The gastrointestinal microbiota in human health and disease

Suk Yee Lam

ISBN: 978-94-6416-718-4
Cover design and layout: Publiss | www.publiss.nl
Printed by: Ridderprint | www.ridderprint.nl

Financial support for printing of this thesis was provided by:

- Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam
- Erasmus University Rotterdam
- ChipSoft.

The Gastrointestinal Microbiota in Human Health and Disease

~

De gastro-intestinale microbiota in ziekte en gezondheid

Thesis

To obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. F.A. van der Duijn Schouten

and in accordance with the decision of the Doctorate Board

The public defense shall be held on

Wednesday 29 September 2021 at 10:30

by

Suk Yee Lam

Born in Amsterdam, The Netherlands

Erasmus University Rotterdam

Erasmus

DOCTORAL COMMITTEE

Promotor:	prof. dr. M.P. Peppelenbosch
Inner Committee:	prof. dr. C.J. van der Woude
	prof. dr. R.K. Weersma
	prof. dr. O. Koren
Copromotor:	dr. G.M. Fuhler

CONTENTS

1

CHAPTER 1

General introduction and outline of this thesis

General introduction

The microbial world of bacteria, viruses, fungi and archaea is not directly apparent to the naked eye, but constitutes a wealth of intrigue and interest for many scientists. While individual (opportunistic) pathogens have shown their capability to cause clinical disease, much is yet to be learned about the microbial communities that reside within our human bodies. Since trillions of microbes are encountered at the protective internal and external epithelial linings on a daily basis, it is of great importance to understand their implications for human health and disease. Culture-independent microbial research has become the method of choice to study the presence of the microbial residents (microbiota) and their genomes (microbiome). The bacteria in particular have gained vast attention from the scientific community as their number within the human body equals the amount of human cells [1]. Many bacterial phyla (>90) have been identified as part of the tree of life [2], but the majority of bacteria colonizing humans are represented by Firmicutes, Bacteroidetes, Actinobacteria and Proteobacteria [3, 4]. Variation in bacterial community structure in healthy subjects is primarily determined by body site (e.g. gut, oral cavity, skin) while strong interpersonal differences exist within the habitats [3, 4]. The microbiome profile of each person is indeed unique, but also influenced by internal (e.g. host genetics) and external (e.g. environment, diet and antibiotics) factors.

The human gastrointestinal (GI) tract is inhabited by the bulk of bacterial residents, especially the gut, which contains the most dense and diverse microbial population. Colonization starts right at birth, where after the GI microbial composition undergoes further changes as it matures over time [5-7]. Since the GI microbiota are involved in multiple important aspects of host physiology (e.g. maturation of the GI tract, modulation of the immune system, energy supply and vitamin production), these microbial residents are considered to function as an exceptional "organ" of the host [8]. While the maintenance of microbial homeostasis is beneficial for human health, the significance of disturbances of GI microbial communities (dysbiosis) in disease warrants further exploration [9]. Both untargeted shotgun metagenomics and targeted 16S rRNA amplicon sequencing have led to more insight into the bacterial microbiota of the GI tract, revealing associations of the GI microbiota with diseases within the GI tract (e.g. colorectal cancer [CRC]) as well as remote organ systems (e.g. autism) that seems more complex [8]. An interplay of microbes with temporal changes in different stages of disease pathogenesis is conceivable, but the precise contribution of individual agents remains to be established and warrants further exploration. Animal models such as rodents are commonly used to study GI microbial dynamics and to provide proof for causality in an experimental set up. Nevertheless, these

models do not mimic the natural setting within humans and therefore the inclusion of human subjects and the collection of preferentially longitudinal biological samples are essential for human microbiome studies.

