HCs: 71 [±SD: 5.4)]	Shanghai, China	°age >75y °usage of antibiotics and additional gastrointestinal disorders °evidence of infection	Age, BMI and gender are not significantly different between groups
Kohoutova, 2014 Case control	Gut microbiome	Colorectal cancer cases	150: 90 in 30 CRC patients (biopsies from caecum, transverse colon and rectum in all patients); 60 in 20 HCs (biopsies from caecum, transverse colon and rectum in all patients)	CRC: 67 years [±SD: 11)] HCs: 55 [±SD: 15)]	Czech Republic	Not mentioned	Not mentioned
Marchesi, 2011 Case control	Gut microbiome	Colorectal cancer cases	6 subjects; 12 samples: 6 samples on tumor and 6 from adjacent non-malignant tissue ("off-tumor")	CRC: 63.5 years (49-71)	Nijmegen	Not mentioned	Not mentioned
Mira-Pascual, 2014 Case control	Gut microbiome	Colorectal cancer cases	20: 15 case (7tumor, 8 polyp); 5 HCs	CRC: 71.1 years [±SD: 10.1] Polyp: 63.3 years [±SD:	Oviedo, Spain	°hereditary syndromes (familial adenomatous polyposis (FAP) and	Age-matched and sex-matched subjects

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
				13.1)] HCs: 52.6 years [±SD: 15.2)]		hereditary non polypoid colorectal cancer (HNPCC))	
McCoy, 2013 (Case control samples	Gut microbiome	Colorectal cancer cases	10 subjects; 19 samples: 10 samples on-tumor and 9 non-malignant samples from adjacent mucosa	37-78 years	Caucasians, USA	Not mentioned	Not mentioned
Warren, 2013 Case control samples	Gut microbiome	Colorectal cancer cases	65 subjects; 130 samples: 65 samples on tumor 65 control samples (off-tumor)	Not mentioned	Canada	Not mentioned	Not mentioned
Youssef, 2018 Case control	Gut microbiome	Stomach cancer Colorectal cancer	83 patients, who were diagnosed with different GIT neoplasms, and 13 healthy individuals	Controls: 43.8 (19–65) Stomach: 69.4 (36–98) Pancreas: 62.3 (57–67) Small intestine: 61.5 (39–79) Colon: 74.8 (64–84) Rectum: 73.6 (39–85) Treated: 66.9 (53–78)	Finland	Not available	Tumor site, Age, Gender
Goedert, 2015 Case control	Gut microbiome	Breast cancer cases	96: 48 breast cancer patients, 48 HCs	62 years [±SD: 6.86]	Colorado (86% non-Hispanic)	antibiotic exposure	Not mentioned
Goedert, 2018 Case control	Gut microbiome	Breast cancer	48 postmenopausal breast cancer cases and 48 contemporaneous, postmenopausal, normal-mammogram, age-matched controls	Cases: median 63 (57.5–67) Controls: median 61 (59–64.25)	USA	history of previous cancer (except non-melanoma skin cancer), inflammatory bowel disease, diverticulitis, or bariatric surgery; other gastrointestinal surgery within 6 months; any antibiotic	age, BMI, characteristics relevant to breast cancer risk (details not available); adjustment on oestrogens

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
						prescription within 6 months; and any oestrogenic or other hormone prescription within 12 months	
Zhu, 2018 Case control	Gut microbiome	Breast cancer	18 premenopausal breast cancer patients, 25 premenopausal healthy controls, 44 postmenopausal breast cancer patients, and 46 postmenopausal healthy controls	Premenopausal cases: 37.06 ± 5.23 Postmenopausal cases: 57.45 ± 7.41	China	diarrhea, diabetes, ulcerative colitis, Crohn's disease, or other infectious diseases antibiotics, steroid hormones, Chinese herbal medicine (including oral, intramuscular, or intravenous injection), or probiotics such as yogurt during the 3 months before fecal sample collection chemotherapy, radiation, or surgery prior to fecal sample collection	Age, BMI, Age at menopause, ethnicity
Alanee, 2019 Cohort	Gut and urinary microbiome	Prostate cancer	30 patients 60 samples, one urine and one fecal per patient	Not available	USA	patients who received antibiotic therapy within 28 days from their prostate biopsy from the study	Not available
Golombos, 2018 Case control	Gut microbiome	Prostate Cancer	20 men with either benign prostatic conditions (n = 8) or intermediate or high risk clinically localized prostate cancer (n = 12)	Median 64 years	USA	recent antibiotic use, significant gastrointestinal disorders, hormonal or systemic therapy for prostate cancer	age, body mass index (BMI), and prostate size

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Liss, 2018 Case control	Gut microbiome	Prostate Cancer	105 samples (64 with cancer, 41 without cancer)	PC: 66.5 (62–70) H: 65 (60–68.5)	USA	Any antibiotic therapy in the last 6 mo	age, BMI, Diabetes, Race, PSA, Abnormal prostate examination, family history, Lower urinary tract symptoms, Number of negative prostate biopsies, active surveillance, Gleason score, Number of aberrant microbiome profiles
Qi, 2019 Case control	Gut microbiome	Gastric cancer	116 gastric cancer patients and 88 healthy controls	Not available	China	other ethnicities except Han Chinese; age under 18 years; living in Shanxi Province for less than 10 years; exposure to antibiotics, probiotics, immunomodulators and acid blockers for the past month; history of gastrointestinal tract infections, such as Campylobacter and Salmonella, for the past month; presence of digestive diseases except gastric cancer; presence of type 2 diabetes, autoimmune diseases and other malignant tumors; and history of gastrointestinal tract surgery, chemoradiotherapy, and cholecystectomy.	adjustment for sex and age
Ren, 2019 Case control	Gut microbiome	Liver cancer	150 patients with HCC, 40 patients with cirrhosis and 131 healthy controls	Liver cancer: 46.95±5.6 Early HCC: 49.67±8.56	China	Not available	age, gender, BMI, clinicopathological data, CT scan, histopathology images and diet habit

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
				Control: 48.65±6.61			
Feng, 2019 Case control	Gut microbiome	Thyroid cancer	30 TC patients and 35 healthy controls	TC: 48.47 6.52 HCs: 46.03 7.96	China	Not available	gender, age, BMI, T status, N status, M status
Gong, 2013 Case control	Laryngeal microbiome	Laryngeal cancer cases	29 pat. with laryngeal squamous cell carcinoma (LSCC), 31 controls with vocal cord polyps (H).	In the (LSCC) group, 11 patients were younger than or exactly 60 years while 18 patients were strictly older than 60 years. In (H), 131 patients were younger than or exactly 60 years while 18 patients were strictly older than 60 years.	China	°pat. with a history of antibiotic use in the previous 3 months °pat. with a history of active bacterial or viral infections in other parts of the body ° controls not free of cancer °controls with evidence of epithelial dysplasia.	No corrections, but each patient is his or her own control.
Shin, 2017 Case control	Oral microbiome	Head and neck cancer	72 tissue samples (normal, primary, metastatic) originating from the oral cavity, larynx, pharynx and lymph nodes of 34 HNSCC subjects	59 ± 5.6 years	USA	Not available	age, gender
Wang, 2017 Case control	Oral microbiome	Head and neck cancer	242 samples from 121 patients	63 ± 11	USA	insufficient reads	age, gender, race, tumor localization, T-stage, N-stage, smoking history, alcohol use, previous treatment
Hayes, 2018 Nested case-control	Oral microbiome	Head and neck cancer	129 incident patient cases of HNSCC and 254 matched controls	CPS-II cohort:71.0 [6.4] years PLCO cohort: 62.7 [4.8] years	USA	Not available	controls matched on age, sex, race/ethnicity, and time since mouthwash collection adjustment on age, sex, race, cohort, smoking, alcohol, and

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
							oral human papillomavirus-16 status
Hu, 2016 Case control	Oral microbiome	Oral cancer	16 patients with OSCC, 10 with OLK, and 19 HCs	Not available	China	Not available	Not available
Banerjee, 2017 Case control	Oral microbiome	Oral cancer	100 OCSCC patient samples as well as 20 cancer adjacent normal controls (matched) and 20 oral tissue (uvula) from healthy individuals (non-matched controls)	Not available	USA	Not available	Clinically normal samples adjacent to the cancers are referred here as "matched controls
Lee, 2017 Case control	Oral microbiome	Oral cancer	Normal (n = 127) Epithelial precursor lesion (n = 124) Cancer (n = 125)	Normal: 52 ± 14 Epithelial precursor lesion 50 ± 11 Cancer 53 ± 10	Taiwan	Not available	sex, age, alcohol consumption, betel chewing, cigarette smoking, and family history of cancer
Mok, 2017 Case control	Oral microbiome	Oral cancer	9 normal, 9, oral potentially malignant disorders, 9 malignant lesions	Normal: 40 OPMD: 54 Cancer: 60	Malaysia	currently having malaise, pregnancy or lactation, infected with HIV, under antibiotic medication within the past 3 months, involve surgery, chemotherapy or radiation treatment within the past 1 month	Not available
Wolf, 2017 Case control	Oral microbiome	Oral cancer	11 patients and 11 healthy controls	Patients: 61.6 yrs., SD = 8.2 yrs Healthy controls: 46.7 yrs., SD = 15.1 yrs	Austria	undergone long-term antibiotic use (as per their medical histories), vaccinated in the six months prior to study inclusion	age, alcohol consumption, tumour size, lymph node status, smoking habit, and tumour HPVpositivity
Zhao, 2017 Case control	Oral microbiome	Oral cancer	80 samples	Not available	China	Not available	Not available

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Hsiao, 2018 Case control	Oral microbiome	Oral cancer	138 OSCC cases and 151 controls	Not available	Taiwan	(i) diagnosis of pathologically confirmed squamous cell carcinoma located in the oral cavity, oropharynx, hypopharynx and larynx (ICD-10 codes: C00-C10, C12-C14, C32); (ii) no previous cancer diagnosis; (iii) age ≧20 years; (iv) Han ethnicity and (v) ability to provide informed consent	sex, age, education, consumption of alcohol, betel quids and cigarettes, and dental care habits
Lim, 2018 Case control	Oral microbiome	Oral cancer	normal healthy controls (n=20), high-risk individuals (n = 11) and OCC and OPC patients (n = 52)	young normal healthy controls: 26 elderly normal healthy controls: 61 high-risk controls: 59 OCC or OPC: 65	Australia	local and/or systemic antibiotics prior to sample collection	Gender, Age, Race, Smoking, Alcohol, Tumor characteristics
Perera, 2018 Case control	Oral microbiome	Oral cancer	25 OSCC cases and 27 FEP controls	Not available	Sri Lanka	Not available	demographic and clinical characteristics
Yost, 2018 Case control	Oral microbiome	Oral cancer	15 samples, including four tumour sites from OSCC subjects, four tumour-adjacent sites from OSCC subjects, four sites from healthy patients who matched the locations of the tumour sites and	Not available Range from 40-64	USA	any treatment, including surgery, radiotherapy and/or chemotherapy, if they were pregnant or nursing, received antibiotics or periodontal therapy in the previous 3 months, if they had any systemic illnesses, including diabetes, or if	Controls matched for age and sex, non-smokers Only male >15 natural teeth and were in good general health

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			three buccal sites in healthy tumour-free subjects that matched the locations of tumouradjacent samples			they had any immunocompromising conditions	
Börnigen, 2017 Case control	Oral microbiome	Oral squamous cell carcinomas	121 oral cancer patients to 242 age- and gendermatched controls	median age 58 years (IQR 53–66)	USA	Case: older than 17 years and not newly diagnosed with a histologically confirmed squamous cell carcinoma of the oral cavity or oropharynx Control: older than 17 years with no history of cancer who were evaluated as an outpatient for any benign condition	age- and gendermatched controls measures of tobacco, alcohol and marijuana use and medical and dental history.
Schmidt, 2014 Case control	Oral microbiome	Oral cancer cases	For study 1 (discovery cohort): 5 pat. with oral cancer (OC1) and 5 healthy controls (H1). For study 2 (confirmation cohort): 10 pat. with oral cancer (OC2), 1 pat. with carcinoma in situ (CIS), 8 pat. with pre-cancer stages (PRE) and 20 healthy controls (H2).	(OC1) 69.2 years [range 62-84], (H1) no age mentioned. (OC2) 59 years [range 39-78], in (CIS) the pat. was 84 years, (PRE) 68.4 years [range 49-79] and (H2) 30 years [range 30-30].	U.S.A.	°cancer pat. from study 1 were excluded from study 2.	No corrections, but each patient is his or her own control.
Berkovitz, 2016 Case control	Oral microbiome	Oral cancer cases	60 subjects: 20 pat. (14 m. and 6 fem.) with oral squamous	(OSCC) 62 years [range 44-86], (H)	Hungary	°controls not free of oral pathology.	

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			cell carcinoma (OSCC), 40 healthy controls (22 m. and 18 fem.)	67 years [range 49-82].			
Henrich, 2014 Case control	Oral microbiome	Oral cancer cases	2 pat. with Fanconi Anaemia and oral squamous cell cancer (FAC)+ (FAC2), 2 pat. with Fanconi Anaemia and benign oral lesion (FAB) + (FAB2), and 5 healthy controls (H).	(FAC) was 41 years, (FAC2) was 27 years, (FAB) was 27 years, (FAB2) was 33 years and (H) 45.8 years [range 32-43].	Germany	/	
Homann, 2000 Case control	Oral microbiome	Oral cancer cases	326 volunteers: 26 pat. with a malignant tumor of the oral cavity (T), 64 alcoholics (A), 24 pat. seeking a dental examination or treatment (DE), 90 unemployed volunteers (UN) and 114 healthy volunteers (H).	93 of the participants younger than 41 years, 162 were between 41-58 years and 74 older than 58 years.	Finland	°ex-smokers with a cessation of less than 5 years °treatment with oral antiseptic or antibiotics in the past month °food or fluid intake, smoking or tooth brushing in the past 90 min. °recent alcohol intake or measurable amount of alcohol in the saliva by head space GC.	Adjustments for age, smoking, alcohol, tooth brushing, tooth loss, eating between meals, periodontitis, frequency of dentist visits, mouthwash use, dentures and self-reported dry mouth and burning mouth.
Marttila, 2013 Case control	Oral microbiome	Oral cancer cases	90 subjects: 30 pat. with oral squamous cell carcinoma (OSCC), 30 pat. with oral lichenoid disease (OLD) and 30 healthy controls (H).	(OSCC) 65.6 years [range 39-85], (OLD) 54 years [range 24-74] and (H) 30.4 years [range 19-56].	Finland	°pat. with antimicrobial therapy within the past 7 days °pat. diagnosed with human immunodeficiency virus or hepatitis virus infection.	Adjustments for impact of drinking and smoking on acetaldehyde production.
Guerrero-Preston, 2016	Oral microbiome	Head and neck cancer cases	17 pat. with head and neck squamous cell carcinoma	(OPSCC) mean age of 62 years, (OCSCC) mean	U.S.A.	/	

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
Cohort study (1 year follow-up of patients and controls nested in cohort study)			(HNSCC), 25 healthy controls (H). 11 pat. with an oropharyngeal squamous cell carcinoma (OPSCC): 7 were HPV+ and 4 HPV-. 6 pat with an oral cavity squamous cell carcinoma (OCSCC), all HPV-.	age of 66. Mean age of controls is not mentioned.			
Wang, 2019 Case control	Oral and lung microbiome	Lung cancer	51 patients with primary bronchogenic carcinoma and 15 healthy controls	Controls: 56.9±6.1 LAC: 54.8±10.7 LSCC: 62.4±8.4 SCLC: 62.4±8.4	China	the participant manifested other basic pulmonary diseases, oral disorders or the presence of removable partial dentures or orthodontic appliances; systemic diseases, such as diabetes mellitus, gastritis, hepatitis and other cancers in addition to PBC; immune-compromising diseases, such as human immunodeficiency virus (HIV) or ongoing immunosuppressive therapy; and other diseases known to affect the oral and airway microbiota	age, gender, BMI, smoking history, drinking history
Lu, 2016 Case control	Oral (Tongue coat) microbiome	Liver cancer	35 patients in the early stages of LC patients with	Cases: 50.30 ± 5.86	China	Cases: diagnosed with intrahepatic cholangiocarcinoma or underwent prior	No significant differences between the groups in age, sex distribution, and body mass index.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			cirrhosis 25 healthy subjects	Controls: 48.20 ± 6.03		anticancer treatment co-infected with HCV, HIV or other secondary organism (bacteria and fungi) infection Controls: hypertension, diabetes, obesity, metabolic syndrome, irritable bowel syndrome, non-alcoholic fatty liver disease, coeliac disease and liver cirrhosis antibiotics, probiotics or both within 8 weeks before enrolment, or consumption of unhealthy substances (including alcohol, cigarette/ tobacco and drug abuse)	
Yang, 2019 Case control	Oral microbiome	colorectal cancer	231 incident CRC cases and 462 controls	Age at enrollment available by category (no mean)	USA	antibiotics treatment during the seven days prior to sample collection	controls matched to cases based on age, race, smoking, season-of-study enrollment and recruitment method
Flemer, 2018 Case control	Oral microbiome Gut microbiome	Colorectal cancer	CRC (99 subjects), colorectal polyps (32) or controls (103)	Stool CRC: 65.3±10.8 Swab CRC: 65.7±10.9	Ireland	Personal history of CRC, IBD and irritable bowel syndrome.	Age, BMI, gender, tumour size, rectal bleeding, alcohol consumption, smoking status
Han, 2014 Case control	Oral microbiome	Colorectal cancer cases	92 subjects: 47 pat. with colorectal cancer (22 with rectal cancer, 25 with colon cancer) (CRC), 45 healthy controls (H). (CRC) was divided into a	(CRC) 53.24 years [±SD 9.70], (H) 51.57 years [±SD 8.01], calculated for the prior bigger study group containing 45 pat. and 47	China	°controls with gastrointestinal diseases, oral diseases, malignant tumor and cancer related symptoms in the last 2 years.	Adjustments for age, weight, smoking, hypertension and diabetes. Chemotherapy and surgical treatment are possible reasons for a thicker tongue coating; no adjustment was made.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			tick group, with 9 cases, and a thin group with 5 cases.	controls. Thick group: 53.78 years [±SD 14.43], thin group: 48.60 years [±SD 8.56], and the 7 healthy controls: 52.14 years [±SD 10.63].			
Han, 2016 Case control	Oral microbiome	Colorectal cancer, lung cancer and gastric cancer cases	386 subjects: 90 pat. with colorectal cancer (CRC), 96 pat. with lung cancer (LC), 100 pat. with gastric cancer (GC) and 100 healthy controls (H).	(CRC) 55.45 years [±SD 1155], (LC) 55.14 years [±SD 9.80], (GC) 56.20 years [±SD 10.24] and (H) 53.57 years [±SD 8.32].	China	°controls with digestive diseases, respiratory diseases, oral disease, malignant tumor and cancer related symptoms in the last two years.	Adjustments for age, sex, BMI, smoking status, hypertension and diabetes.
Peters, 2017 Case control	Oral microbiome	Esophageal Cancer	EAC cases: 81 EAC matched controls: 160 ESCC cases: 25 ESCC matched controls: 50	EAC cases: 68.0 6.7 EAC matched controls: 68.0 6.6 ESCC cases: 6.6 6.5 ESCC matched controls: 66.8 6.4	USA	Not available	adjusting for BMI, smoking, and alcohol
Chen, 2015 (Case control	Oral microbiome	Esophageal cancer cases	235 subjects: 87 pat. with esophageal squamous cell carcinoma (ESCC), 63 subjects with dysplasia (DYS) and 85 healthy controls (H).	(ESCC) 64.8 years [±SD 8.0], (DYS) 65.5 years [±SD 7.6], (H) 66 years [±SD 7.3].	China	°cases collected during November of 2010 and March of 2011 to avoid confounders (i.e. ambient temperatures and different dietary habits during different seasons) °cases with no histopathological confirmation °cases with no complete	Case-control matching for sex and age. Adjustments for education, smoking, alcohol drinking, family history of ESCC, MFT, times of tooth brushing per day, daily consumption of pickled vegetables and daily consumption of fresh fruits.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
						questionnaire or no saliva samples.	
Olson, 2017 Case control	Oral microbiome	Pancreatic cancer	40 newly diagnosed PDAC patients, 39 IPMN patients, and 58 controls,	Most of case >=70 years	USA	current smokers and users of antibiotics	gender and age
Fan, 2018 Case control	Oral microbiome	Pancreatic cancer	361 incident adenocarcinoma of pancreas and 371 matched controls	CPS II cohort: 73.7 PLCO cohort: 63.8	USA	Not available	Controls were matched to cases by cohort, age (5-year), sex, race (white, other) and calendar year of oral wash collection. Confounders: age, sex, race, body mass index (BMI), smoking status, alcohol consumption and history of diabetes
Fan, 2016 Prospective nested case control	Oral microbiome	Pancreatic cancer cases	CPS II cohort: 170 cases with primary pancreatic adenocarcinoma (PAD1), 170 matched controls (H1). PLCO: 191 cases with primary pancreatic adenocarcinoma (PAD2), 201 matched controls (H2).	(PAD1) 73.7 years [±SD 5.7], (H1) 73.7 years [±SD 5.7]. (PAD2) 63.8 years [±SD 5.2], (H2) 63.8 years [±SD 5.4].	U.S.A.	°cases with a history of cancer prior to pancreatic adenocarcinoma (except non-melanoma skin cancer) °controls with cancer prior to selection.	Case-control matching for age, sex, race and calendar year. Adjustments for pancreatic cancer status, race, BMI, smoking status, alcohol consumption status and history of diabetes.
Farrell, 2012 Case control	Oral microbiome	Pancreatic cancer cases	For discovery phase: 10 pat. with pancreatic cancer (PC1) and 10 matched controls (H1). For independent validation phase: 28 pat. with pancreatic cancer (PC2), 28 matched controls	(PC1) 66.5 years [±SD 8.9], (H1) 66.4 years [±SD 10.5]. (PC2) 69.9 years [±SD 11.6], (H2) 65.1 years [±SD 10.1] and (CP) 57.8 years [±SD 11.0].	U.S.A.	°cases with evidence of locally advanced pancreas cancer due to arterial involvement or direct extension into adjacent organs, metastatic pancreatic cancer, chemotherapy or radiation therapy prior to saliva collection °cases with diagnosis of	Case-control matching for age, gender and ethnicity. Corrections for smoking and drinking history.