The investigation of the GI microbiota often relies on the collection of fresh or frozen biopsy and stool samples. Although the mucosal microbiota are thought to be more representative for the local microbiota at near proximity to human epithelial cells, the fecal content is often used as a proxy for the lower GI tract [10], offering an opportunity to study metabolic profiles and to identify new non-invasive biomarkers. Fecal immunochemical tests (FITs) are currently applied in CRC screening programs to detect subjects at risk for (pre)malignant transformation of the colon. These stool-based tests rely on the measurement of occult blood (i.e. hemoglobin), but not all subjects with (pre)malignant lesions are detected [11-13], and the relatively high false positivity rate of this assay means that many patients unnecessarily undergo endoscopic surveillance procedures. The application of the GI microbiome as additional non-invasive biomarker seems promising, but microbiome profiling by high throughput sequencing is not cost-effective and labor intensive for screening purposes. Targeted analysis of metagenomics markers in stool derived samples would be more feasible [14-17], but it is worthwhile to examine the possibility of using FITs of lower biomass instead of whole stool samples in future efforts [18]. The selection of microbial biomarkers remains to be determined for different disease stages and warrants validation in study populations with different ethnic background and in presence of other (GI) comorbidities. Since environmental and genetic host factors are significant contributors to host susceptibility to disease, these factors also require further investigation.

Microbiome studies have enabled the identification of species that were previously not defined, providing a more complete overview of the taxonomic and functional repertoire of the human GI microbiota [8]. We have started to unravel the significance of the GI microbiota within this new scientific era with ongoing technical advances, but their precise contribution in the pathological basis of disease and their utility in medical practice deserves further exploration.

Outline of this thesis

The importance of the GI microbiota in health and disease has gained attention in recent years, and many disease entities are now linked to changes in the gut microbial composition. In Chapter 2, we provide a detailed description of the GI microbiota in human healthy physiology and its emerging role in GI oncogenesis. This review of literature

indicates that different studies together have contributed to our general knowledge regarding the bacterial GI residents in the healthy state, but many of them have focused on a section of the GI tract only. In Chapter 3, we conducted 16S rRNA amplicon sequencing to explore the bacterial composition along nine mucosal sites within the GI tract of 14 subjects using biopsies obtained during double balloon enteroscopy. We demonstrate that both the bacterial load and α-diversity were higher in the lower GI tract, the bacterial communities of which cluster separately from the upper GI tract. While fresh or frozen specimens (e.g. biopsies and feces) are commonly used to answer specific microbe-related health questions, collection of these samples for research involving specific disease sites, rare diseases or a long follow-up time of patients may be difficult. In these situations, tissues from pathology archives may be useful. The application of formalin-fixed paraffin embedded (FFPE) tissues for deoxyribonucleic acid (DNA) isolation remains difficult due to the detrimental effect of tissue fixation procedures on DNA integrity and the inherent low bacterial biomass, but these specimens might be the only source directly available. In Chapter 4, we show the feasibility of conducting bacterial analysis of FFPE sections under specific conditions, but also demonstrate that despite rigorous technical conditions, 16S rRNA amplicon sequencing of this material remains challenging. Our recommendations to achieve quality control and to account for unintentional bacterial contamination might be helpful for future research using low biomass specimens, but other microbiome sources are also of interest. In Chapter 5, we demonstrate that specific bacterial markers can be stably measured up to six days in FITs from the CRC screening program and that bacterial contamination is overall low when analyzed by quantitative polymerase chain reaction (qPCR). Thus, the application of FITs to analyze the GI microbiome for the detection of (pre)malignant lesions is promising. The clinical application of the microbiome also holds promise in radiation enteropathy (RE) as stated by study of Ferreira *et al* [19]. In Chapter 6, we discuss their findings in the context of recent literature and the parallels of RE with inflammatory bowel disease (IBD), a disease group known to have an altered microbial composition [20, 21] and to be at increased risk of developing CRC. Interestingly, human genomic DNA can also be extracted from stool alongside microbial DNA, and utilized for the measurement of methylated levels of specific host loci (i.e. *BMP3* and *VAV3* genes). This may hold promise for the screening of high grade dysplasia and CRC in IBD patients [22], though we discuss the effect of IBD medication on such diagnostic measurements in Chapter 7. Since IBD is associated with several extra-intestinal complaints, including skin diseases psoriasis [23, 24] and hidradenitis suppurativa (HS) [25-27], a role for the GI microbiota in processes beyond the gut is increasingly being speculated upon. While the concept of a gut-skin axis implicating the GI microbiota to skin diseases has been