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			(H2) 27 pat. with chronic pancreatitis (CP).			other malignancies within 5 years from the time of saliva collection.	
Torres, 2015 Case control	Oral microbiome	Pancreatic cancer cases	108 subjects: 8 with pancreatic cancer (PC), 78 with other diseases (OD) and 22 healthy controls (H).	(PC) 71.1 years, other ages are not (clearly) mentioned.	U.S.A.	°participants undergoing active chemotherapy or radiation therapy or use of antibiotics two weeks prior to saliva collection °participants with invasive surgery in the past year.	
Hu, 2015 Case control	Oral microbiome	Gastric cancer cases	For tongue images, 74 pat. with gastric cancer (GC) and 72 healthy controls (H). For the samples, 34 pat. with gastric cancer, from who 16 had thin coatings (GCtn) and 18 had thick tongue coatings (GCtk), and 17 healthy controls (Hs).	(GC) 57.46 years [±SD 8.43], (H) 54.55 years [±SD 9.63].	China	°controls with stomach discomfort over the past three years, malignant tumors, oral diseases or gastric diseases °subjects that had used any antibiotics within the past two months.	Adjustments for chemotherapeutics and surgery, BMI, diabetes, hypertension, smoking and drinking.
Banerjee, 2017 Case control	Ovarian microbiome	Ovarian cancer	99 ovarian cancer samples and 20 matched (tissue adjacent to the tumor deemed noncancerous by pathological analysis) and 20 unmatched control samples	Not available	USA	Not available	tissue adjacent to the tumor deemed noncancerous by pathological analysis referred as matched controls
Gong, 2017 Case control	Pharyngeal microbiome	Laryngeal carcinoma	68 subjects with laryngeal cancer and	57.1 (±11.5) years	China	history of use of hormones or antibiotics	Not available

RESEARCH ARTICLE	ANALYSED MICROBIOME	DEFINITION OF THE CASES	SAMPLE SIZE	MEAN AGE OF PARTICIPANTS	ORIGIN OF SUBJECTS	EXCLUSIONS	CONFOUNDERS/ADJUSTMENTS
			28 subjects with vocal cord polyps			use during the previous three months or an active bacterial or viral infection in another part of the body	
Cavarretta, 2017 Case control	Prostate microbiome	Prostate cancer	16 participants: tumor, peri-tumor, and nontumor tissues after radical prostatectomy	Not available	Italy	self-reported lower urinary tract symptoms or a pathologic International Prostate Symptom Score were excluded	diet, lifestyle
Salava, 2016 Case control	Skin microbiome	Skin cancer	15 cutaneous melanomas and 17 benign melanocytic nevi	Melanomas: 69.1 ± 16.9 Melanocytic nevi: 52.9 ± 17.6	Finland	unsuitable lesion type (dysplastic nevus or other), or insufficient amount of DNA in the samples.	Not available
Bučević Popović, 2018 Case control	Urinary microbiome	Bladder cancer	12 male patients diagnosed with bladder cancer, and from 11 healthy, age-matched individuals	Not available	Croatia	positive history of sexually transmitted or recent urinary infections, diabetes and obesity	Not available
Bi, 2019 Case control	Urinary microbiome	Bladder cancer	29 bladder cancer patients and 26 non-cancer patients	67.4 years	China	Not available	Not available

Supplementary Table S4: The results of the articles included in the systematic review, presenting the quality assessment, microbiome measurement and composition and the differences between cases and controls.

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Liu, 2018 **Case control**	(LC) Lung cancer (H) Healthy controls	Protected bronchial specimen brushing samples	Airway microbiome	16S rRNA gene sequencing (V3-V4 region)	(LC) ↑ Streptococcus (H) ↑ Staphylococcus	lung cancer-associated microbiota profile is distinct from that found in healthy controls, and the altered cancer-associated microbiota is not restricted to tumor tissue	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Tsay, 2018 **Case control**	(LC) Lung cancer (NLC) Non-cancer diagnoses (H) Healthy controls	Airway brushings	Airway microbiome	16S rRNA gene sequencing (V4 region)	(LC) ↑ Streptococcus and Veillonella	several transcriptomic signatures previously identified as relevant to lung cancer pathogenesis are associated with enrichment of the lower airway microbiota with oral commensals	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Lee, 2016 **Cohort**	(LC) Lung cancer (BML) Begnin mass-like lesion	Bronchoalveolar fluid	Lung microbiome	16S rRNA gene sequencing (V1-V3 region)	(LC) ↑ Firmicutes, TM7, Veillonella, Megasphaera	Number of OTUs: higher in LC compared to BML Chao 1 and Shannon: higher in LC compared to BML	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Carpagnano, 2014 **Case control**	(NSCLC) Non-small cell lung cancer, (H) Healthy controls.	Exhaled breath condensate (EBC) and bronchial brushing.	Lung microbiome	Subculturing of the colonies and incubated using 3 different agars.	12 of the (NSCLC): colonized with fungi. 0 of the (H) colonized. Mostly Aspergillus niger (5), by Aspergillus ochraceus (3) and Penicillum spp. (4).	Fungal colonization by Aspergillus niger, Aspergillus ochraceus and Penicillium ssp. in the EBC of (NSCLC) and not of the (H).	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Hosgood, 2014 Case control	(LC) Lung cancer, (H) Healthy controls.	Buccal and sputum samples	Lung microbiome	DNA extraction and 16S rRNA gene analysis of the V1-V2 region.	All buccal samples: mainly Bacteroidetes, Firmicutes, Proteobacteria, Actinobacteria, and Fusobacteria. Sputum of (LC): ↑ Granulicatella, Abiothrophia and Streptococcus. No difference in the lung microbiota between the controls from Laibin and of Reshui, but a difference in the cancer pat. from those two villages. Reshui: ↑ Proteobacteria, ↑ Neisseria, ↓ Bacilli and Streptococcus.	Clear differences between the sputum samples of (LC) and (H).	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Cameron, 2017 Case control	(LC) Lung cancer	Sputum samples	sputum microbiome	16S rRNA gene sequencing	(LC) ↑ Granulicatella adiacens correlated with Enterococcus sp. 130, Streptococcus intermedius, Escherichia coli, S. viridans, Acinetobacterjunii, and Streptococcus sp. 6	Spontaneous sputum appears to be a viable source of bacterial biomarkers which may have utility as biomarkers for LC status and stage.	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Aviles-Jimenez, 2016 Case control	(ECCA) Extrahepatic cholangiocarcinoma (BBP) benign pathology of the common bile duct	Epithelial cells of the bile duct	Bile duct microbiome	DNA extraction and 16S rRNA gene analysis of the V4 region + PCR.	All groups: mainly Proteobacteria. (ECCA) ↑ Fusobacteria, Acidobacteria and Planctomycetes, ↑ Methylophilaceae, Fusobacterium, Prevotella, Helicobacter and Campylobacter, ↓ Nesterenkonia, Rothia and Mesorhizobium.	Significant separation between (ECCA) and (BBP) was observed.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Chng, 2016 Case control	(CCA) Cholangiocarcinoma (OVa) Opisthorchis viverrini associated	Tissue samples of the liver, of the bile duct, bile fluid samples and gastric mucosa samples.	Bile duct microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V6 region.	(CCA) dominated by Dietziaceae, Pseudomonadaceae, Oxalobacteraceae. (CCA) normal tissue ↑ Enterobacteriaceae, Lachnospiraceae, Sphingomonadaceae and Bifidobacteriaceae. (CCA) cancer tissue ↑ Stenothrophomonas.	Gastric tissue microbiome was clearly distinguishable from the bile duct. (CCA) tumor tissue ~ adjacent normal tissue. Significant differences with the controls.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Hieken 2016 Case control	(BBD-non-atypia) Benign breast disease without atypia, (IBC) Invasive breast cancer	Breast tissue	Breast tissue microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V5 region.	(IBC) ↑ Fusobacterium, Atopobium, Hydrogenophaga, Gluconacetobacter and Lactobacillus	Significant differences in microbial composition of breast tissue in benign vs. malignant disease.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Urbaniak 2016 Case control	(BBC) Benign breast cancer (IBC) Invasive breast cancer (H) Healthy controls	Breast tissue	Breast tissue microbiome	DNA extraction and 16S rRNA gene analysis of the V6 region.	(IBC) ↑ Enterobacteriaceae, Bacillus and Staphylococcus	Significant differences in microbial composition of breast tissue in healthy controls vs. malignant disease.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Audirac-Chalifour, 2016 Case control	(CC) Cervical cancer, (SIL) Squamous intraepithelial lesions (H) Healthy controls with no cervical lesions	Cervical scraping swabs and fresh cell biopsies.	Cervical microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V4 region.	(CC) characterised by presence of Sneathia spp. and Fusobacterium spp. and absence of organisms from the Bifidobacteriaceae family. (CC) ↓ Lactobacillus crispatus, Lactobacillus iners and Gardnerella vaginalis, ↑ Fusobacterium necrophorum. (SIL) dominated by Fusobacterium spp, Sneathia spp., Shuttleworhia satelles and Megasphaera spp.	Higher alpha diversity in the (SIL) group and (CC) group than in the (H) + notably different beta diversity in every stage of cervical cancer.	Selection: **** Comparability: ** Exposure: * Total NOS score: 7
Oh, 2015 Case control	(CIN) Cervical intraepithelial neoplasia (H) Healthy controls.	Cervical swab	Cervical microbiome	DNA extraction and 16S rRNA gene analysis of the V1-V3 region.	All groups: dominated by Firmicutes, Actinobacteria, Bacteroidetes, Proteobacteria, Tenericutes, Fusobacteria, and candidate division TM7. Predominance of A. vaginae, L. iners and G. vaginalis and paucity of L. crispatus = risky microbial pattern. Synergistic effect of risky microbial pattern with high risk HPV infection on CIN risk. Different A. vaginae	(CIN): higher numbers of OTU.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
					and L. crispatus ratio between (H) and (CIN).		
Seo, 2016 Case control	(CIN) Cervical intraepithelial neoplasia	Cervical swab	Cervical microbiome	DNA extraction and 16S rRNA gene analysis of the V1-V3 region.	Lactobacillus iners-dominant microbial type B + A. vaginae-dominant microbial type C: ↑ risk of CIN. Synergistic effect between semi-Western diet and microbial type C. No synergistic effect between semi-Western diet and microbial type B.	Semi-Western diet: ↑ risk of CIN + synergistically ↑ risk with the dominance of A. vaginae.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Mannell, 1983 Case control	(EC) Esophagus carcinoma (H) Healthy controls.	Esophageal aspirates	Esophageal microbiome	Incubated and then aerobic and anaerobic subculturing onto selective agar media. Identified by their morphology and biochemical reactivity.	(H) and (EC): mainly Streptococcus viridans, Haemophilus influenza and Neisseria catarrhalis, Klebsiella pneumoniae and Streptococcus group B.	No significant difference in the number and type of bacterial species between (H) and (EC).	Selection: *** Comparability: Exposure: * Total NOS score: 4

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Wang, 2016 Case control	(GC) Gastric cancer	gastric mucosa	Gastric microbiome	16S rRNA gene sequencing (V1-V3 region)	(GC) ↑ Lactobacillus, Escherichia– Shigella, Nitrospirae, Burkholderia fungorum,and Lachnospiraceae	Markedly increased bacterial load in GC ;	Selection: *** Comparability: * Exposure: ** Total NOS score: 6
Castaño-Rodríguez, 2017 Case control	(GC) Gastric cancer	Antral gastric biopsies	gastric microbiome	16S rRNA gene sequencing	(GC) ↑ Lactococcus, Veilonella, and Fusobacteriaceae	enrichment of pro-inflammatory oral bacterial species, increased abundance of lactic acid producing bacteria, and enrichment of short chain fatty acid production pathways in GC	Selection: **** Comparability: * Exposure: ** Total NOS score: 7
Li, 2017 Case control	(GC) Gastric cancer	Endoscopic gastric biopsies	Gastric microbiome	16S rRNA gene sequencing (V3-V4 region)	(GC) ↑ Flavobacterium, Klebsiella, Serratia marcescens, Stenotrophomnonas, Achromobacter and Pseudomonas	GC samples tended to have lower bacterial diversity compared with other samples with similar H. pylori levels. H. pylori colonization results in alterations of gastric microbiota and reduction in bacterial diversity.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Yu, 2017 Case control	(GC) Gastric cancer	gastric tissue	Gastric microbiome	16S rRNA gene sequencing	(High tumor grade vs. Low tumor grade) ↑ Helicobacter pylori ↓ Bacteroidetes & Lactobacillales	gastric non-malignant tissue microbiota features were associated with a known gastric cancer risk factor (family history of UGI cancer) and clinical features (tumor grade and metastasis)	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Yu, 2017 Case control	(GC) Gastric cancer	Paired non-malignant and tumor tissues	Gastric microbiome	16S rRNA gene sequencing (V3-V4 region)	(GC) ↑ Bacteroidetes, Firmicutes, Fusobacteria, and Spirochaetes ↓ Proteobacteria	Comparison with other body sites suggested that stomach microbiota resembled oral microbiota in phylum-level taxonomical profiles, but not in functional profiles.	Selection: **** Comparability: * Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Coker, 2018 Case control	(SG) Superficial gastritis (AG) atrophic gastritis (IM) intestinal metaplasia (GC) Gastric cancer	gastric mucosa	Gastric microbiome	16S rRNA gene sequencing (V4 region)	(GC) ↑ 21 bacterial taxa, including Peptostreptococcus stomatis, Streptococcus anginosus, Parvimonas micra, Slackia exigua and Dialister pneumosintes ↓ 10 bacterial taxa compared to SG	members of oral pathogenic taxa were over-represented and formed strong co-occurrence network in GC compared with other precancerous stages	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Hu, 2018 Case control	(SG) Superficial gastritis (GC) Gastric adenocarcinoma	gastric wash samples	Gastric microbiome	Shotgun Metagenomic sequencing	(GC) ↑ Neisseria, Alloprevotella, and Aggregatibacter ↓ Sphingobium yanoikuyae	GC patients was characterized by reduced species richness, enrichment of 13 bacterial taxa and depletion of 31 taxa	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Liu, 2019 Cohort	(GC) Gastric cancer	Gastric tissues	Gastric microbiome	16S rRNA gene sequencing (V3-V4 region)	(GC) ↑ Prevotella melaninogenica, Streptococcus anginosus and Propionibacterium acnes ↓ Helicobacter pylori, Prevotella copri and Bacteroides uniformis	the stomach microhabitats determined the overall structure and composition ofthe gastric microbiota, regardless ofdifferent GC stage and type.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Aviles-Jimenez, 2014 Case control	(NAG) Non-atrophic gastritis, (IM) Intestinal metaplasia (GC) Intestinal-type of gastric cancer.	Gastric samples	Gastric microbiome	DNA extraction and 16S rRNA gene analysis (no region mentioned).	Gastric tissue of all groups: mainly Proteobacteria and Firmicutes. (GC) ↑Lactobacillus coleohominis and Lachnospiraceae, ↓ 2 TM7, 2 Porphyromonas and Neisseria.	Differences between (GC) and (NAG), but not between (IM) and (GC) or (IM) and (NAG). From (NAG) to (IM) to (GC), the microbiota diversity showed a trend to diminish.	Selection: **** Comparability: * Exposure: * Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Seo, 2014 Case control samples	(GC) Gastric cancer (H) The adjacent normal gastric mucosa.	Gastric tumor tissue and adjacent normal mucosa	Gastric microbiome	DNA extraction and high-throughput RNA sequencing (no further details mentioned).	11/16 samples dominated by H. pylori. (GC) ↓ H. pylori, Propionibacteirum spp., Staphylococcus spp., and Corynebacterium spp. ↑ Clostridium spp. and Prevotella spp.	Significant differences between the (GC) and (H) tissue.	Selection: ** Comparability: Exposure: ** Total NOS score: 4
Nasrollahzadeh, 2015 Case control	(ESCC) Esophageal squamous cell carcinoma, (ED) Esophageal sqoamous dysplasia, (DC) Diseased controls, (H) Healthy controls.	Gastric tissue	Gastric microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V4 region.	All groups: mainly Firmicutes, Bacteroidetes and Proteobacteria. (EC) ↑ Clostridiales and Erysipelotrichales, ↓ Helicobacteraceae.	Significant differences in gastric mucosa of (ESCC) and (ESD) compared to (H).	Selection: *** Comparability: * Exposure: ** Total NOS score: 6
Baxter, 2016 Case control	(CRA) Adenomas (CRC) Carcinomas (H) Healthy	Stool	Gut microbiome	16S rRNA gene sequencing (V4 region)	(CRC) ↑ Porphyromonas assaccharolytica, Peptostreptococcus stomatis, Parvimonas micra, and Fusobacterium nucleatum ↓ Ruminococcoaceae and Lachnospiraceae families	The gut microbiota can be used to differentiate healthy individuals from those with colonic lesions.	Selection: **** Comparability: * Exposure: ** Total NOS score: 7
Sinha, 2016 Case control	(CRC) Colorectal cancer	lyophilized feces	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(CRC) ↓ Clostridia, Lachnospiraceae ↑ Fusobacterium, Porphyromonas	CRC cases had strong microbe-metabolite correlations that were predominated by Proteobacteria and Actinobacteria.	Selection: **** Comparability: * Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Thomas, 2016 Case control	(CRC) Colorectal cancer (H) Healthy controls	tissue samples	Gut microbiome	16S rRNA gene sequencing (V4-V5 region)	(CRC) ↑ Bacteroides and Dorea ; Parcubacteria) ;Bacteroides, Phascolarctobacterium, Parabacteroides, Desulfovibrio, and Odoribacter (H) ↑ Planctomycetes ; Pseudomonas, Escherichia, Acinetobacter, Lactobacillus, and Bacillus	Total number of OTUs and the Shannon and Simpson higher in CRC patients	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Amitay, 2017 Case control	(CRC) Colorectal cancer	Stool	Gut microbiome	16S rRNA gene sequencing (V1-V2 region)	(CRC) ↑ Fusobacterium nucleatum	Fusobacterium is a passenger that multiplies in the more favorable conditions caused by the malignant tumor rather than a causal factor in colorectal cancer development	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Flemer, 2017 Case control	(CRC) Colorectal cancer	faecal and mucosal samples	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(CRC) ↑ Bacteroidetes Cluster 2, Firmicutes Cluster 2, Pathogen Cluster and Prevotella Cluster ↓ Bacteroidetes Cluster 1 and Firmicutes Cluster 1	CRC-associated microbiota profiles differ from those in healthy subjects	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Gao, 2017 Case control	(CRC) Colorectal cancer	Tissue samples including tumor and adjacent normal mucosal tissue	Gut microbiome	16S rRNA gene sequencing (V4 region)	(CRC) ↑ Fusobacteria ↓ Firmicutes and Actinobacteria	specific mucosa-associated microbiota signature and function are significantly changed in the gut ofCRC patients	Selection: ** Comparability: ** Exposure: * Total NOS score: 5