widely proposed [28-31], we are among the first to characterize the fecal microbiome in HS patients and to demonstrate some interesting bacterial taxonomic alterations in patients compared to healthy controls in Chapter 8. Removal of inflamed intestine through ileocecal resection is a successful treatment strategy for the IBD sub-entity Crohn's disease (CD), but recurrence of disease may still occur. In Chapter 9, we used FFPE tissue resection specimens to explore whether microbial biomarkers might be helpful to predict CD patients at risk of postoperative surgical recurrence. We show that bacterial markers (*Faecalibacterium prausnitzii* and adherent invasive *Escherichia coli*) were not able to distinguish between subjects with and without re-resection in retrospective. Host genetic factors (i.e. single nucleotide variants in the autophagy gene *ATG16L1*) have been shown to affect host bacterial handling by affecting Paneth cell function [32]. Nevertheless, neither histopathological markers for Paneth cell function nor *ATG16L1* status were predictive for disease recurrence and the search for suitable future markers thus continues. In Chapter 10, we turned our attention to the role of host genetic factors in the bacterial handling of *Helicobacter pylori* (*H. pylori*), a known risk factor for gastric carcinogenesis [33, 34]. We investigated the genetic association between *Toll-Like receptor 1* (*TLR1*) locus and anti-*H. pylori* IgG levels, which was not uniformly confirmed across cohorts, and demonstrate that antibody decay and TLR1-independent effects of *H. pylori* may complicate these genetic association studies in different cohorts.

References

1. Sender R, Fuchs S and Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol, 2016. 14(8): p. e1002533.
2. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ *et al*. A new view of the tree of life. Nat Microbiol, 2016. 1: p. 16048.
3. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, and Knight R. Bacterial community variation in human body habitats across space and time. Science, 2009. 326(5960): p. 1694-7.
4. Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. Nature, 2012. 486(7402): p. 207-14.
5. Nagpal R, Mainali R, Ahmadi S, Wang S, Singh R, Kavanagh K *et al*. Gut microbiome and aging: Physiological and mechanistic insights. Nutr Healthy Aging, 2018. 4(4): p. 267-285.
6. Odamaki T, Kato K, Sugahara H, Hashikura N, Takahashi S, Xiao JZ *et al*. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. BMC Microbiol, 2016. 16: p. 90.
7. Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, and Rakoff-Nahoum S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature, 2021.
8. Sekirov I, Russell SL, Antunes LC and Finlay BB. Gut microbiota in health and disease. Physiol Rev, 2010. 90(3): p. 859-904.
9. Scott AJ, Alexander JL, Merrifield CA, Cunningham D, Jobin C, Brown R *et al*. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. Gut, 2019. 68(9): p. 1624-1632.
10. Vaga S, Lee S, Ji B, Andreasson A, Talley NJ, Agréus L *et al*. Compositional and functional differences of the mucosal microbiota along the intestine of healthy individuals. Sci Rep, 2020. 10(1): p. 14977.
11. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á *et al*. Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. N Engl J Med, 2012. 366(8): p. 697-706.
12. van Doorn SC, Stegeman I, Stroobants AK, Mundt MW, de Wijkerslooth TR, Fockens P *et al*. Fecal immunochemical testing results and characteristics of colonic lesions. Endoscopy, 2015. 47(11): p. 1011-7.
13. Katsoula A, Paschos P, Haidich AB, Tsapas A and Giouleme O. Diagnostic Accuracy of Fecal Immunochemical Test in Patients at Increased Risk for Colorectal Cancer: A Meta-analysis. JAMA Intern Med, 2017. 177(8): p. 1110-1118.
14. 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(1): p. 70-78.
15. Wong SH, Kwong TNY, Chow TC, Luk AKC, Dai RZW, Nakatsu G *et al*. Quantitation of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. Gut, 2017. 66(8): p. 1441-1448.
16. Liang Q, Chiu J, Chen Y, Huang Y, Higashimori A, Fang J *et al*. Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. Clin Cancer Res, 2017. 23(8): p. 2061-2070.
17. Liang JQ, Li T, Nakatsu G, Chen YX, Yau TO, Chu E *et al*. A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut, 2020. 69(7): p. 1248-1257.
18. Baxter NT, Koumpouras CC, Rogers MA, Ruffin MTt and Schloss PD. DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. Microbiome, 2016. 4(1): p. 59.