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Xu, 2017 **Case control**	(CRC) Carcinoma (A) Adenoma (H) Healthy controls	Biopsy on the mucosa	Gut microbiome	16S rRNA gene sequencing	(CRC) ↑ Firmicutes and Fusobacteria ↓ Proteobacteria	Chao1 Richness Index for normal, adenoma, and cancer group were 66.4±32.6, 61.9±29.3, and 87.8±37.5, respectively, p=0.0006 unweighted UniFrac distance: ADONIS, normal-adenoma: R2 =0.015, p=0.05; normal-cancer: R2 enoma-cancer: R2 =0.051, p=0.01; ad=0.059, p=0.01	Selection: *** Comparability: Exposure: ** Total NOS score: 4
Yu, 2017 **Case control**	(CRC) Colorectal cancer	Stool	Gut microbiome	Metagenomic sequencing	(CRC) ↑ Fusobacterium nucleatum and Peptostreptococcus stomatis, Parvimonas micra and Solobacterium moorei	consistent faecal microbial changes in CRC across four cohorts, identification of novel bacterial candidates that may be involved in the development and progression of CRC,	Selection: **** Comparability: Exposure: ** Total NOS score: 6
Allali, 2018 **Case control**	(CRC) Colorectal cancer	Stool	Gut microbiome	16S rRNA gene sequencing (V1-V2 region)	(CRC) ↑ Porphyromonas, Clostridium, Ruminococcus, Selenomonas, and Fusobacterium	Phylogenetic diversity: a moderate but statistically significant effect was observed when we compared control and CRC groups in the second age category (ages 50–79)	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Burns, 2018 **Case control**	(CRC) Colorectal cancer	tumors and matched normal tissues	Gut microbiome	16S rRNA gene sequencing (V5-V6 region)	Changes in abundances of 31 microbial taxa including several taxa within the phylum Bacteroidetes by tumor stage	CRC microbiomes are correlated with tumor mutational profiles, pointing towards possible mechanisms of molecular interaction	Selection: **** Comparability: Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Dai, 2018 Case control	(CRC) Colorectal cancer	Stool	Gut microbiome	Shotgun Metagenomic sequencing	(CRC) ↑ Bacteroides fragilis, Fusobacterium nucleatum, Porphyromonas asaccharolytica, Parvimonas micra, Prevotella intermedia, Alistipes finegoldii,and Thermanaerovibrio acidaminovorans	Identification of a group of bacteria that is consistently associated with CRC and shows potential in the diagnosis of CRC across multiple populations despite technical and biological variations.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Hale, 2018 Case control	(CRC) Colorectal cancer (MMR) Mismatch repair (dMMR) deficient MMR or (pMMR) proficient MMR	Tumor and normal-adjacent tissue	Gut microbiome	16S rRNA gene sequencing (V3-V5 region)	(dMMR CRC) ↑ Bacteroides fragilis and sulfidogenic Fusobacterium nucleatum (pMMR CRC) ↓ B. fragilis	Tumor MMR status strongly predicted microbial community variance and was associated with distinct microbial, metabolic, and interaction profiles.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Loke, 2018 Case control	(CRC) colorectal cancer (H) Healthy controls	paired tumor and unaffected (normal) surgical biopsy tissues	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(H) ↑ 24 OTUs including Alistipes, Oscillibacter, Bacteroides, Pseudoflavonifracor and Succinivibrio (top five)	local microbiome dysbiosis may contribute to functional changes at the cancer sites	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Mori, 2018 Case control	(HP) hyperplastic polyps, (LRA) low-risk adenomas, (HRA) high-risk adenomas (ADK) adenocarcinomas	Stool	Gut microbiome	16S rRNA gene sequencing (V4 region)	(HP) ↑ Firmicutes and Actinobacteria, Lachnospiraceae (ADK) ↑ Sutterella and Escherichia/ Shigella	identification of specific microbial biomarkers associated with each carcinogenic stage	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Richard, 2018 Case control	(CAC) Colitis associated cancer, (SC) Sporadic cancer, (H) Healthy subjects	mucosa on the tumor site and normal mucosa near the tumor	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(CAC) ↑ Enterobacteriacae family and Sphingomonas genus (CAC) ↓ Fusobacterium and Ruminococcus genus	Chao1: higher in H compared to CAC and normal mucosa near the tumor in CAC	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Shah, 2018 Case control	(CRC) tumor biopsies	tumor biopsy and tumor-adjacent biopsy or fecal samples	Gut microbiome	16S rRNA gene sequencing	(CRC) ↑ Fusobacterium, Parvimonas and Streptococcus (CRC) ↓ Faecalibacterium and Ruminococcaceae	With the exception of these limited taxa, the majority of findings from individual studies were not confirmed by other 16S rRNA gene-based datasets.	Selection: *** Comparability: * Exposure: ** Total NOS score: 6
Zhang, 2018 Case control	(CRC) Colorectal cancer (A-CRA) Advanced colorectal adenoma	Stool	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(CRC) ↑ 24 species ↓ 10 species	gut microbiome and inflammation may gradually form a microenvironment to promote the development of CRC	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Saito, 2019 Case control	(CRC) Advanced colorectal carcinoma (CRA) Colorectal adenoma	Colonoscopy aspirates	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(CRA/CRC) ↑ Fusobacterium, Parvimonas, and Atopobium (Controls) ↑ Lachnobacterium, Salmonella, and Moraxellaceae	Two distinct aggregations were observed in the data ofUnweighted UniFrac PCoA.	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Yachida, 2019 Cohort	(CRC) Colorectal cancer (IC) Intramucosal carcinomas (PA) Polypoid adenomas	Stool	Gut microbiome	Shotgun Metagenomic sequencing	(CRC) ↑ Fusobacterium nucleatum spp. (IC/PA) ↑ Atopobium parvulum and Actinomyces odontolyticus	observation of dynamic shifts in microbial composition, gene abundance in gut microbiota and metabolites during multistep CRC progression	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Ahn, 2013 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	16S rRNA gene analysis of the V3-V4 region (454 pyrosequencing) and qPCR	(CRC) ↓ Firmicutes (most prominent for the class Clostridia; family Lachnospiraceae; genus Coprococcus) ↑ genera Fusobacterium, Atopobium and Porphyromonas	Shannon: lower in cases Eveness: NS	Selection: ** Comparability:** Exposure: ** Total NOS score: 6
Amiot, 2015 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	real time qPCR of 16S rRNA	(CRC) ↑ genera Lactobacillus, Leuconostoc, Pediococcus, Blautia coccoides ↑ species E. coli, Faecalibacterium prauznitzii ↑ genera Bacteroides, Prevotella ↓ species Clostridium	Not mentioned	Selection: * Comparability: ** Exposure: *** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Chen, 2012 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	PCR amplification of the V1-V3 region of bacterial 16S rDNA was performed using universal primers incorporating the FLX Titanium adapters and a sample barcode sequence.	(CRC) ↑ families Erysipelotrichaceae, Prevotellaceae, Coriobacteriaceae, Peptostreptococcaceae ↑ genera Peptostreptococcus and Anaerotruncus	Shannon: NSChao index: NSPCA analysis based on unweighted UniFrac: Significant	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Feng, 2015 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	metagenomic shotgun sequencing (Illumina)	(CRC) ↑ Bacteroides, Alistipes (putredenis), Esherichia, Parvimonas, Bilophila (wadsworthia) and Fusobacterium ↓ Genera Ruminococcus, Bifidobacterium and Streptococcus	Shannon: NS Richness: higher in cases (Significant)	Selection: *** Comparability: ** Exposure: *** Total NOS score: 8
Fukugaiti, 2015 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	real time qPCR of 16S rRNA	(CRC) ↑ Fusobacterium nucleatum and C. difficile	Not mentioned	Selection: *** Comparability: * Exposure: *** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Kanazawa, 1996 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	culture	(CRC) ↑ Clostridium lecithinase negative, and Lactobacillus ↓ yeast	The total amount of bacteria is more abundant in the high risk group	Selection: *** Comparability: * Exposure: *** Total NOS score: 7
Kasai, 2016 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	T-RFLP and 16S RNA gene analysis of the V3-V4region (Illumina)	(CRC) ↑ Fusobacteria ↑ genera Actinomyces, Atopobium, Fusobacterium and Heamophilus ↓ genus Slackia ↑ species Actinomyces odontolyticus, Bacteroides fragilis, Clotridium nexile, Fusobacterium varium, Heamophilus parainfluenzae, Prevotella stercorea, Streptococcus gordonii, and Veillonella dispar ↓ species Eubacterium coprostanoligens	Initial T-RFLP analysis did not reveal any statistically significant differences between control and cancer subjects. NGS found that the gut microbiota differs between control and cancer subjects.	Selection: *** Comparability: ** Exposure: *** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Mastromarino, 1978 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	culture	No significant findings	No clearly discernible statistically significant relationships between the fecal profile of the bacterial flora and colon cancer were evident	Selection: **** Comparability: * Exposure: *** Total NOS score: 8
Mira-Pascual, 2014 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	16S rRNA gene analysis (454 pyrosequencing) and real-time qPCR of specific bacteria	(CRC) ↑ Methanobacteriales and Methanobrevibacterium	PCoA based on UniFrac: NS	Selection: *** Comparability: * Exposure: ** Total NOS score: 6
Ohigashi, 2013 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	real time-qPCR	(CRC) ↓ Clostridium coccoides group, Clostridium leptum subgroup, Bifidobacterium, Atopobium cluster, Enterobacteriaceae and Staphylococcus	Total bacterial counts in the CRC group were significantly lower than those in the healthy group.	Selection: ** Comparability: ** Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Vogtmann, 2016 Case control (same samples as Ahn 2013)	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	whole-genome shotgun sequencing (illumine)	(CRC) ↑ Fusobacteria, Fusobacterium and Porphyromonas	Shannon: NS Evenness and richness: NS	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7
Weir, 2013 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	16S rRNA gene analysis of the V4 region (454 pyrosequencing)	(CRC) The composition and relative abundance of the major phyla were similar ↑ Acidaminobacter, Phascolarctobacterium, Akkermansia muciniphila, citrobacter farmeri ↓ Bacteroides finegoldii, Bacteroides intestinalis, Bacteroides capillosus, Prevotella copri, Prevotella oris, Ruminococcus obeum, Ruminococcus albus, Dorea formicigenerans, Lachnobacterium bovis, Lachnospira pectinoschiza, Pseudobutyrivibrio ruminis, Dialister invisus, Dialister pneumosintes, Megamonas hypermegale	Shannon: NS Simpson: NS Evenness: NS unweighted Jaccard and Yue and Clayton's index (beta diversity): NS	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Wu, 2013 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	16S rRNA gene analysis of the V3 region (454 pyrosequencing) and RT-qPCR	(CRC) ↑ Fusobacteria ↑ Families Eubacteriaceae, clostridiales Family XI Incertae sedis and Staphylococcaceae, ↑ Campylobacteraceae, Porphyromonadaceae, Fusobacteriaceae ↑ genera Fusobacterium, Campylobacter, Kingella, Leptotrichia, Solobacterium, Holdemania, Parvimonas, Anaerococcus, Oscillibacter, Fastidiosipila, Peptostreptococcus, Eubacterium, Gemella, Porphyromonas, Parabacteroides, Odoribacter ↑ 17 OTUs belonging to the genus Bacteroides and 5 OTUs that were closely related to Ruminococcus ↓ 24OTUs, most related to Faecalibacterium prausnitzii and Roseburia	Shannon: NS Chao 1 index (richness): NS PCoA based on the unweighted Unifrac: Significant	Selection: ** Comparability: ** Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Yu, 2015 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	whole genome shotgun sequencing (Illumina)	(CRC) ↑ phylum Fusobacteria ↑ Fusobacterium ↓ genera Eubacterium ↑ B. fragilis, Parvimonas micra, Solobacterium moorei, F. nucleatum, Peptostreptococcus stomatis ↓ Eubacterium ventriosum	Shannon: NS Richness: lower in cases (However, these differences exhibited p>0,5 after correcting for colonoscopy)	Selection: ** Comparability: * Exposure: *** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Zackular, 2014 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	16S rRNA gene analyis of V4 region (Illumina)	(CRC) ↑ OTUs associated with Fusobacterium, Porphyromonas, Enterobacteriaceae ↓ OTUs associated with Bacteroides, Porphyromonadaceae, Lachnospiraceae	Relative abundance data from the human gut microbiome differentiates healthy colons from carcinomas	Selection: **** Comparability: ** Exposure: *** Total NOS score: 9
Zeller, 2014 Case control	(CRC) Colorectal cancer (H) Healthy controls	Fecal samples	Gut microbiome	whole genome shotgun sequencing (Illumina)	(CRC) ↑ Fusobacteria, Proteobacteria and Bacteroidetes ↓ Actinobacteria and Firmicutes ↓ genera Eubacterium, Ruminococcus, Bifidobacterium, Campylobacter, Acinetobacter ↑ genera Fusobacterium, Pseudoflavonifractor, Peptostreptococcus, Leptotrichia, Porphyromonas, Desulfovibrio, Parvimonas, Bilophila	Shannon: NS Evenness: NS PCoA: NS	Selection: **** Comparability: ** Exposure: *** Total NOS score: 9

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Alexander, 2016 Case control	(CRC) Colorectal cancer	Sampling of tumors is done without the use of bowel preparation and acquired samples were immediately transferred to a freezer at -80°C.	Gut microbiome	16S rRNA gene analysis (454 pyrosequencing), qPCR (fusobacterium nucleatum)	(CRC) ↑ Fusobacteria and eProteobacteria ↑ Fusobacterium nucleatum	Shannon: NS PCA: NS	Selection: ** Comparability: * Exposure: *** Total NOS score: 6
Allali, 2015 Prospective nested case control samples	(CRC) Colorectal cancer	Samples were collected in sterile Eppendorf tubes, immediately snap-frozen in isopentane and kept at -80°C until analysis.	Gut microbiome	16S rRNA gene analysis of the V1-V2 region (454 pyrosequencing)	(CRC) ↑ genus Eikenella	Phylogenetic Diversity: NS Species Richness: NS weighted and unweighted Unifrac PCA: NS	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7
Allali, 2015 Case control (Spain)	(CRC) Colorectal cancer	Samples were collected in sterile Eppendorf tubes, immediately snap-frozen in isopentane and kept at -80°C until analysis.	Gut microbiome	16S rRNA gene analysis of the V1-V2 region (454 pyrosequencing)	(CRC) ↑ relative abundance of phylum Fusobacteria ↑ genera Bulleidia, Gemella, Streptococcus, Parvimonas, Fusobacterium, and Campylobacter ↓ genera Lachnospira, Faecalibacterium, Ruminococcus and Blautia	Phylogenetic Diversity: NS Species Richness: NS weighted and unweighted Unifrac PCA: NS	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Burns, 2015 Case control samples	(CRC) Colorectal cancer	Tissue pairs were resected concurrently, rinsed with sterile water, flash frozen in liquid nitrogen and characterized by staff pathologists. Total DNA was isolated from the flash-frozen tissue samples and their associated microbiomes by adapting an established nucleic acid extraction protocol.	Gut microbiome	16S rRNA gene analysis of the V5-V6 region (Illumina) and qPCR (Fusobacterium nucleatum, Profidencia)	(CRC) ↓ families Lachnospiraceae, Ruminococcaceae, Rikenellaceae ↓ genus Bacteroides ↓ species Faecalibacterium prausnitzii, Bacteroides uniformis ↑ genera Fusobacterium and Provindencia	Shannon: higher in cases	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7
Chen, 2012 Case control samples	(CRC) Colorectal cancer	After a standard bowel cleansing, tissue samples were collected during surgery before surgery. All samples were frozen and stored at -80°C until further use.	Gut microbiome	16S rRNA gene analysis of V1-V3 region (454 pyrosequencing)	(CRC) ↑ Bacilli (main component Lactobacillales) ↓ family Ruminococcaceae (in pa10t) ↓ genera Faecalibacterium, Phascolarctobacterium, Ochrobactrum, Paraprevotella, Parabacteroides, Methylobacterium, Acidocella (in pa10t)	Shannon: higher in controls (10cm from the tumor) Chao index: NS unweighted Unifrac PCA analysis: NS	Selection: *** Comparability:** Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Chen, 2012 Case control	(CRC) Colorectal cancer (H) Healthy controls	Swab samples were collected from each subject prior to bowel cleansing. All samples were frozen and stored at -80°C until further use.	Gut microbiome	16S rRNA gene analysis of V1-V3 region (454 pyrosequencing)	(CRC) ↑ families Porphyromonadaceae, Fusobacteriaceae and peptostreptococcaceae ↓ Bifidobacteriaceae and Alcaligenaceae ↑ genera Gemella, Catonella,Porphyromonas, Filifactor, Fusobacterium, Peptostreptococcus and Mogibacterium ↓ genera Anaerostipes, Catenibacterium, Faecalibacterium, Blautia, Gardnerella and Bifidobacterium	Shannon index: NS Chao index: NS unweighted Unifrac PCA analysis: Significant	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Flemer, 2016 Case control	(CRC) Colorectal cancer (H) Healthy controls	colon tissue samples	Gut microbiome	real-time quantitative PCR of total DNA and 16S rRNA gene analysis of the V3-V4 region (Illumina)	No significant findings were reported	Unweighted UniFrac PCA analysis: Significant	Selection: *** Comparability: * Exposure: *** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Gao, 2015 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	After a standard bowel cleansing, colorectal tissue samples were collected during surgery/colonoscopy. All were placed in liquid nitrogen immediately and transported to the laboratory within 30 min of collection.	Gut microbiome	16S rRNA gene analysis of the V3 region (454 pyrosequencing)	(CRC) ↑ phylum Fusobacteria ↑ genera Peptostreptococcus, Fusobacterium ↓ class Flavobacteria, Sphingobacteria, Alphaproteobacteria, Gammaproteobacteria ↓ genera Epilithonimonas, Flavobacterium, Pedobacter, Sphingobacterium, Caulobacter, Brevundimonas, Sphingomonas, Acidovorax, Janthinobacterium, Buttiauxella, Rachnella, Acinetobacter, Enhydrobacter, Psychrobacter, Pseudomonas, Nevskia, Stenotrophomonas, Nesterenkonia, Propionibacterium	Chao and ACE indices: NS Shannon: higher in controls	Selection: * Comparability: ** Exposure: *** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Kohoutova, 2014 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	The usual bowel preparation was either polyethylene glycol or sodium phosphate solution. Sterile biopsy forceps (Olympus) were used for every single biopsy. Each biopsy specimen of the colonic mucosa was immediately inserted into a transport liver-enriched broth.	Gut microbiome	culture and PCR	(CRC) ↑ Escherichia coli strains ↑ E. coli phylogroup D	Not mentioned	Selection: **** Comparability:- Exposure:*** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Marchesi, 2011 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	tissue samples were collected during surgery	Gut microbiome	DGGE fingerprinting and 16S rRNA gene analysis of V1-V3 region (454 pyrosequencing)	(CRC) ↑ genera Slackia, Roseburia and Collinsella, Faecalibacterium, Fusobacterium ↓ members of Enterobacteriaceae, such as Citrobacter, Shigella, Cronobacter, Kluyvera, Serratia and Salmonella	Shannon: NS	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7
Mira-Pascual, 2014 **Case control**	(CRC) Colorectal cancer (H) Healthy controls	bowel cleaning: (a) low-fiber diet for 72h before the colonoscopy; (b) full liquid diet during the day before the colonoscopy; (c) ingestion of two tablets of bisacodyl-5mg, the day before the colonoscopy around 7p.m. ; (d) drinking of 1l polyethylene glycol+ ascorbic acid at 8p.m. on the day before and repeated at the same dose and quantity at 6a.m. the same day of the	Gut microbiome	16S rRNA gene analysis (454 pyrosequencing), together with real-time quantitative PCR of specific bacteria	(CRC) ↑ Enterobacteriaceae ↓ Staphylococcus ↑ Blautia (coccoides), Prevotella	Alpha-diversity (rarefaction curves): higher in cases (Significant) PCoA: significant	Selection: *** Comparability: * Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
		colonoscopy. Biopsy samples were taken from the normal mucosa (rectum) in controls and from the patients.					
McCoy, 2013 Case control samples	(CRC) Colorectal cancer	samples were obtained from UNC Tissue Procurement Facility to confirm previously reported studies	Gut microbiome	16S rRNA gene analysis of V1-V3 region (454 Pyrosequencing) and qPCR	(CRC) ↑ Fusobacterium	Shannon diversity and richness: higher in cases (NM)	Selection: ** Comparability: ** Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Warren, 2013 Case control samples	(CRC) Colorectal cancer	total RNA was isolated from frozen surgical sections of CRC and matched unaffected control tissue from 65 subjects.	Gut microbiome	whole genome shotgun sequencing (Illumina)	(CRC) ↓ Ruminococcus, Parabacteroides, Pseudoflavonifractor and Holdemania ↓ family Ruminococcaceae ↑ Fusobacterium, Campylobacter and Leptotrichia	Principal Component Analysis (beta diversity): NS	Selection: ** Comparability: ** Exposure: *** Total NOS score: 7
Youssef, 2018 Case control	(GIT) gastrointestinal tract (RN) rectal neoplasm (CN) colon neoplasm (GN) stomach neoplasm	Stool	Gut microbiome	16S rRNA gene sequencing (V2, V4, V8, V3, V6–7, and V9 region)	(GN) ↑ Enterobacteriaceae (GIT) ↑Ruminococcus, Subdoligranulum ↓ Lachnoclostridium and Oscillibacter (RN) ↓ Bifidobacteriaceae (CN) ↓ Lactobacillaceae	Alpha diversity (Shannon index) and observed richness did not have significant differences between the groups (based on location of neoplasm). Beta diversity was not significantly different between the controls and the patients, either for the treated or the nontreated groups.	Selection: *** Comparability: * Exposure: ** Total NOS score: 6
Goedert, 2015 Case control	(BC) Breast cancer	Fecal samples	Gut microbiome	16S rRNA gene sequencing of the V3-V4 region (Illumina)	(BC) ↑ Clostridiaceae, Faecalibacterium and Ruminococcaceae ↓ Dorea and Lachnospiraceae	Shannon: lower in BC cases Chao1 index: lower in cases Beta-diversity: higher in cases (unweighted UniFrac)	Selection: **** Comparability: ** Exposure: *** Total NOS score: 9
Goedert, 2018 Case control	(BC) Breast cancer	Stool	Gut microbiome	16S rRNA gene sequencing (V4 region)	↓ alpha diversity	Cases had significantly reduced alpha diversity and altered composition of both their IgA-positive and IgA-negative faecal microbiota	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Zhu, 2018 Case control	(BC) Breast cancer	Stool	Gut microbiome	Shotgun Metagenomic sequencing	(postmenopausal BC) ↑ 38 species including Escherichia coli, Klebsiella sp_1_1_55, Prevotella amnii, Enterococcus gallinarum, Actinomyces sp. HPA0247, Shewanella putrefaciens,and Erwinia amylovora ↓ 7 species including Eubacterium eligens and Lactobacillus vaginalis	Relative species abundance in gut microbiota did not differ significantly between premenopausal breast cancer patients and premenopausal controls	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Alanee, 2019 Cohort	(PC) Prostate cancer	Rectal swab and first voided urine after a prostate massage	Gut and urinary microbiome	16S rRNA gene sequencing (V3-V5 region)	(PC) ↓ faecalibacterium, lactobaccili,and Actinetobacter ↑ Veillonella, Streptococcus, and Bacteroides (H) ↑ clostridium XVIII & IV, lachnospira, acetanaerobacterium, faecalibacterium	The majority of patients had similar bacterial communities within their urinary sample profile. Analysis of the bacterial taxonomies of the fecal samples did not reveal any clustering in concordance with benign or malignant prostate biopsies.	Selection: **** Comparability: Exposure: ** Total NOS score: 6
Golombos, 2018 Case control	(PC) Prostate Cancer (H) Controls	Stool	Gut microbiome	Shotgun Metagenomic sequencing	(PC) ↑ Bacteriodes massiliensis (H) ↑ Faecalibacterium prausnitzii and Eubacterium rectalie	Significant differences exist in the gut microbial composition of men with prostate cancer compared to benign controls	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Liss, 2018 **Case control**	(PC) Prostate Cancer (H) Controls	Rectal swab	Gut microbiome	16S rRNA gene sequencing (V1-V2 region)	(PC) ↑ Bacteroides and Streptococcus species	Mean beta-diversity distances (BrayCurtis/Unweighted UniFrac): Patients with prostate cancer are significantly more similar in taxonomic composition compared with the non-cancer group	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Qi, 2019 **Case control**	(GC) Gastric cancer	Stool	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(GC) ↑ Lactobacillus, Escherichia, and Klebsiella ↓ butyrate-producing bacteria	Intestinal microbiota in gastric cancer patients was characterized by increased species richness (Sobs index)	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Ren, 2019 **Case control**	(HCC) hepatocellular carcinoma	Stool	Gut microbiome	16S rRNA gene sequencing	(HCC) ↑ Actinobacteria (vs cirrhosis), and 13 genera including Gemmiger and Parabacteroides (vs controls)	Faecal microbial diversity was decreased from healthy controls to cirrhosis, but it was increased from cirrhosis to early HCC with cirrhosis.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Feng, 2019 Case control	(TC) thyroid carcinoma (H) healthy controls	Stool	Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(BC) ↑ 19 genera ↓ 8 genera	Eight metabolites combined with 5 genera were more effective in distinguishing TC patients from H	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Gong, 2013 Case control	(LSCC) Laryngeal squamous cell carcinoma, (H) Controls with vocal cord polyps.	Tissue samples of the larynx	Laryngeal microbiome	DNA extraction and 16S rRNA gene analysis of the V1-V3 region.	All samples: mainly Firmicutes, Fusobacteria, Bacteroidetes, Proteobacteria, and Actinobacteria (in that order of frequency) and at genera level Streptococcus, Fusobacterium, Prevotella, Neisseria and Gemella. (LSCC) ↑ Fusobacterium, Prevotella and Gemella ↓ Streptoccocus and Rothia. Prevotella and Solobacterium were significantly more prevalent in T3-T4 tumors than T1-T2.	Significantly different microbiomes between (LSCC) and controls.	Selection: **** Comparability: Exposure: * Total NOS score: 5
Shin, 2017 Case control	(HNSCC) Head and neck squamous cell carcinoma	Tissues from the oral cavity, larynx-pharynx, and lymph nodes	Oral microbiome	16S rRNA gene sequencing (V4 region)	(HNSCC) ↑ Fusobacteria and Proteobacteria ↓ Firmicutes, Actinobacteria and Streptococcus	Alpha and beta diversity analyses revealed that normal tissues had the greatest richness in community diversity, while the metastatic populations were most closely related to one another.	Selection: **** Comparability: * Exposure: ** Total NOS score: 7
Wang, 2017 Case control	(HNSCC) Head and neck squamous cell carcinoma	Tumor resection specimens	Oral microbiome	16S rRNA gene sequencing (V1-V4 region)	(HNSCC) ↑ Parvimonas ↓ Actinomyces	Shannon, phylogenetic divesity: no sgnifican difference Weighted, unweighted Unifrac: lower in tumor	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Hayes, 2018 Nested case-control	(HNSCC) Head and neck squamous cell carcinoma	mouthwash samples	Oral microbiome	16S rRNA gene sequencing (V3-V4 region)	(HNSCC) ↓ Corynebacterium, Kingella	oral microbiome may be associated with subsequent risk ofHNSCC,with the strongest links for larynxcancer and those with a history of tobacco use	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Hu, 2016 Case control	(OSCC) oral squamous cell carcinoma (OLK) oral leukoplakia (H) healthy controls	Nonstimulated saliva	Oral microbiome	16S rRNA gene sequencing	(OSCC) ↑ Bacillus (OLK) ↑ Bacteroidetes, TM7, Haemophilus ↓ Firmicutes (H) ↑ Streptococcus and Abiotrophia	Chao index: higher richness in OLK compared to HC Shannon and Simpson: higher diversity in OSCC compared to HC	Selection: * Comparability: Exposure: ** Total NOS score: 3
Banerjee, 2017 Case control	(OCSCC/OPSCC) oropharyngeal squamous cell carcinomas	Carcinomas taken from tongue, base of tongue, tonsil, floor of mouth, cheek and predominantly oropharynx	Oral microbiome	Pan-pathogen array technology (PathoChip) coupled with next-generation sequencing	(OCSCC) ↑ Proteobacterias Eshcherichia, Brevundimonas, Comamonas, Alcaligenes, Caulobacter, Cardiobacterium, Plesiomonas, Serratia, Edwardsiella, Haemophilus, Frateuria along with Actinobacteria Rothia and Bacteroidetes Peptoniphilus	The screening of OCSCC samples as well as matched and non-matched controls have identified distinct viral and other microbial signature patterns associated with OCSCC	Selection: **** Comparability: Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Lee, 2017 Case control	(OC) oral cancer	saliva samples	Oral microbiome	16S rRNA gene sequencing (V4 region)	(OC) ↑ Bacillus, Enterococcus, Parvimonas, Peptostreptococcus, and Slackia	the oral microbiome compositions revealed significant differences between epithelial precursor lesion and cancer patients	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Mok, 2017 Case control	(C) Cancer (OPMD) Potentially malignant disorders (H) Normal	Swab samples	Oral microbiome	16S rRNA gene sequencing (V6-V9 region)	(C/OPMD) ↑ Neisseria, Gemella and Granulicatella (H) ↑ Streptococcus and Veillonella	significant difference between the normal and the cancer associated oral microbiota but not between the OPMD and the other two groups	Selection: **** Comparability: Exposure: ** Total NOS score: 6
Wolf, 2017 Case control	(OSCC) Oral squamous cell carcinomas (H) Healthy	saliva samples	Oral microbiome	16S rRNA gene sequencing	(H) ↑ Bacteroidetes (Prevotella), Proteobacteria (e.g. Haemophilus and Neisseria) and Firmicutes (e.g. Streptococcus and Veilonella) (OSCC) ↑ Actinomyces (Actinobacteria), Schwartzia (Firmicutes), Treponema (Spirochaetes) and Selenomonas (Firmicutes)	differences in microbial abundance and diversity might inform disease status in OSCC patients	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Zhao, 2017 Case control	(OSCC) Oral squamous cell carcinomas	Lesion samples and anatomically matched normal samples	Oral microbiome	16S rRNA gene sequencing (V4-V5 region)	(OCSCC) ↑ Fusobacterium, Dialister, Peptostreptococcus, Filifactor, Peptococcus, Catonella and Parvimonas	Significantly greater bacterial diversity was observed in the cancer samples than in the normal samples	Selection: **** Comparability: Exposure: ** Total NOS score: 6
Hsiao, 2018 Case control	(OSCC) Oral squamous cell carcinomas (H) Healthy	saliva samples	Oral microbiome	16S rRNA gene sequencing (V3-V5 region)	(OSCC) ↑ Prevotella tannerae, Fusobacterium nucleatum, and Prevotella intermedia	positive association between periodontopathogenic bacteria and OSCC risk and this relationship may be influenced by lifestyle and genetic factors	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Lim, 2018 Case control	(OCC) Oral cavity cancer and (OPC) oropharyngeal cancers	Oral rinse samples	Oral microbiome	16S rRNA gene sequencing (V6-V8 region)	(OCC/OPC) ↑ Actinobacillus, Actinomyces, Aggregatibacter, Capnocytophaga, Fusobacterium, Oribacterium, Rothia, Haemophilus, Leptotrichia, Neisseria, Porphyromonas and Veillonella ↓ Paludibacter, Corynebacterium	the oral microbiome is able to predict the presence of OCC and OPC with sensitivity and specificity of 100 and 90%	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Perera, 2018 Case control	(OSCC) oral squamous cell carcinoma (FEP) fibroepithelial polyp	fresh biopsies	Oral microbiome	16S rRNA gene sequencing (V1-V3 region)	(OSCC) ↑ Capnocytophaga, Pseudomonas, and Atopobium, Campylobacter concisus, Prevotella salivae, Prevotella loeschii, and Fusobacterium oral taxon 204 (FEP) ↑ Lautropia, Staphylococcus, and Propionibacterium, Streptococcus mitis, Streptococcus oral taxon 070, Lautropia mirabilis, and Rothia dentocariosa	OSCC tissues tended to have lower species richness and diversity	Selection: **** Comparability: * Exposure: ** Total NOS score: 7
Yost, 2018 Case control	(OSCC) Oral squamous cell carcinoma	tumor biopsy and tumor-adjacent biopsy	Oral microbiome	Metatranscriptomic analysis	(OSCC) ↑ Fusobacteria	Fusobacteria may be the leading phylogenetic group responsible for the increase in expression of virulence factors in the oral microbiome of OSCC patients	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Börnigen, 2017 Case control	(OSCC) Oral squamous cell carcinomas	Oral rinse samples	oral microbiome	16S rRNA gene sequencing (V4 region)	(OSCC) ↑ Dialister ↓ Scardovia	Alterations following tooth loss, itself a major risk factor for oral cancer, are likely a result of severe ecological disruption due to habitat loss but may also contribute to the development of the disease.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Schmidt, 2014 Case control	(OC) Oral cancer, (HC) Healthy controls. (CIS) Carcinoma in situ, (PRE) Pre-cancer stages	Oral swab	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V4 region.	All samples: mainly Firmicutes, Bacteroidetes, Proteobacteria, Fusobacteria and Actinobacteria. (OC2) ↑ Fusobacteria and Bacteroidetes, ↓ Firmicutes and Actinobacteria, ↓ Streptoccocus and Rothia, ↑ Fusobacterium and Prevotella. (PRE) ↑ Bacteroidetes and ↓ Streptoccocus.	Significant differences between the groups and within the same patient (oral lesions vs. anatomically matched samples).	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Berkovitz, 2016 Case control	(OSCC) Oral squamous cell carcinoma. (H) Healthy controls	Oral swab	Oral microbiome	Yeast colonization first by agar plates and incubation, then by macro- and microscopic morphology, catalase test and CHROMagar Candida plates, and finally MALDI-TOF analysis.	All samples: predominant fungal genus was Candida. (OSCC) ↑ frequency of oral yeast colonization, more yeast cells and ↑ fungal burden.	(OSCC): ↑ diversity of fungi. No significant differences in lipase and protease activity.	Selection: **** Comparability: Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Henrich, 2014 Case control	(FAC) Fanconi Anaemia and oral squamous cell cancer. (FAB) Fanconi Anaemia and benign oral lesion (H) Healthy controls.	Oral swab	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V1-V2 region, as well as a Candida-specific qPCR.	All (FAC) samples: mainly Bacteroidetes, Firmicutes, Proteobacteria and Tenericutes. All (FAB) and (H): mainly Actinobacteria, Bacteriodetes, Firmicutes and Proteobacteria. (FAC) ↑ Mycoplasmataceae (M. salivarium), Pseudomonadaceae, P. salivae and Prevotella spp., ↓ Streptococcus, Rothia mucilaginosa. All (FAC): Candida positive, all other samples Candida negative.	(FAC) ↓ diversity, microbiota was getting less diverse the more the samples are moving into a tumourous state.	Selection: * Comparability: Exposure: ** Total NOS score: 3
Homann, 2000 Case control	(T) Malignant tumor of the oral cavity. (A) Alcoholics. (DE) Dental examination or treatment. (UN) Unemployed volunteers. (H) Healthy volunteers.	Saliva	Oral microbiome	Head space gas chromatography (GC) + several specific and non-specific agar media.	High acetaldehyde producers (especially smokers and heavy drinkers): ↑ count of aerobes like Streptococcus salivarius, Streptococcus viridans, Corynebacterium sp., Stomatococcus sp., and yeasts (higher concentration + more frequently), also ↑ count of anaerobes. No bacterial species was significantly more frequent among the low producers.	Significant differences between high and low producers. Smoking and heavy alcohol intake are strong predictors of microbial acetaldehyde production.	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Marttila, 2013 Case control	(OSCC) Oral squamous cell carcinoma. (OLD) Oral lichenoid disease. (H) Healthy controls.	Microbial sample of the oral mucosa	Oral microbiome	Gas chromatography + several agar media.	(OSCC) ↑ numbers of microbes, especially aerobic bacteria. Lesions site of the (OSCC): ↑ amount of anaerobic bacteria. Lesion sites of (OSCC) and (OLD): ↑ frequency and density of candida colonization + significantly more frequently mutagenic amounts of acetaldehyde. Cultures producing mutagenic concentrations of acetaldehyde: ↑ Candida colonization.	The majority (68%) of the cultures from all groups: mutagenic levels of acetaldehyde. No correlation between acetaldehyde levels and total amount of cultivable microbes in any patient group or sample site. Smokers: ↑ mean acetaldehyde production. Non-smokers: ↑ diversity.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Guerrero-Preston, 2016 Cohort study (1 year follow-up of patients and controls nested in cohort study)	(HNSCC) Head and neck squamous cell carcinoma, (HC) Healthy controls. (OPSCC) Oropharyngeal squamous cell carcinoma. (OCSCC) Oral cavity squamous cell carcinoma	Saliva and tumor samples	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V5 region.	All samples: mainly Firmicutes, Bacteroidetes, Protebacteria, Actinobacteria and Fusobacteria (in order of frequency in (H) group). (HNSCC) ↑ Firmicutes, ↓ Bacteroidetes and Proteobacteria, ↑ Streptoccocus and Lactobacillus, ↓ Agreggatibacter, Lautropia, Haemophillus, Neisseria, Prevotella, Gemellaceae and Leptotricha. (OCSCC) ↑ Neisseria and ↓ Citrobacter than (OPSCC).	(HNSCC): ↓ diversity. (OCSCC): ↑ diversity than (OPSCC).	Selection: *** Comparability: * Exposure: ** Total NOS score: 6

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Wang, 2019 Case control	(BALF) fluid bronchoalveolar lavage (LC) Lung cancer	Buccal (saliva) and lower respiratory tract BALF samples	Oral and lung microbiome	16S rRNA gene sequencing (V4 region)	(LC) ↓ Treponema and Filifactor in BALF ↓ Filifactor in saliva	Patients with lung cancer have lower microbial diversity than healthy controls in both saliva and BALF samples.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Lu, 2016 Case control	(LC) liver carcinoma patients with cirrhosis (H) Healthy	Tongue coat	Tongue coat microbiome	16S rRNA gene sequencing (V3-V4 region)	(LC) ↑ Oribacterium and Fusobacterium	Shannon, Simpson and Chao 1 indexes: higher diversity in LC	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Yang, 2019 Case control	(CRC) Colorectal cancer	mouth rinse samples	Oral microbiome	16S rRNA gene sequencing (V4 region)	(CRC) ↑ Treponema denticola and Prevotella intermedia, Actinobacteria, Bifidobacteriaceae, Bacteroidetes, Prevotella denticola and Prevotella sp. oral taxon 300 ↓ Prevotella melaninogenica, Firmicutes, Carnobacteriaceae, Streptococcaceae, Erysipelotrichaceae, Streptococcus, Solobacterium, Streptococcus sp. oral taxon 058 and Solobacterium moorei	multiple oral bacterial taxa were associated with subsequent CRC risk	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Flemer, 2018 Case control	(CRC) colorectal cancer	oral swabs, colonic mucosae and stool	Oral microbiome Gut microbiome	16S rRNA gene sequencing (V3-V4 region)	(CRC) ↓ Lachnospiraceae ↑ Streptococcus and Prevotellas pp., putative oral biofilm forming bacteria	The heterogeneity of CRC may relate to microbiota types that either predispose or provide resistance to the disease, and profiling the oral microbiome may offer an alternative screen for detecting CRC.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Han, 2014 Case control	(CRC) Colorectal cancer. (H) Healthy controls	Tongue coating samples and images.	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V2-V4 region.	Thick (CRC): ↓ OTU's, ↑ Prevotella, Leptotrichia and Actinomyces, ↓ Gemella, compared to thin (CRC) and (H). Thin (CRC): ↓ Veilonella compared to thick (CRC) and (H). General (CRC): ↑ Streptoccocus and ↓ Haemophilius than in (H).	Significant thicker tongue coating (CRC). Different bacteria depending on the thickness of tongue coating.	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Han, 2016 Case control	(CRC) Colorectal cancer. (LC) Lung cancer. (GC) Gastric cancer. (H) Healthy controls	Tongue coating samples and images.	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V2-V4 region.	Cancer group: ↓ Neisseria, Haemophilus, Fusobacterium and Porphyromonas (i.e. Fusobacterium periodonticum, Haemophilus parainfluenzae, Peptostreptococcaceae bacterium, Prevotella aurantiaca, Prevotella salivae and a TM7).	Significant ↑ mirror-like tongues and thicker tongue coating in the cancer group. Significant bacterial changes.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Peters, 2017 Case control	(EAC) esophageal adenocarcinoma and (ESCC) esophageal squamous cell carcinoma	mouthwash samples	Oral microbiome	16S rRNA gene sequencing (V4 region)	(EAC) ↑ Tannerella forsythia, Neisseria , Streptococcus pneumoniae (ESCC) ↑ Porphyromonas gingivalis	specific bacterial pathogens may play a role in esophageal cancer risk, whereas other bacterial types may be associated with reduced risk	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Chen, 2015 Case control	(ESCC) Esophageal squamous cell carcinoma. (DYS) Dysplasia. (H) Healthy controls.	Saliva	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V4 region.	All samples: mainly Bacteroidetes, Firmicutes, Protebacteria, Fusobacteria, and Actinobacteria (in order of frequency). (ESCC): ↑ Prevotella, Streptococcus and Porphyromonas, ↓ Megasphaera, Aggregatibacter, Atopobium, Lautropia, Actinobacillus, Bulleidia, Catonella, Filifactor, Corynebacterium, TG5, Acholeplasma, Moryella, Butyrivibrio, Dialister, Peptococcus, and Cardiobacterim. (ESCC) compared to (DYS) ↓Lautropia, Bulleidia, Catonella, Corynebacterium, Moryella, Peptococcus and Cardiobacterium.	Significant difference in OTU diversity and richness. (ESCC): ↓ microbial diversity. ↓ Streptococcus ~ ↑ TNM stage, ↑ Lactobacillus ~ ↑ TNM stages.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Olson, 2017	(PDAC) pancreatic ductal adenocarcinoma (IPMN) intraductal papillary mucinous neoplasms (H) healthy controls	Saliva samples	Oral microbiome	16S rRNA gene sequencing (V4-V5 region)	(PDAC) ↑ Firmicutes (H) ↑ Proteobacteria	no differences in diversity of the oral microbiota among these groups, there were differences in the mean relative proportions of some taxa	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Fan, 2018 Case control	(PC) pancreatic cancer	oral wash samples	Oral microbiome	16S rRNA gene sequencing (V3-V4 region)	(PC) ↑ Porphyromonas gingivalis and Aggregatibacter actinomycetemcomitans ↓ Fusobacteria, Leptotrichia	The oral microbiota may play a role in the aetiology of pancreatic cancer.	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Fan, 2016 Prospective nested case control	(PAD) Primary pancreatic adenocarcinoma. (H) Matched controls.	Oral mouthwash	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V3-V4 region.	P.gingivalis: associated with ↑ risk of panreatic cancer (OR 1.6) for low relative abundance and high relative abundance, thus showing a dose-response relationship. A. actinomycetemcomitans: associated with ↑ risk of pancreatic cancer (OR 2.20), more in ever-drinkers (OR 3.03) than in never-drinkers (OR 0,47). Fusobacteria and its genus Leptotrichia: associated with ↓ pancreatic cancer risk (OR 0.94).	Carriage of the periodontal pathogens P. gingivalis and A. actinomycetemcomitans and ↓relative abundance of Fusobacteria and its genus Leptotrichia, were associated with subsequent risk of pancreatic cancer, unlikely due to smoking or other confounders.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Farrell, 2012 Case control	(PC) Pancreatic cancer. (H) Matched controls. (CP) Chronic pancreatitis	Saliva	Oral microbiome	DNA extraction and 16S rRNA gene analysis using universal primers + HOMIM array was used for profiling.	All samples: mainly Firmicutes, Proteobacteria, Bacteroidetes and Actinobacteria. (PC1): ↑ 31 species/clusters and ↓ 25. Potential biomarker candidates: Streptococcus, Prevotella, Campylobacter, Granulicatella, Atopobium and Neisseria. (PC2): ↓ Streptococcus and Neisseria, ↑ Granulicatella. (CP): ↓ Streptococcus and ↑ Granulicatella, compared to (PC2).	Significant differences between (PC) and (H). Using Streptococcus and Neisseria as biomarker: 96.4% sensitivity and 82.1% specificity for distinguishing patients with pancreatic cancer from healthy subjects. Using Streptococcus and Granulicatella: 85.7% sensitivity and 55.6% specificity.	Selection: *** Comparability: ** Exposure: ** Total NOS score: 7
Torres, 2015 Case control	(PC) Pancreatic cancer. (OD) Other diseases. (H) Healthy controls	Saliva	Oral microbiome	DNA extraction and 16S rRNA gene analysis using a 'universal' bacterial primer 515F + qPCR.	All samples: mainly Proteobacteria, Actinobacteria, Bacteroidetes, Firmicutes, and Fusobacteria. (PC): ↑ Leptotrichia, ↓ Porphyromonas and Neisseria. No difference in Streptococcus or Granulicatella between (PC) and (H). (PC): ↑ abundance ratio of Neisseria to Porphyromonas.	Significant differences in microbiome between (PC) and (H). No differences among the 3 main groups in beta diversity or alpha diversity. The Leptotrichia to Porphyromonas ratio as a potential diagnostic biomarker for pancreas cancer.	Selection: *** Comparability: Exposure: ** Total NOS score: 5

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Hu, 2015 **Case control**	(GC) Gastric cancer, (H) Healthy controls, (GCtn) Gastric cancer, thin coatings, (GCtk) Gastric cancer, thick tongue coatings	Tongue coating samples and images	Oral microbiome	DNA extraction and 16S rRNA gene analysis of the V2-V4 region.	All samples: mainly Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Fusobacteria, and TM7. (GC): ↑ Actinobacteria and ↓ Proteobacteria, ↓ Fusobacterium, Neisseria, Haemophilus and Porphyromonas. (GCtn): mainly Prevotella, Veillonella, Leptotrichia, Lactococcus, and Streptococcus. (GCtk): mainly Prevotella, Streptococcus, Actinomyces, Veillonella, and Leptotrichia, ↑ Actinomyces and Streptococcus compared to the others.	Significant difference in the thickness of tong coating. (GCtk) significantly ↓ microbial community diversity.	Selection: **** Comparability: ** Exposure: ** Total NOS score: 8
Banerjee, 2017 **Case control**	(OC) Ovarian cancer	Ovarian cancer samples	Ovarian microbiome	Pan-pathogen array technology (PathoChip) coupled with next-generation sequencing	(OC) ↑ Streptococcus, Staphylococcus, Bacillus, Mycoplasma, Chlamydophila, Pediococcus, Chyseobacterium, Fusobacterium, Prevotella, Escherichia, Treponema	A distinct group of viral, bacterial, fungal and parasitic signatures of high significance in ovarian cases	Selection: **** Comparability: Exposure: ** Total NOS score: 6
Gong, 2017 **Case control**	(LC) laryngeal carcinoma (VCP) vocal cord polyps	swab samples	Pharyngeal microbiome	16S rRNA gene sequencing (V1-V2 and incomplete V3 region)	no difference in bacterial communities in the pharynx from patients with LC and VCP	Inverse correlations were found between Streptococcus and other bacterial communities, suggesting that potential antagonism may exist among pharyngeal microbiota	Selection: ** Comparability: * Exposure: ** Total NOS score: 5

RESEARCH ARTICLE	Abbreviations for cases & controls	ANALYSED SPECIMEN	ANALYSED MICROBIOME	MEASUREMENT OF THE MICROBIOME	MICROBAL COMPOSITION ALTERATION	OUTCOME	QUALITY OF THE ARTICLE
Cavarretta, 2017 Case control	(T) Tumor (N) Non tumor	radical prostatectomy-specimens	Prostate microbiome	Massive ultradeep pyrosequencing	(T) ↑ Staphylococcaceae, Staphylococcus ↓ Lactobacillales, Streptococcaceae, Streptococcus	possible pathophysiological correlation between the composition of the local microbial niche and the presence of the tumor itself	Selection: *** Comparability: * Exposure: ** Total NOS score: 6
Salava, 2016 Case control	(M) Melanomas (MN) Melanocytic nevi	Non-invasive swab specimens from melanocytic skin lesions	Skin microbiome	16S rRNA gene sequencing (V1-V3 region)	No significant differences between M and MN	No significant differences in the relative abundances of bacterial genera or bacterial diversity between the patient groups	Selection: ** Comparability: * Exposure: ** Total NOS score: 5
Bučević Popović, 2018 Case control	(BC) Bladder cancer	Urine	Urinary microbiome	16S rRNA gene sequencing (V4 region)	(BC) ↑ Fusobacterium (H) ↑ Veillonella, Streptococcus and Corynebacterium	microbial diversity and overall microbiome composition were not significantly different between groups	Selection: **** Comparability: * Exposure: ** Total NOS score: 7
Bi, 2019 Case control	(BC) Bladder cancer (H) Healthy control	Urine	Urinary microbiome	16S rRNA gene sequencing (V3-V4 region)	(BC) ↑ Actinomyces (HC) ↑ Streptococcus, Bifidobacterium, Lactobacillus, Veillonella	Alpha diversity of the bladder cancer group was higher than that of the control group	Selection: *** Comparability: * Exposure: ** Total NOS score: 6

NS: Non-significant
NM: Not mention

Supplementary Table S5: Overview of all bacteria reported in the studies included in this systematic review and their direction of association with cancer risk

Taxonomic level - Phylum	Cancer and Result	Number of studies		Microbiota body site	Reference
Acidobacteria	Increased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Actinobacteria	Increased in GC	1		Oral	Hu, 2015
Actinobacteria	Decreased in OC	1		Oral	Schmidt, 2014
Actinobacteria	Increased in HCC	1		Gut	Ren, 2019
Actinobacteria	Decreased in HNSCC	1		Oral	Shin, 2017
Actinobacteria	Increased in CRC	1		Oral	Yang, 2019
Actinobacteria	Decreased in CRC	2		Gut	Gao, 2017; Zeller, 2014
Bacteroidetes	Increased in GC	1		Gastric	Yu, 2017
Bacteroidetes	Increased in CRC	1		Gut	Zeller, 2014
Bacteroidetes class1	Decreased in CRC	1		Gut	Flemer, 2017
Bacteroidetes class 2	Increased in CRC	1		Gut	Flemer, 2017
Bacteroidetes	Increased in OC	3	*	Oral	Banerjee, 2017; Hu, 2016; Schmidt, 2014
Bacteroidetes	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Firmicutes	Increased in GC	1		Gastric	Yu, 2017
Firmicutes	Decreased in OC	2	*	Oral	Hu, 2016; Schmidt, 2014
Firmicutes	Increased in CRC	1		Gut	Xu, 2017
Firmicutes	Decreased in CRC	3		Gut & Oral	Gao, 2017; Zeller, 2014; Yang, 2019
Firmicutes class 1	Decreased in CRC	1		Gut	Flemer, 2017
Firmicutes class II	Increased in CRC	1		Gut	Flemer, 2017
Firmicutes	Increased in LC	1		Lung	Lee, 2016
Firmicutes	Decreased in PC	1		Oral	Olson, 2017
Firmicutes	Increased in HNSCC	1		Oral	Guerrero-Preston, 2016
Firmicutes	Decreased in HNSCC	1		Oral	Shin, 2017
Fusobacteria	Increased in GC	1		Gastric	Yu, 2017
Fusobacteria	Increased in ECCA	1		Bile duct	Aviles-Jimenez, 2016

Taxonomic level - Phylum	Cancer and Result	Number of studies		Microbiota body site	Reference
Fusobacteria	Increased in CRC	10	*	Gut	Kasai, 2016; Vogtmann, 2016; Wu, 2013; Yu, 2015; Zeller, 2014; Alexander, 2016; Gao, 2015; McCoy, 2013; Gao 2017; Xu 2017
Fusobacteria	Increased in HNSCC	1		Oral	Shin, 2017
Fusobacteria	Increased in OC	2	*	Oral	Schmidt, 2014; Yost, 2018
Fusobacteria	Associated with low risk of PC or decreased	2	*	Oral	Fan, 2016; Fan, 2018
Planctomycetes	Increased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Proteobacteria	Decreased in CRC	1		Gut	Xu, 2017
Proteobacteria	Increased in CRC	2		Gut	Alexander, 2016; Zeller, 2014
Proteobacteria	Increased in HNSCC	1		Oral	Shin, 2017
Proteobacteria	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Proteobacteria	Decreased in GC	2		Oral & Gastric	Hu, 2015; Yu, 2017
Proteobacteria	Decreased in PC	1		Oral	Olson, 2017
Proteobacteria	Increased in OC	1		Oral	Banerjee, 2017
Spirochaetes	Increased in GC	1		Gastric	Yu, 2017
TM7	Increased in LC	1		Lung	Lee, 2016
TM7	Increased in OC	1		Oral	Hu, 2016

*All studies in same direction

CRC= Colorectal cancer

ECCA= Extrahepatic cholangiocarcinoma

GC= Gastric Cancer

HCC= Hepato cellular carcinoma

HNSCC= Head and neck squamous cell carcinoma

LC= lung cancer

OC= Oral cancer

PC= Pancreatic cancer

Taxonomic level - Class	Cancer and Result	Number of studies	Microbiota body site	
Alphaproteobacteria	Decreased in CRC	1	Gut	Gao, 2015
Bacilli	Increased in CRC	1	Gut	Chen, 2012
Clostridia	Decreased in CRC	1	Gut	Sinha, 2016
Gammaproteobacteria	Decreased in CRC	1	Gut	Gao, 2015
Lactobacilli	Decreased in PRO	1	Gut & Urinary	Alanee, 2019
Sphingobacteria	Decresed in CRC	1	Gut	Gao, 2015

*All studies in same direction

CRC= Colorectal cancer
PRO= Prostate cancer

Taxonomic level - Order	Cancer and Result	Number of studies	Microbiota body site	
Clostridiales	Increased in ESCC	1	Gastric	Nasrollahzadeh, 2015
clostridiales Family XI Incertae sedis	Increased in CRC	1	Gut	Wu, 2013
Erysipelotrichales	Increased in ESCC	1	Gastric	Nasrollahzadeh, 2015
Lactobacillales	Decreased in PRO	1	Prostate	Cavarretta, 2017
Methanobacteriales	Increased in CRC	1	Gut	Mira-Pascual, 2014
Planctomycetes	Decreased in CRC	1	Gut	Thomas, 2016

*All studies in same direction

CRC= Colorectal cancer
ECCA= Extrahepatic cholangiocarcinoma
PRO= Prostate cancer

Taxonomic level - Family	Cancer and Result	Number of studies		Microbiota body site	Reference
Alcaligenaceae	Decreased in CRC	1		Gut	Chen, 2012
Bifidobacteriaceae	Decreased in rectal cancer	1		Gut	Youssef, 2018
Bifidobacteriaceae	Increased in CRC	1		Gut	Yang, 2019
Bifidobacteriaceae	Decreased in CRC	1		Gut	Chen, 2012
Campylobacteraceae	Increased in CRC	1		Gut	Wu, 2013
Clostridiaceae	Increased in BC	1		Gut	Goedert, 2015
Coriobacteriaceae	Increased in CRC	1		Gut	Chen, 2012
Enterobacteriaceae	Decreased in CRC	1		Gut	Ohigashi, 2013
Enterobacteriaceae	Increased in CRC	2		Gut	Mira-Pascual, 2014; Zackular, 2014
Erysipelotricaceae	Increased in CRC	1		Gut	Chen, 2012
Erysipelotricaceae	Decreased in CRC	1		Oral	Yang, 2019
Eubacteriaceae	Increased in CRC	1		Gut	Wu, 2013
Fusobacteriaceae	Increased in GC	1		Gut	Castaño-Rodríguez, 2017
Fusobacteriaceae	Increased in CRC	2	*	Gut	Chen, 2012; Wu, 2013
Gemellaceae	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Helicobacteraceae	Decreased in ESCC	1		Gastric	Nasrollahzadeh, 2015
Lachnospiraceae	Decreased in BC	1		Gut	Goedert, 2015
Lachnospiraceae	Decreased in CRC	6	*	Gut & Oral	Ahn, 2013; Baxter, 2016; Burns, 2015; Flemer, 2018; Sinha, 2016; Zackular, 2014
Lachnospiraceae	Increased in GC	2	*	Gastric	Aviles-Jimenez, 2014; Wang, 2016
Lactobacillaceae	Decreased in colon cancer	1		Gut	Youssef, 2018
Methylophilaceae	Increased ECCA	1		Bile duct	Aviles-Jimenez, 2016
Moraxellaceae	Decreased in CRC	1		Gut	Saito, 2019
Peptostreptococcaceae	Increased in CRC	2	*	Gut & Oral	Chen, 2012; Chen, 2012
Porphyromonadaceae	Increased in CRC	2	*	Gut	Chen, 2012; Wu, 2013
Prevotellaceae	Increased in CRC	1		Gut	Chen, 2012

Taxonomic level - Family	Cancer and Result	Number of studies		Microbiota body site	Reference
Pseudomonadaceae	Increased in FAC	1		Oral	Henrich, 2014
Rikenellaceae	Decreased in CRC	1		Gut	Burns, 2015
Ruminococcaceae	Decreased in CRC	5	*	Gut	Burns, 2015; Chen, 2012; Shah, 2018; Warren, 2013; Baxter, 2016 Goedert, 2015
Ruminococcaceae	Increased in BC	1		Gut	Wu, 2013
Staphylococcaceae	Increased in CRC	1		Gut	Cavarretta, 2017
Staphylococcaceae	Increased in PRO	1		Prostate	Yang, 2019
Streptococcaceae	Decreased in CRC	1		Oral	Cavarretta, 2017
Streptococcaceae	Decreased in PRO	1		Prostate	

*All studies in same direction

BC= Breast cancer

CRC= Colorectal cancer

ECCA= Extrahepatic cholangiocarcinoma

FAC= Fanconi Anaemia and oral squamous cell cancer

GC= Gastric cancer, HNSCC= Head and neck squamous cell carcinoma

PRO= Prostate cancer

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Achromobacter	Increased in GC	1		Gastric	Li, 2017
Acholeplasma	Decreased in ESCC	1		Oral	Chen, 2015
Actinobacillus	Increased in OC	1		Oral	Lim, 2018
Actinobacillus	Decreased in ESCC	1		Oral	Chen, 2015
Abiothrophia	Increased in LC	1		Lung	Hosgood, 2014
Abiothrophia	Decreased in OC	1		Oral	Hu, 2016
Acetanaerobacterium	Decreased in PRO	1		Gut & Urinary	Alanee, 2019
Acidaminobacter	Increased in CRC	1		Gut	Weir, 2013
Acidocella	Decreased in CRC	1		Gut	Chen, 2012
Acidovorax	Decreased in CRC	1		Gut	Gao, 2015
Acinetobacterjunii	Increased in LC	1		Lung	Cameron, 2017
Acinetobacter	Decreased in CRC	3	*	Gut	Thomas, 2016; Zeller, 2014; Gao, 2015
Actinomyces	Increased in OC	2	*	Oral	Lim, 2018; Wolf, 2017
Actinomyces	Increased in bladder cancer	1		Urinary	Bi, 2019
Actinomyces	Decreased in HNSCC	1		Oral	Wang, 2017
Aggregatibacter	Increased in GC	1		Gastric	Hu, 2018
Aggregatibacter	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Aggregatibacter	Decreased in ESCC	1		Oral	Chen, 2015
Aggregatibacter	Increased in OC	1		Oral	Lim, 2018
Aggregatibacter	Increased in PC	1		Oral	Fan, 2018
Alcaligenes	Increased in OC	1		Oral	Banerjee, 2017
Alistipes	Increased in CRC	1		Gut	Loke, 2018
Alloprevotella	Increased in GC	1		Gastric	Hu, 2018
Anaerococcus	Increased in CRC	1		Gut	Wu, 2013
Anaerostipes	Decreased in CRC	1		Gut	Chen, 2012

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Anaerotruncus	Increased in CRC	1		Gut	Chen, 2012
Atopium	Increased in IBC	1		Breast tissue	Hieken 2016
Atopobium	Increased in CRC	4	*	Gut	Ahn, 2013; Kasai, 2016; Vogtmann, 2016; Saito, 2019
Atopobium	Decreased in ESCC	1		Oral	Chen, 2015
Atopobium	Increased in PC	1		Oral	Farrell, 2012
Atopobium	Increased in OC	1		Oral	Perera, 2018
Bacillus	Increased in OC	1		Oral	Hu, 2016
Bacillus	Decreased in CRC	1		Gut	Thomas, 2016
Bacillus	Increased in OVC	1		Ovarian	Banerjee, 2017
Bacteroides	Increased in CRC	4		Gut	Amiot, 2015; Feng, 2015; Loke, 2018; Thomas, 2016
Bacteroides	Decreased in CRC	2		Gut	Burns, 2015; Loke, 2018
Bacteroides	Increased in PRO	2		Gut & Urinary	Alanee, 2019; Liss, 2018
Bifidobacterium	Decreased in bladder cancer	1		Urinary	Bi, 2019
Bifidobacterium	Decreased in CRC	4	*	Gut	Chen, 2012; Feng, 2015; Ohigashi, 2013; Zeller, 2014
Bilophila	Increased in CRC	1		Gut	Zeller, 2014
Blautia	Decreases in CRC	2	*	Gut	Allali, 2015; Chen, 2012
Brevundimonas	Increased in OC	1		Oral	Banerjee, 2017
Brevundimonas	Decreased in CRC	1		Gut	Gao, 2015
Bulleidia	Increased in CRC	1		Gut	Allali, 2015
Bulleidia	Decreased in ESCC	1		Oral	Chen, 2015
Butyrivibrio	Decreased in ESCC	1		Oral	Chen, 2015
Buttiauxella	Decreased in CRC	1		Gut	Gao, 2015
Campylobacter	Increased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Campylobacter	Decreased in CRC	1		Gut	Zeller, 2014
Campylobacter	Increased in CRC	2		Gut	Allali, 2015; Warren, 2013
Campylobacter	Decreased in PC	1		Oral	Farrell, 2012
Cardiobacterim	Decreased in ESCC	1		Oral	Chen, 2015

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Capnocytophaga	Increased in OC	2	*	Oral	Lim, 2018; Perera, 2018
Catenibacterium	Decreased in CRC	1		Gut	Chen, 2012
Catonella	Increased in OC	1		Oral	Zhao, 2017
Catonella	Increased in CRC	1		Gut	Chen, 2012
Catonella	Decreased in ESCC	1		Gut	Chen, 2015
Caulobacter	Increased in OC	1		Oral	Banerjee, 2017
Caulobacter	Decreased in CRC	1		Gut	Gao, 2015
Citrobacter	Decreased in CRC	1		Gut	Marchesi, 2011
Chlamydophila	Increased in OVC	1		Ovarian	Banerjee, 2017
Chyseobacterium	Increased in OVC	1		Ovarian	Banerjee, 2017
Clostridium	Increased in CRC	1		Gut	Allali, 2018
Clostridium XVIII & IV	Decreased in PRO	1		Gut & Urinary	Alanee, 2019
Collinsella	Increased in CRC	1		Gut	Marchesi, 2011
Comamonas	Increased in OC	1		Oral	Banerjee, 2017
Coprococcus	Decreased in CRC	1		Gut	Ahn, 2013
Corynebacterium	Decreased in bladder cancer	1		Urinary	Bučević Popović, 2018
Corynebacterium	Decreased in HNSCC	1		Oral	Hayes, 2018
Corynebacterium	Decreased in ESCC	1		Oral	Chen, 2015
Corynebacterium	Decreased in OC	1		Oral	Lim, 2018
Cronobacter	Decreased in CRC	1		Gut	Marchesi, 2011
Desulfovibrio	Increased in CRC	2	*	Gut	Thomas, 2016; Zeller, 2014
Dialister	Increased in OC	2	*	Oral	Börnigen, 2017; Zhao, 2017
Dialister	Decreased in ESCC	1		Oral	Chen, 2015
Dorea	Increased in CRC	1		Gut	Thomas, 2016
Dorea	Decreased in BC	1		Gut	Goedert, 2015
Edwardsiella	Increased in OC	1		Oral	Banerjee, 2017

Taxonomic level - Genus	Cancer and Result	Number of studies	Microbiota body site	Reference
Eikenella	Increased in CRC	1	Gut	Allali, 2015
Enhydrobacter	Decreased in CRC	1	Gut	Gao, 2015
Enterococcus	Increased in OC	1	Oral	Lee, 2017
Epilithonimonas	Decreased in CRC	1	Gut	Gao, 2015
Escherichia	Increased in GC	1	Gut	Qi, 2019
Esherichia	Increased in adenocarcinoma	1	Gut	Mori, 2018
Escherichia	Decreased in CRC	1	Gut	Thomas, 2016
Escherichia	Increased in OVC	1	Ovarian	Banerjee, 2017
Escherichia	Increased in OC	1	Oral	Banerjee, 2017
Eubacterium	Increased in CRC	1	Gut	Wu, 2013
Eubacterium	Decreased in CRC	2	Gut	Zeller, 2014; Yu, 2015
Faecalibacterium	Decreased in CRC	3	Gut	Allali, 2015; Chen, 2012; Shah, 2018
Faecalibacterium	Increased in CRC	1	Gut	Marchesi, 2011
Faecalibacterium	Decreased in PRO	1	Gut & Urinary	Alanee, 2019
Faecalibacterium	Increased in BC	1	Gut	Goedert, 2015
Fastidiosipila	Increased in CRC	1	Gut	Wu, 2013
Filifactor	Increased in CRC	1	Gut	Chen, 2012
Filifactor	Decreased in ESCC	1	Oral	Chen, 2015
Filifactor	Increased in OC	1	Oral	Zhao, 2017
Filifactor	Decreased in LC	1	Oral	Wang, 2019
Flavobacterium	Increased in GC	1	Gastric	Li, 2017
Flavobacterium	Decreased in CRC	1	Gut	Gao, 2015
Frateuria	Increased in OC	1	Oral	Banerjee, 2017
Fusobacterium	Increased in ECCA	1	Bile duct	Aviles-Jimenez, 2016
Fusobacterium	Increased in bladder cancer	1	Urinary	Bučević Popović, 2018
Fusobacterium	Increased in IBC	1	Breast tissue	Hieken 2016

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Fusobacterium	Increased in liver cancer	1		Tongue coat	Lu, 2016
Fusobacterium	Increased in OC	4	*	Oral	Lim,2018; Perera,2018; Schmidt, 2014; Zhao,2017
Fusobacterium	Increased in CRC	16	*	Gut & Oral	Allali 2018; Alexander, 2016; Burns, 2015; Chen, 2012; Feng, 2015; Gao, 2015; Marchesi, 2011; McCoy, 2013; Saito 2019; Shah 2018; Sinha, 2016; Vogtmann, 2016; Warren, 2013; Wu, 2013; Zackular, 2014
Fusobacterium	Increased in LSCC	1		Laryngeal	Gong, 2013
Fusobacterium	Decreased in GC	1		Oral	Hu, 2015
Fusobacterium	Increased in OVC	1		Ovarian	Banerjee, 2017
Gardnerella	Decreased in CRC	1		Gut	Chen, 2012
Gemella	Increased in CRC	3	*	Gut & Oral	Allali, 2015; Chen, 2012; Wu, 2013
Gemella	Increased in OC	1		Oral	Mok, 2017
Gemella	Increased in LSCC	1		Laryngeal	Gong, 2013
Gemmiger	Increased in HCC	1		Gut	Ren, 2019
Gluconacetobacter	Increased in IBC	1		Breast tissue	Hieken 2016
Granulicatella	Increased in PC	1		Oral	Farrell, 2012
Granulicatella	Increased in OC	1		Oral	Mok, 2017
Haemophilus	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Haemophilus	Decreased in CRC	1		Oral	Han, 2014
Haemophilus	Decreased in GC	1		Oral	Hu, 2015
Haemophilus	Decreased in OC	1		Oral	Wolf, 2017
Haemophilus	Increased in OC	3		Oral	Banerjee, 2017; Lim, 2018; Hu, 2016
Helicobacter	Increased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Holdemania	Increased in CRC	1		Gut	Wu, 2013
Holdemania	Decreased in CRC	1		Gut	Warren, 2013
Hydrogenophaga	Increased in IBC	1		Breast tissue	Hieken 2016
Janthinobacterium	Decreased in CRC	1		Gut	Gao, 2015
Kingella	Increased in HNSCC	1		Oral	Hayes, 2018
Kingella	Increased in CRC	1		Gut	Wu, 2013

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Klebsiella	Increased in GC	2	*	Gut	Li, 2017; Qi, 2019
Kluyvera	Decreased in CRC	1		Gut	Marchesi, 2011
Lachnobacterium	Decreased in CRC	1		Gut	Saito, 2019
Lachnoclostridium	Decreased in gastro intestinal tract	1		Gut	Youssef, 2018
Lachnospira	Decreased in CRC	1		Gut	Allali, 2015
Lachnospira	Decreased in PRO	1		Gut & Urinary	Alanee, 2019
Lactobacillus	Increased in GC	1		Gut	Qi, 2019
Lactobacillus	Increased in IBC	1		Breast tissue	Hieken 2016
Lactobacillus	Increased in CRC	1		Gut	Amiot, 2015
Lactobacillus	Decreased in CRC	1		Gut	Kanazawa, 1996
Lactobacillus	Increased in HNSCC	1		Oral	Guerrero-Preston, 2016
Lactobacillus	Decreased in bladder cancer	1		Urinary	Bi, 2019
Lactobacillus	Increased in GC	2	*	Gastric	Qi, 2019; Wang, 2016
Lactococcus	Increased in GC	1		Oral	Hu, 2015
Lautropia	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Lautropia	Decreased in ESCC	1		Oral	Chen, 2015
Leptotrichia	Associated with low risk of PC	2		Oral	Fan, 2016; Fan, 2018
Leptotrichia	Increased in PC	1		Gut	Torres, 2015
Leptotrichia	Increased in CRC	3	*	Gut	Han, 2014; Warren, 2013; Zeller, 2014
Leptotrichia	Increased in OC	1		Oral	Lim, 2018
Leuconostoc	Increased in CRC	1		Gut	Amiot, 2015
Megasphaera	Decreased in ESCC	1		Oral	Chen, 2015
Megasphaera	Increased in LC	1		Oral	Lee, 2016
Mesorhizobium	Decreased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Methylobacterium	Decreased in CRC	1		Gut	Chen, 2012

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Mogibacterium	Increased in CRC	1		Gut	Chen, 2012
Moryella	Decreased in ESCC	1		Oral	Chen, 2015
Mycoplasma	Increased in OVC	1		Ovarian	Banerjee, 2017
Neisseria	Decreased in GC	2		Gastric & Oral	Aviles-Jimenez, 2014; Hu, 2015
Neisseria	Increased in GC	1		Gastric	Hu, 2018
Neisseria	Increased in LC	1		Lung	Hosgood, 2014
Neisseria	Increased in esophageal cancer	1		Oral	Peters, 2017
Neisseria	Decreased in HNSCC	1		Oral	Guerrero-Preston, 2016
Neisseria	Decreased in OC	1		Oral	Wolf, 2017
Neisseria	Increased in OC	2		Oral	Lim, 2018; Guerrero-Preston, 2016
Neisseria	Increased in PC	1		Oral	Torres, 2015
Neisseria	Decreased in PC	1		Oral	Farrell, 2012
Nesterenkonia	Decreased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Nesterenkonia	Decreased in CRC	1		Gut	Gao, 2015
Nevskia	Decreased in CRC	1		Gut	Gao, 2015
Nitrospirae	Increased in GC	1		Gastric	Wang, 2016
Ochrobactrum	Decreased in CRC	1		Gut	Chen, 2012
Odoribacter	Increased in CRC	2	*	Gut	Thomas, 2016; Wu, 2013
Oribacterium	Increased in OC	1		Oral	Lim, 2018
Oribacterium	Increased in liver cancer	1		Tongue coat	Lu, 2016
Oscillibacter	Decreased in gastro intestinal tract	1		Gut	Youssef, 2018
Oscillibacter	Increased in CRC	2	*	Gut	Loke, 2018; Wu, 2013
Paludibacter	Decreased in OC	1		Oral	Lim, 2018
Parabacteroides	Increased in CRC	1		Gut	Thomas, 2016
Parabacteroides	Decreased in CRC	2		Gut	Chen, 2012; Warren, 2013

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Parabacteroides	Increased in HCC	1		Gut	Ren, 2019
Paraprevotella	Decreased in CRC	1		Gut	Chen, 2012
Parcubacteria	Increased in CRC	1		Gut	Thomas, 2016
Parvimonas	Increased in CRC	5	*	Gut	Feng, 2015; Saito, 2019; Shah, 2018; Wu, 2013; Zeller, 2014
Parvimonas	Increased in OC	2	*	Oral	Lee, 2017; Zhao, 2017
Parvimonas	Increased in HNSCC	1		Oral	Wang, 2017
Pediococcus	Increased in CRC	1		Gut	Amiot, 2015
Pediococcus	Increased in OVC	1		Ovarian	Banerjee, 2017
Pedobacter	Decreased in CRC	1		Gut	Gao, 2015
Peptococcus	Increased in OC	1		Oral	Zhao, 2017
Peptococcus	Decreased in ESCC	1		Oral	Chen, 2015
Peptoniphilus	Increased in OC	1		Oral	Banerjee, 2017
Peptostreptococcus	Increased in CRC	5	*	Gut & Oral	Chen, 2012; Chen, 2012; Gao, 2015; Wu, 2013; Zeller, 2014;
Peptostreptococcus	Increased in OC	2	*	Oral	Lee, 2017; Zhao, 2017
Phascolarctobacterium	Increased in CRC	2		Gut	Thomas, 2016; Weir, 2013
Phascolarctobacterium	Decreased in CRC	1		Gut	Chen, 2012
Plesiomonas	Increased in OC	1		Oral	Banerjee, 2017
Porphyromonas	Increased in CRC	8	*	Gut	Ahn, 2013; Allali 2018; Chen, 2012; Sinha, 2016; Vogtmann, 2016; Wu, 2013; Zackular, 2014; Zeller, 2014
Porphyromonas	Increased in ESCC	1		Oral	Chen, 2015
Porphyromonas	Increased in PC	1		Oral	Torres, 2015
Porphyromonas	Decreased in GC	1		Oral	Hu, 2015
Porphyromonas	Decreased in OC	1		Oral	Lim, 2018
Prevotella	Increased in ECCA	1		Bile duct	Aviles-Jimenez, 2016
Prevotella	Decreased in OC	1		Oral	Wolf, 2017
Prevotella	Increased in OVC	1		Ovarian	Banerjee, 2017
Prevotella	Increased in OC	1		Oral	Schmidt, 2014
Prevotella	Increased in CRC	2	*	Gut	Amiot, 2015; Mira-Pascual, 2014

Taxonomic level - Genus	Cancer and Result	Number of studies	Microbiota body site	Reference
Prevotella	Increased in LSCC	1	Laryngeal	Gong, 2013
Prevotella	Increased in ESCC	1	Oral	Chen, 2015
Prevotella	Decreased in PC	1	Oral	Farrell, 2012
Propionibacterium	Decreased in CRC	1	Gut	Gao, 2015
Provindencia	Increased in CRC	1	Gut	Burns, 2015
Pseudoflavonifractor	Increased in CRC	1	Gut	Zeller, 2014
Pseudoflavonifractor	Decreased in CRC	1	Gut	Warren, 2013
Pseudomonas	Increased in OC	1	Oral	Perera, 2018
Pseudomonas	Decreased in CRC	1	Gut	Thomas, 2016
Pseudomonas	Increased in GC	1	Gastric	Li, 2017
Psychrobacter	Decreased in CRC	1	Gut	Gao, 2015
Rachnella	Decreased in CRC	1	Gut	Gao, 2015
Roseburia	Increased in CRC	1	Gut	Marchesi, 2011
Rothia	Decreased in ECCA	1	Bile duct	Aviles-Jimenez, 2016
Rothia	Increased in OC	2	Oral	Banerjee, 2017; Lim, 2018
Rothia	Decreased in OC	1	Oral	Schmidt, 2014
Ruminococcus	Increased in gastro intestinal tract	1	Gut	Youssef, 2018
Ruminococcus	Increased in CRC	1	Gut	Allali, 2018
Ruminococcus	Decreased in CRC	4	Gut	Allali, 2015; Feng, 2015; Zeller, 2014; Warren, 2013
Schwartzia	Increased in OC	1	Oral	Wolf, 2017
Selenomonas	Increased in OC	1	Oral	Wolf, 2017
Selenomonas	Increased in CRC	1	Gut	Allali, 2018
Salmonella	Decreased in CRC	2	Gut	Marchesi, 2011; Saito, 2019
Scardovia	Decreased in OC	1	Oral	Börnigen, 2017
Serratia	Decreased in GC	1	Gastric	Li, 2017
Serratia	Decreased in CRC	1	Gut	Marchesi, 2011
Serratia	Increased in OC	1	Oral	Banerjee, 2017

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Shigella	Increased in GC	1		Gastric	Wang, 2016
Shigella	Decreased in CRC	1		Gut	Marchesi, 2011
Slackia	Increased in OC	1		Gut	Lee, 2017
Slackia	Increased in CRC	1		Gut	Marchesi, 2011
Slackia	Decreased in CRC	1		Gut	Kasai, 2016
Solobacterium	Increased in CRC	1		Gut	Wu, 2013
Solobacterium	Decreased in CRC	1		Oral	Yang, 2019
Succinivibrio	Increased in CRC	1		Gut	Loke, 2018
Subdoligranulum	Increased in gastro intestinal tract	1		Gut	Youssef, 2018
Sphingobacterium	Decreased in CRC	1		Gut	Gao, 2015
Sphingomonas	Increased in CRC	1		Gut	Richard, 2018
Sphingomonas	Decreased in CRC	1		Gut	Gao, 2015
Staphylococcus	Decreased in CRC	2	*	Gut	Mira-Pascual, 2014; Ohigashi, 2013
Staphylococcus	Decreased in LC	1		Lung	Liu, 2018
Staphylococcus	Increased in PRO	1		Prostate	Cavarretta, 2017
Staphylococcus	Increased in OVC	1		Ovarian	Banerjee, 2017
Stenothrophomonas	Increased in CCA cancer tissue	1		Bile duct	Chng, 2016
Stenotrophomonas	Decreased in CRC	1		Gut	Gao, 2015
Stenotrophomonas	Increased in GC	1		Gastric	Li, 2017
Streptococcus	Decreased in OC	4	*	Oral	Hu, 2016; Mok, 2017; Schmidt, 2014; Wolf, 2017
Streptococcus	Decreased in CRC	2		Gut	Yang, 2019; Feng, 2015
Streptococcus	Increased in CRC	3		Gut & Oral	Allali, 2015; Flemer, 2018; Shah, 2018
Streptococcus	Increased in OVC	1		Ovarian	Banerjee, 2017
Streptococcus	Decreased in HNSCC	1		Oral	Shin, 2017
Streptococcus	Increased in HNSCC	1		Oral	Guerrero-Preston, 2016

Taxonomic level - Genus	Cancer and Result	Number of studies		Microbiota body site	Reference
Streptococcus	Decreased in bladder cancer	2		Urinary	Bučević Popović, 2018; Bi, 2019
Streptococcus	Decreased in FAC	1		Oral	Henrich, 2014
Streptococcus	Decreased in PC	1		Oral	Farrell, 2012
Streptococcus	Increased in ESCC	1		Oral	Chen, 2015
Streptococcus, Streptococcus species	Increased in PRO	2	*	Gut & Urinary	Alanee, 2019; Liss, 2018
Streptococcus	Increased in LC	2	*	Airway & Lung	Hosgood, 2014; Tsay, 2018
TG5	Decreased in ESCC	1		Oral	Chen, 2015
Treponema	Increases in OC	1		Oral	Wolf, 2017
Treponema	Increased in OVC	1		Ovarian	Banerjee, 2017
Treponema	Decreased in LC	1		Oral & Lung	Wang, 2019
Veillonella	Increased in GC	1		Gastric	Castaño-Rodríguez, 2017
Veillonella	Increased in LC	2	*	Airway & Lung	Lee, 2016; Tsay, 2018
Veillonella	Increased in OC	1		Oral	Lim, 2018
Veillonella	Decreased in OC	2		Oral	Wolf, 2017; Mok, 2017
Veillonella	Increased in PRO	1		Gut & Urinary	Alanee, 2019
Veillonella	Decreased in bladder cancer	2	*	Urinary	Bi, 2019; Bučević Popović, 2018

*All studies in same direction

ESCC= Esophageal squamous cell carcinoma

GC= Gastric cancer

BC= Breast cancer

LC= Lang cancer

CCA= Cholangiocarcinoma

OV= ovarian cancer

CRC= Colorectal cancer

PC= pancreatic cancer

ECCA= Extrahepatic cholangiocarcinoma

PRO= prostate cancer

FAC= Fanconi Anaemia and oral squamous cell cancer

HCC= hepato cellular carcinoma

HNSCC= Head and neck squamous cell carcinoma

IBC= Invasive Breast Cancer

LSCC= Laryngeal squamous cell carcinoma

OC= Oral cancer

OCSCC= Oral cavity squamous cell carcinoma

Taxonomic level - Species	Cancer and Result	Number of studies		Microbiota body site	Reference
Actinomyces sp. HPA0247	Increased in BC	1		Gut	Zhu, 2018
Aggregatibacter actinomycetemcomitans	Associated with PC	1		Oral	Fan, 2018
Alistipes finegoldii	Increased in CRC	1		Gut	Dai, 2018
Alistipes putredenis	Increased in CRC	1		Gut	Feng, 2015
Atopobium cluster	Decreased in CRC	1		Gut	Ohigashi, 2013
Bacteroides capillosus	Decreased in CRC	1		Gut	Weir, 2013
Bacteroides finegoldii	Decreased in CRC	1		Gut	Weir, 2013
Bacteroides fragilis	Increased in CRC	4	*	Gut	Dai, 2018; Hale, 2018; Kasai, 2016; Yu, 2015
Bacteroides massiliensis	Increased in PRO	1		Gut	Golombos, 2018
Bacteroides intestinalis	Decreased in CRC	1		Gut	Weir, 2013
Bacteroides uniformis	Decreased in GC	1		Gastric	Liu, 2019
Bacteroides uniformis	Decreased in CRC	1		Gut	Burns, 2015
Bacteroides species	Increased in PRO	1		Gut	Liss, 2018
Bacteroides sulfinogenic	Increased in CRC	1		Gut	Hale, 2018
Bilophila wadsworthia	Increased in CRC	1		Gut	Feng, 2015
Blautia coccoides	Increased in CRC	2	*	Gut	Amiot, 2015; Mira-Pascual, 2014
Burkholderia fungorum	Increased in GC	1		Gastric	Wang, 2016
Campylobacter concisus	Increased in OC	1		Oral	Perera, 2018
Citrobacter farmeri	Increased in CRC	1		Gut	Weir, 2013
Clostridium coccoides group	Decreased in CRC	1		Gut	Ohigashi, 2013
Clostridium leptum subgroup	Decreased in CRC	1		Gut	Ohigashi, 2013
Dialister invisus	Decreased in CRC	1		Gut	Weir, 2013
Dialister pneumosintes	Increased in GC	1		Gastric	Coker, 2018
Dialister pneumosintes	Decreased in CRC	1		Gut	Weir, 2013
Dorea formicigenerans	Decreased in CRC	1		Gut	Weir, 2013

Taxonomic level - Species	Cancer and Result	Number of studies		Microbiota body site	Reference
Enterococcus gallinarum	Increased in BC	1		Gut	Zhu, 2018
Erwinia amylovora	Increased in BC	1		Gut	Zhu, 2018
Escherichia coli	Increased in LC	1		Oral	Cameron, 2017
Escherichia coli	Increased in BC	1		Gut	Zhu, 2018
Escherichia coli	Increased in CRC	2	*	Gut	Fukugaiti, 2015; Kohoutova, 2014
Eubacterium eligens	Decreased in BC	1		Gut	Zhu, 2018
Eubacterium rectale	Decreased in PRO	1		Gut	Golombos, 2018
Eubacterium ventriosum	Decreased in CRC	1		Gut	Yu, 2015
Faecalibacterium prauznitzii	Increased in CRC	1		Gut	Amiot, 2015
Faecalibacterium prauznitzii	Decreased in CRC	1		Gut	Wu, 2013
Faecalibacterium prauznitzii	Decreased in PRO	1		Gut	Golombos, 2018
Fusobacterium necrophorum	Increased in CC	1		Cervical	Audirac-Chalifour, 2016
Fusobacterium nucleatum	Increased in OC	1		Oral	Hsiao, 2018
Fusobacterium nucleatum	Increased in CRC	11	*	Gut	Yu, 2015; Alexander, 2016; Amitay, 2017; Baxter, 2016; Dai, 2018; Fukugaiti, 2015; Hale, 2018; Mira-Pascual, 2014; Yachida, 2019; Yu, 2017
Gardnerella vaginalis	Decreased in CC	1		Cervical	Audirac-Chalifour, 2016
Granulicatella adiacens	Increased in LC	1		Oral	Cameron, 2017
Klebsiella sp_1_1_55	Increased in BC	1		Gut	Zhu, 2018
Helicobacter pylori	Decreased in GC	1		Gastric	Liu, 2019
Lachnobacterium bovis	Decreased in CRC	1		Gut	Weir, 2013
Lachnospira pectinoschiza	Decreased in CRC	1		Gut	Weir, 2013
Lactobacillus coleohominis	Increased in GC	1		Gastric	Aviles-Jimenez, 2014
Lactobacillus crispatus	Decreased in CC	1		Cervical	Audirac-Chalifour, 2016
Lactobacillus iners	Decreased in CC	1		Cervical	Audirac-Chalifour, 2016
Megamonas hypermegale	Decreased in CRC	1		Gut	Weir, 2013
Methanobrevibacterium	Increased in CRC	1		Gut	Mira-Pascual, 2014
Mycoplasma salivarium	Increased in FAC	1		Oral	Henrich, 2014

Taxonomic level - Species	Cancer and Result	Number of studies		Microbiota body site	Reference
Parvimonas micra	Increased in GC	1		Gastric	Coker, 2018
Parvimonas micra	Increased in CRC	4	*	Gut	Baxter, 2016; Dai, 2018; Yu, 2015; Yu, 2017
Peptostreptococcus stomatis	Increased in GC	1		Gastric	Coker, 2018
Peptostreptococcus stomatis	Increased in CRC	3	*	Gut	Baxter, 2016; Yu, 2015; Yu, 2017
Porphyromonas asaccharolytica	Increased in CRC	2	*	Gut	Baxter, 2016; Dai, 2018
Porphyromonas gingivalis	Increased in EC	1		Oral	Peters, 2017
Porphyromonas gengivalis	Increased in PC	1		Oral	Fan, 2018
Prevotella melaninogenica	Increased in GC	1		Gastric	Liu, 2019
Prevotella melaninogenica	Decreased in CRC	1		Oral	Yang, 2019
Prevotella amnii	Increased in BC	1		Gut	Zhu, 2018
Prevotella copri	Decreased in GC	1		Gastric	Liu, 2019
Prevotella copri	Decreased in CRC	1		Gut	Weir, 2013
Prevotella denticola	Increased in CRC	1		Oral	Yang, 2019
Prevotella loeschi	Increased in OC	1		Oral	Perera, 2018
Prevotella oris	Decreased in CRC	1		Gut	Weir, 2013
Prevotella intermedia	Increased in OC	1		Oral	Hsiao, 2018
Prevotella intermedia	Increased in CRC	2	*	Gut & Oral	Dai, 2018; Yang, 2019
Prevotella melaninogenica	Decreased in CRC	1		Gut	Yang, 2019
Prevotella salivae	Increased in OC	1		Oral	Perera, 2018
Prevotella salivae	Increased in FAC	1		Oral	Henrich, 2014
Prevotella sp	Increased in CRC	2	*	Oral	Flemer, 2018; Henrich, 2014
Prevotella spp	Increased in FAC	1		Oral	Henrich, 2014
Prevotella tannerae	Increased in OC	1		Oral	Hsiao, 2018
Propionibacterium acnes	Increased in GC	1		Gastric	Liu, 2019
Pseudobutyrivibrio ruminis	Decreased in CRC	1		Gut	Weir, 2013
Rothia mucilaginosa	Decreased in FAC	1		Oral	Henrich, 2014
Ruminococcus albus	Decreased in CRC	1		Gut	Weir, 2013

Taxonomic level - Species	Cancer and Result	Number of studies		Microbiota body site	Reference
Ruminococcus obeum	Decreased in CRC	1		Gut	Weir, 2013
Serratia marcescens	Increased in GC	1		Gastric	Li, 2017
Shewanella putrefaciens	Increased in BC	1		Gut	Zhu, 2018
Slackia exigua	Increased in GC	1		Gastric	Coker, 2018
Solobacterium moorei	Decreased in CRC	1		Oral	Yang, 2019
Solobacterium moorei	Increased in CRC	2		Gut	Yu, 2015; Yu, 2017
Sphingobium yanoikuyae	Decreased in GC	1		Gastric	Hu, 2018
Streptococcus anginosus	Increased in GC	2	*	Gastric	Coker, 2018; Liu, 2019
Streptococcus intermedius	Increased in LC	1		Oral	Cameron, 2017
Streptococcus pneumoniae	Increased in EC	1		Oral	Peters, 2017
Streptococcus sp6	Increased in LC	1		Oral	Cameron, 2017
Tannerella forsythia	Increased in EC	1		Oral	Peters, 2017
Thermanaerovibrio acidaminovorans	Increased in CRC	1		Gut	Dai, 2018
Treponema denticola	Increased in CRC	1		Oral	Yang, 2019

*All studies in same direction

BC=Breast cancer

CC= Cervical cancer

CRC= Colorectal cancer

EC= Esophageal cancer

FAC= Fanconi Anaemia and oral squamous cell cancer

GC= Gastric cancer

PC= Pancreatic cancer

Findings for Alpha diversity	Number of studies		Microbiota body site	Reference
Higher alpha diversity in Cervical Cancer	1		Cervical	Audirac-Chalifour, 2016
Higher alpha diversity in Cervical intraepithelial neoplasia	1		Cervical	Oh, 2015
Lower alpha diversity in CRC	5		Gut	Ahn, 2013; Gao, 2015; Ohigashi, 2013; Richard, 2018; Yu, 2015
Higher alpha diversity in CRC	7		Gut	Burns, 2015; Chen, 2012; Feng, 2015; McCoy, 2013; Mira-Pascual, 2014; Thomas, 2016; Xu, 2017
Lower alpha diversity in BC	2	*	Gut	Goedert, 2015; Goedert, 2018
Lower alpha diversity in GC at species level	1		Gastric	Hu, 2018
Higher alpha diversity in GC at species level	1		Gut	Qi, 2019
Lower alpha diversity in GC	1		Oral	Hu, 2015
Higher alpha diversity in HCC	1		Gut	Ren, 2019
Higher alpha diversity in LC	1		Oral	Lee, 2016
Lower alpha diversity in LC	1		Oral	Wang, 2019
Higher alpha diversity in OC	2		Oral	Hu, 2016; Zhao 2017
Lower alpha diversity in OC	1		Oral	Perera, 2018
Higher alpha diversity in liver cancer	1		Oral	Lu, 2016
Higher alpha diversity in bladder cancer	1		Urinary	Bi, 2019
Lower alpha diversity in HNSC	1		Oral	Shin, 2017

*All studies in same direction

BC= Breast cancer
CRC= colorectal cancer
GC= gastric cancer
HCC= hepatocellular carcinoma
LC= lung cancrer
OC= oral cancer
HNSC= head neck squamous cell carcinoma

Findings for Beta diversity	Number of studies		Microbiota body site	Reference
Significant for CRC	7	*	Gut	Baxter, 2016; Chen, 2012; Flemer, 2016; Mira-Pascual, 2014; Saito, 2019; Wu, 2013; Xu, 2017
Significant for BC (higher in cases)	1		Gut	Goedert, 2015
Significant for OSCC	1		Oral	Wolf, 2017
Signifcant for OSCC (lower in cases)	1		Oral	Perera, 2018
Significant for FAC (lower in cases)	1		Oral	Henrich, 2014
Significant for HNSCC (lower in cases)	2	*	Oral	Shin, 2017; Wang, 2017
Significant for ESCC (lower in cases)	1		Oral	Chen, 2015
Significant for GC (lower in cases)	1		Oral	Hu, 2015
Significant for PRO	1		Gut	Liss, 2018

*All studies in same direction

BC= Breast cancer
CRC= colorectal cancer
ESCC=Esophageal squamous cell carcinoma
FAC= Fanconi Anaemia and oral squamous cell cancer
GC= Gastric cancer
OCSCC= Oral cavity squamous cell carcinoma
PRO= prostate cancer

Supplementary material: Comparison of fecal sample collection methods for microbial analysis embedded within colorectal cancer screening programs

Table S2a – Results from qPCR, sample frozen immediately with no solution.

Sample name	Bacterial DNA concentration (ng/µl)	Sample mass (mg)	Bacterial quantity/g	log(Bacterial quantity/g)
Cryotube1 01 (-80°C)	74.24	424		
Cryotube1 02 (-80°C)	84.1	239		
Cryotube1 03 (-80°C)	91.78	216		
Cryotube1 04 (-80°C)	181.45	246		
Cryotube1 05 (-80°C)	55.28	244		
Cryotube1 06 (-80°C)	110.2	299		
Cryotube1 07 (-80°C)	64.24	275		
Cryotube1 08 (-80°C)	98.1	187		
Cryotube1 09 (-80°C)	101.89	254		
Cryotube1 10 (-80°C)	37.75	45		
Cryotube1 11 (-80°C)	111.94	333		
Cryotube1 12 (-80°C)	75.34	191		
Cryotube1 13 (-80°C)	99.58	414		
Cryotube1 14 (-80°C)	30.23	441		
Cryotube1 15 (-80°C)	106.75	191		
Cryotube1 16 (-80°C)	99.05	277		
Cryotube1 17 (-80°C)	60.62	496		
Cryotube1 18 (-80°C)	117.56	343		
Cryotube1 19 (-80°C)	71.03	290		
Cryotube2 01 (-80°C)	101.77	192	1.39E+11	11.14
Cryotube2 02 (-80°C)	36.15	129	4.70E+10	10.67
Cryotube2 03 (-80°C)	101.02	219	8.79E+10	10.94
Cryotube2 04 (-80°C)	67.67	103	1.11E+11	11.04
Cryotube2 05 (-80°C)	73.2	209	5.11E+10	10.71
Cryotube2 06 (-80°C)	102.85	159	2.19E+11	11.34
Cryotube2 07 (-80°C)	28.87	75	7.38E+10	10.87
Cryotube2 08 (-80°C)	43.74	88	9.42E+10	10.97
Cryotube2 09 (-80°C)	36.73	116	2.04E+10	10.31
Cryotube2 10 (-80°C)	88.39	113	1.76E+11	11.25
Cryotube2 11 (-80°C)	71.05	109	2.17E+11	11.34
Cryotube2 12 (-80°C)	55.14	109	9.06E+10	10.96
Cryotube2 13 (-80°C)	22.29	118	2.04E+10	10.31
Cryotube2 14 (-80°C)	21.76	171	1.27E+10	10.10
Cryotube2 15 (-80°C)	80.43	100	1.24E+11	11.09
Cryotube2 16 (-80°C)	53.4	116	7.39E+10	10.87
Cryotube2 17 (-80°C)	71.78	167	8.87E+10	10.95
Cryotube2 18 (-80°C)	98.29	113	1.56E+11	11.19
Cryotube2 19 (-80°C)	38.34	98	9.51E+10	10.98

Table S2b – Results from qPCR, specimen collection cards.

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ slices	log(Bacterial quantity/slices)
GenSaver 01 (-80°C)	28.69	2.85E+10	10.46
GenSaver 02 (-80°C)	42.52	3.66E+10	10.56
GenSaver 03 (-80°C)	44.10	4.09E+10	10.61
GenSaver 04 (-80°C)	43.09	2.52E+10	10.40
GenSaver 05 (-80°C)	48.76	4.01E+10	10.60
GenSaver 06 (-80°C)	37.77	4.49E+10	10.65
GenSaver 07 (-80°C)	40.75	4.50E+10	10.65
GenSaver 08 (-80°C)	53.93	6.16E+10	10.79
GenSaver 09 (-80°C)	38.49	3.76E+10	10.58
GenSaver 10 (-80°C)	21.25	2.06E+10	10.31
GenSaver 11 (-80°C)	16.17	9.96E+09	10.00
GenSaver 12 (-80°C)	53.27	5.66E+10	10.75
GenSaver 13 (-80°C)	52.56	6.57E+10	10.82
GenSaver 14 (-80°C)	12.37	8.34E+09	9.92
GenSaver 15 (-80°C)	50.14	3.95E+10	10.60
GenSaver 16 (-80°C)	87.88	1.10E+11	11.04
GenSaver 17 (-80°C)	46.83	5.08E+10	10.71
GenSaver 18 (-80°C)	49.92	4.81E+10	10.68
GenSaver 19 (-80°C)	50.71	6.49E+10	10.81
GenSaver 01 (RT)	27.94	1.75E+10	10.24
GenSaver 02 (RT)	38.75	1.74E+10	10.24
GenSaver 03 (RT)	25.90	1.21E+10	10.08
GenSaver 04 (RT)	46.60	1.79E+10	10.25
GenSaver 05 (RT)	42.35	1.17E+10	10.07
GenSaver 06 (RT)	20.64	1.17E+10	10.07
GenSaver 07 (RT)	30.37	1.49E+10	10.17
GenSaver 08 (RT)	60.90	2.71E+10	10.43
GenSaver 09 (RT)	41.14	1.88E+10	10.28
GenSaver 10 (RT)	18.60	8.33E+09	9.92
GenSaver 11 (RT)	39.20	2.30E+10	10.36
GenSaver 12 (RT)	45.11	3.52E+10	10.55
GenSaver 13 (RT)	46.01	2.32E+10	10.37
GenSaver 14 (RT)	7.51	3.03E+09	9.48
GenSaver 15 (RT)	31.27	1.43E+10	10.15
GenSaver 16 (RT)	71.45	4.64E+10	10.67
GenSaver 17 (RT)	18.26	9.40E+09	9.97
GenSaver 18 (RT)	29.20	1.51E+10	10.18
GenSaver 19 (RT)	54.59	3.77E+10	10.58
GenCollect 01 (-80°C)	15.68	1.73E+10	10.24
GenCollect 02 (-80°C)	27.49	2.79E+10	10.44
GenCollect 03 (-80°C)	44.97	5.52E+10	10.74
GenCollect 04 (-80°C)	70.51	6.26E+10	10.80

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ slices	log(Bacterial quantity/slices)
GenCollect 05 (-80°C)	55.42	4.27E+10	10.63
GenCollect 06 (-80°C)	41.59	5.70E+10	10.76
GenCollect 07 (-80°C)	27.61	3.12E+10	10.49
GenCollect 08 (-80°C)	45.63	5.11E+10	10.71
GenCollect 09 (-80°C)	33.67	3.70E+10	10.57
GenCollect 10 (-80°C)	16.82	1.79E+10	10.25
GenCollect 11 (-80°C)	35.18	1.97E+10	10.30
GenCollect 12 (-80°C)	69.91	4.96E+10	10.70
GenCollect 13 (-80°C)	62.19	5.26E+10	10.72
GenCollect 14 (-80°C)	8.29	2.47E+09	9.39
GenCollect 15 (-80°C)	69.78	3.88E+10	10.59
GenCollect 16 (-80°C)	110.60	8.80E+10	10.94
GenCollect 17 (-80°C)	43.43	2.59E+10	10.41
GenCollect 18 (-80°C)	71.65	3.75E+10	10.57
GenCollect 19 (-80°C)	66.12	5.11E+10	10.71
GenCollect 01 (RT)	26.67	1.24E+10	10.09
GenCollect 02 (RT)	42.69	1.97E+10	10.29
GenCollect 03 (RT)	15.80	8.20E+09	9.91
GenCollect 04 (RT)	45.24	1.59E+10	10.20
GenCollect 05 (RT)	44.84	2.18E+10	10.34
GenCollect 06 (RT)	24.81	1.42E+10	10.15
GenCollect 07 (RT)	27.90	1.77E+10	10.25
GenCollect 08 (RT)	56.88	3.95E+10	10.60
GenCollect 09 (RT)	31.05	1.73E+10	10.24
GenCollect 10 (RT)	8.70	4.24E+09	9.63
GenCollect 11 (RT)	28.24	1.06E+10	10.03
GenCollect 12 (RT)	55.23	3.31E+10	10.52
GenCollect 13 (RT)	35.33	3.91E+10	10.59
GenCollect 14 (RT)	9.55	4.77E+09	9.68
GenCollect 15 (RT)	41.18	3.49E+10	10.54
GenCollect 16 (RT)	81.09	1.39E+11	11.14
GenCollect 17 (RT)	50.51	9.27E+10	10.97
GenCollect 18 (RT)	43.55	4.40E+10	10.64
GenCollect 19 (RT)	53.37	1.08E+11	11.03

Table S2c – Results from qPCR, FIT tubes.

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ mL	log(Bacterial quantity/ mL)
OC-Auto Sampling 01 (-80°C)	2.44	1.04E+09	9.02
OC-Auto Sampling 02 (-80°C)	5.27	1.93E+09	9.29
OC-Auto Sampling 03 (-80°C)	4.69	2.84E+09	9.45
OC-Auto Sampling 04 (-80°C)	3.11	4.69E+08	8.67
OC-Auto Sampling 05 (-80°C)	7.31	3.46E+09	9.54
OC-Auto Sampling 06 (-80°C)	4.44	2.89E+09	9.46
OC-Auto Sampling 07 (-80°C)	5.38	2.07E+09	9.32
OC-Auto Sampling 08 (-80°C)	5.66	8.47E+08	8.93
OC-Auto Sampling 09 (-80°C)	4.24	2.73E+09	9.44
OC-Auto Sampling 10 (-80°C)	3.00	1.01E+09	9.01
OC-Auto Sampling 11 (-80°C)	4.71	1.32E+09	9.12
OC-Auto Sampling 12 (-80°C)	4.13	6.99E+09	9.84
OC-Auto Sampling 13 (-80°C)	3.32	1.70E+09	9.23
OC-Auto Sampling 14 (-80°C)	1.34	9.35E+07	7.97
OC-Auto Sampling 15 (-80°C)	3.93	1.75E+09	9.24
OC-Auto Sampling 16 (-80°C)	7.07	6.45E+09	9.81
OC-Auto Sampling 17 (-80°C)	3.21	6.95E+08	8.84
OC-Auto Sampling 18 (-80°C)	4.39	1.41E+09	9.15
OC-Auto Sampling 19 (-80°C)	4.73	4.75E+09	9.68
OC-Auto Sampling 01 (+4°C)	5.18	6.51E+08	8.81
OC-Auto Sampling 02 (+4°C)	2.49	9.66E+08	8.99
OC-Auto Sampling 03 (+4°C)	3.23	1.42E+09	9.15
OC-Auto Sampling 04 (+4°C)	3.27	6.19E+08	8.79
OC-Auto Sampling 05 (+4°C)	4.15	1.57E+09	9.20
OC-Auto Sampling 06 (+4°C)	3.06	1.44E+09	9.16
OC-Auto Sampling 07 (+4°C)	3.42	1.59E+09	9.20
OC-Auto Sampling 08 (+4°C)	3.71	2.81E+09	9.45
OC-Auto Sampling 09 (+4°C)	4.69	5.47E+07	7.74
OC-Auto Sampling 10 (+4°C)	4.13	4.66E+09	9.67
OC-Auto Sampling 11 (+4°C)	3.66	1.81E+09	9.26
OC-Auto Sampling 12 (+4°C)	3.74	5.93E+09	9.77
OC-Auto Sampling 13 (+4°C)	2.51	5.23E+08	8.72
OC-Auto Sampling 14 (+4°C)	2.28	1.39E+08	8.14
OC-Auto Sampling 15 (+4°C)	5.22	1.04E+09	9.02
OC-Auto Sampling 16 (+4°C)	5.07	3.33E+09	9.52
OC-Auto Sampling 17 (+4°C)	3.86	1.65E+09	9.22
OC-Auto Sampling 18 (+4°C)	2.84	1.18E+09	9.07
OC-Auto Sampling 19 (+4°C)	4.04	2.30E+09	9.36
OC-Auto Sampling 01 (+30°C)	1.19	3.42E+08	8.53
OC-Auto Sampling 02 (+30°C)	5.34	3.23E+08	8.51
OC-Auto Sampling 03 (+30°C)	3.10	2.07E+09	9.32
OC-Auto Sampling 04 (+30°C)	2.60	4.16E+08	8.62

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ mL	log(Bacterial quantity/ mL)
OC-Auto Sampling 05 (+30°C)	2.61	1.26E+09	9.10
OC-Auto Sampling 06 (+30°C)	2.78	7.33E+08	8.87
OC-Auto Sampling 07 (+30°C)	5.17	1.67E+09	9.22
OC-Auto Sampling 08 (+30°C)	3.49	2.02E+09	9.31
OC-Auto Sampling 09 (+30°C)	3.74	3.35E+09	9.53
OC-Auto Sampling 10 (+30°C)	2.76	1.35E+09	9.13
OC-Auto Sampling 11 (+30°C)	3.08	1.33E+09	9.12
OC-Auto Sampling 12 (+30°C)	3.65	3.12E+09	9.49
OC-Auto Sampling 13 (+30°C)	3.23	2.27E+09	9.36
OC-Auto Sampling 14 (+30°C)	1.55	1.30E+09	9.11
OC-Auto Sampling 15 (+30°C)	4.11	7.88E+08	8.90
OC-Auto Sampling 16 (+30°C)	3.77	1.94E+09	9.29
OC-Auto Sampling 17 (+30°C)	2.57	9.67E+08	8.99
OC-Auto Sampling 18 (+30°C)	4.69	2.13E+09	9.33
OC-Auto Sampling 19 (+30°C)	5.14	2.21E+09	9.34
OC-Auto Sampling (screening) 01	1.41	2.32E+08	8.37
OC-Auto Sampling (screening) 02	4.26	8.73E+08	8.94
OC-Auto Sampling (screening) 03	4.18	1.47E+09	9.17
OC-Auto Sampling (screening) 04	3.16	2.77E+08	8.44
OC-Auto Sampling (screening) 05	3.02	7.82E+08	8.89
OC-Auto Sampling (screening) 06	2.02	3.72E+08	8.57
OC-Auto Sampling (screening) 07	3.05	5.27E+08	8.72
OC-Auto Sampling (screening) 08	5.17	6.21E+08	8.79
OC-Auto Sampling (screening) 09	2.70	9.03E+08	8.96
OC-Auto Sampling (screening) 10	2.67	6.81E+08	8.83
OC-Auto Sampling (screening) 11	3.32	3.11E+08	8.49
OC-Auto Sampling (screening) 12	4.05	1.51E+09	9.18
OC-Auto Sampling (screening) 13	2.84	6.69E+08	8.83
OC-Auto Sampling (screening) 14	1.17	7.31E+07	7.86
OC-Auto Sampling (screening) 15	1.94	1.84E+08	8.27
OC-Auto Sampling (screening) 16	4.09	1.61E+09	9.21
OC-Auto Sampling (screening) 17	2.55	4.36E+08	8.64
OC-Auto Sampling (screening) 18	2.90	7.79E+08	8.89
OC-Auto Sampling (screening) 19	5.00	1.52E+09	9.18
Hemotrust 01 (-80°C)	2.72	3.39E+09	9.53
Hemotrust 02 (-80°C)	1.93	3.94E+09	9.60
Hemotrust 03 (-80°C)	3.88	4.69E+09	9.67
Hemotrust 04 (-80°C)	5.99	4.87E+09	9.69
Hemotrust 05 (-80°C)	3.15	3.43E+09	9.53
Hemotrust 06 (-80°C)	2.56	5.23E+09	9.72
Hemotrust 07 (-80°C)	2.95	4.82E+09	9.68
Hemotrust 08 (-80°C)	4.45	7.10E+09	9.85
Hemotrust 09 (-80°C)	3.40	6.27E+09	9.80

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ mL	log(Bacterial quantity/ mL)
Hemotrust 10 (-80°C)	3.40	5.95E+09	9.77
Hemotrust 11 (-80°C)	3.31	4.80E+09	9.68
Hemotrust 12 (-80°C)	5.79	1.03E+10	10.01
Hemotrust 13 (-80°C)	2.50	7.15E+09	9.85
Hemotrust 14 (-80°C)	-0.68	4.12E+08	8.61
Hemotrust 15 (-80°C)	2.24	3.32E+09	9.52
Hemotrust 16 (-80°C)	5.49	1.18E+10	10.07
Hemotrust 17 (-80°C)	1.95	3.52E+09	9.55
Hemotrust 18 (-80°C)	3.66	4.75E+09	9.68
Hemotrust 19 (-80°C)	3.97	6.41E+09	9.81
Hemotrust 01 (RT)	1.68	2.02E+09	9.31
Hemotrust 02 (RT)	4.91	5.93E+09	9.77
Hemotrust 03 (RT)	8.42	1.19E+10	10.08
Hemotrust 04 (RT)	2.57	2.19E+09	9.34
Hemotrust 05 (RT)	4.55	2.98E+09	9.47
Hemotrust 06 (RT)	4.69	3.58E+09	9.55
Hemotrust 07 (RT)	3.31	2.60E+09	9.41
Hemotrust 08 (RT)	4.44	3.82E+09	9.58
Hemotrust 09 (RT)	4.89	3.96E+09	9.60
Hemotrust 10 (RT)	2.20	2.67E+08	8.43
Hemotrust 11 (RT)	1.84	2.40E+09	9.38
Hemotrust 12 (RT)	6.29	8.45E+09	9.93
Hemotrust 13 (RT)	3.79	3.55E+09	9.55
Hemotrust 14 (RT)	-0.43	1.43E+08	8.15
Hemotrust 15 (RT)	2.15	1.41E+09	9.15
Hemotrust 16 (RT)	2.74	5.52E+08	8.74
Hemotrust 17 (RT)	2.06	2.23E+09	9.35
Hemotrust 18 (RT)	4.21	3.68E+09	9.57
Hemotrust 19 (RT)	3.94	4.51E+09	9.65
Specimen Collection Container A 01 (-80°C)	1.06	8.44E+08	8.93
Specimen Collection Container A 02 (-80°C)	1.41	7.91E+08	8.90
Specimen Collection Container A 03 (-80°C)	2.19	3.11E+08	8.49
Specimen Collection Container A 04 (-80°C)	1.30	3.87E+08	8.59
Specimen Collection Container A 05 (-80°C)	1.50	4.59E+08	8.66
Specimen Collection Container A 06 (-80°C)	2.13	1.48E+09	9.17
Specimen Collection Container A 07 (-80°C)	1.37	1.06E+09	9.02
Specimen Collection Container A 08 (-80°C)	2.62	1.59E+09	9.20
Specimen Collection Container A 09 (-80°C)	3.35	1.15E+08	8.06
Specimen Collection Container A 10 (-80°C)	0.83	7.25E+08	8.86
Specimen Collection Container A 11 (-80°C)	1.47	7.40E+08	8.87
Specimen Collection Container A 12 (-80°C)	1.86	1.33E+09	9.12
Specimen Collection Container A 13 (-80°C)	1.74	1.43E+09	9.15
Specimen Collection Container A 14 (-80°C)	0.27	6.46E+07	7.81

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ mL	log(Bacterial quantity/ mL)
Specimen Collection Container A 15 (-80°C)	1.72	2.17E+08	8.34
Specimen Collection Container A 16 (-80°C)	2.52	1.91E+09	9.28
Specimen Collection Container A 17 (-80°C)	2.00	1.55E+09	9.19
Specimen Collection Container A 18 (-80°C)	2.08	1.22E+09	9.09
Specimen Collection Container A 19 (-80°C)	1.79	1.05E+09	9.02
Specimen Collection Container A 01 (RT)	1.52	4.29E+08	8.63
Specimen Collection Container A 02 (RT)	1.41	1.53E+08	8.19
Specimen Collection Container A 03 (RT)	2.35	1.12E+09	9.05
Specimen Collection Container A 04 (RT)	0.94	2.22E+08	8.35
Specimen Collection Container A 05 (RT)	3.30	4.68E+08	8.67
Specimen Collection Container A 06 (RT)	0.89	4.66E+08	8.67
Specimen Collection Container A 07 (RT)	2.55	3.68E+08	8.57
Specimen Collection Container A 08 (RT)	1.97	5.64E+08	8.75
Specimen Collection Container A 09 (RT)	1.90	2.46E+08	8.39
Specimen Collection Container A 10 (RT)	0.89	1.85E+08	8.27
Specimen Collection Container A 11 (RT)	1.41	1.13E+08	8.05
Specimen Collection Container A 12 (RT)	2.41	6.92E+08	8.84
Specimen Collection Container A 13 (RT)	1.29	4.80E+08	8.68
Specimen Collection Container A 14 (RT)	1.20	3.13E+07	7.50
Specimen Collection Container A 15 (RT)	1.19	2.22E+08	8.35
Specimen Collection Container A 16 (RT)	2.14	2.97E+08	8.47
Specimen Collection Container A 17 (RT)	1.43	1.28E+08	8.11
Specimen Collection Container A 18 (RT)	1.96	2.67E+08	8.43
Specimen Collection Container A 19 (RT)	1.79	4.64E+08	8.67
One-Step FOB 01 (-80°C)	3.20	4.70E+09	9.67
One-Step FOB 02 (-80°C)	-0.87	1.98E+09	9.30
One-Step FOB 03 (-80°C)	-0.12	1.89E+09	9.28
One-Step FOB 04 (-80°C)	4.58	4.63E+09	9.67
One-Step FOB 05 (-80°C)	4.58	4.33E+09	9.64
One-Step FOB 06 (-80°C)	0.76	5.10E+09	9.71
One-Step FOB 07 (-80°C)	0.94	3.03E+09	9.48
One-Step FOB 08 (-80°C)	2.94	7.46E+09	9.87
One-Step FOB 09 (-80°C)	-1.57	1.34E+09	9.13
One-Step FOB 10 (-80°C)	0.43	3.14E+09	9.50
One-Step FOB 11 (-80°C)	0.30	2.34E+09	9.37
One-Step FOB 12 (-80°C)	3.91	6.93E+09	9.84
One-Step FOB 13 (-80°C)	-1.32	5.25E+09	9.72
One-Step FOB 14 (-80°C)	-5.18	3.00E+08	8.48
One-Step FOB 15 (-80°C)	5.01	4.94E+09	9.69
One-Step FOB 16 (-80°C)	7.92	4.59E+09	9.66
One-Step FOB 17 (-80°C)	-1.81	2.83E+09	9.45
One-Step FOB 18 (-80°C)	4.79	8.12E+09	9.91
One-Step FOB 19 (-80°C)	2.53	6.99E+09	9.84

Sample name	Bacterial DNA concentration (ng/µl)	Bacterial quantity/ mL	log(Bacterial quantity/ mL)
One-Step FOB 01 (RT)	-2.77	7.15E+08	8.85
One-Step FOB 02 (RT)	-1.99	5.87E+08	8.77
One-Step FOB 03 (RT)	1.27	3.19E+09	9.50
One-Step FOB 04 (RT)	-3.43	7.21E+08	8.86
One-Step FOB 05 (RT)	-0.11	1.41E+09	9.15
One-Step FOB 06 (RT)	-0.97	1.16E+09	9.06
One-Step FOB 07 (RT)	0.41	2.03E+09	9.31
One-Step FOB 08 (RT)	5.57	8.00E+09	9.90
One-Step FOB 09 (RT)	-1.82	1.92E+09	9.28
One-Step FOB 10 (RT)	-1.72	3.44E+09	9.54
One-Step FOB 11 (RT)	-3.76	3.66E+08	8.56
One-Step FOB 12 (RT)	0.15	5.94E+09	9.77
One-Step FOB 13 (RT)	2.65	4.30E+09	9.63
One-Step FOB 14 (RT)	-4.11	1.61E+08	8.21
One-Step FOB 15 (RT)	-2.00	2.62E+09	9.42
One-Step FOB 16 (RT)	-1.07	1.83E+09	9.26
One-Step FOB 17 (RT)	0.78	2.35E+09	9.37
One-Step FOB 18 (RT)	3.66	7.04E+09	9.85
One-Step FOB 19 (RT)	4.87	4.46E+09	9.65