19. Reis Ferreira M, Andreyev HJN, Mohammed K, Truelove L, Gowan SM, Li J *et al*. Microbiota- and Radiotherapy-Induced Gastrointestinal Side-Effects (MARS) Study: A Large Pilot Study of the Microbiome in Acute and Late-Radiation Enteropathy. Clin Cancer Res, 2019. 25(21): p. 6487-6500.
20. Kostic AD, Xavier RJ and Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology, 2014. 146(6): p. 1489-99.
21. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM *et al*. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol, 2017. 2: p. 17004.
22. Kisiel JB, Klepp P, Allawi HT, Taylor WR, Giakoumopoulos M, Sander T *et al*. Analysis of DNA Methylation at Specific Loci in Stool Samples Detects Colorectal Cancer and High-Grade Dysplasia in Patients With Inflammatory Bowel Disease. Clin Gastroenterol Hepatol, 2019. 17(5): p. 914-921 e5.
23. Eppinga H, Poortinga S, Thio HB, Nijsten TEC, Nuij V, van der Woude CJ *et al*. Prevalence and Phenotype of Concurrent Psoriasis and Inflammatory Bowel Disease. Inflamm Bowel Dis, 2017. 23(10): p. 1783-1789.
24. Yang BR, Choi NK, Kim MS, Chun J, Joo SH, Kim H *et al*. Prevalence of extraintestinal manifestations in Korean inflammatory bowel disease patients. PLoS One, 2018. 13(7): p. e0200363.
25. van der Zee HH, de Winter K, van der Woude CJ and Prens EP. The prevalence of hidradenitis suppurativa in 1093 patients with inflammatory bowel disease. Br J Dermatol, 2014. 171(3): p. 673-5.
26. Shalom G, Freud T, Ben Yakov G, Khoury R, Dreiher J, Vardy DA *et al*. Hidradenitis Suppurativa and Inflammatory Bowel Disease: A Cross-Sectional Study of 3,207 Patients. J Invest Dermatol, 2016. 136(8): p. 1716-1718.
27. Deckers IE, Benhadou F, Koldijk MJ, Del Marmol V, Horváth B, Boer J *et al*. Inflammatory bowel disease is associated with hidradenitis suppurativa: Results from a multicenter cross-sectional study. J Am Acad Dermatol, 2017. 76(1): p. 49-53.
28. O'Neill CA, Monteleone G, McLaughlin JT and Paus R. The gut-skin axis in health and disease: A paradigm with therapeutic implications. Bioessays, 2016. 38(11): p. 1167-1176.
29. Vaughn AR NM, Clark AK, Sivamani RK. Skin-gut axis: The relationship between intestinal bacteria and skin health. World J Dermatol 2017; 6(4): 52-58.
30. Salem I, Ramser A, Isham N and Ghannoum MA. The Gut Microbiome as a Major Regulator of the Gut-Skin Axis. Front Microbiol, 2018. 9: p. 1459.
31. Ellis SR, Nguyen M, Vaughn AR, Notay M, Burney WA, Sandhu S *et al*. The Skin and Gut Microbiome and Its Role in Common Dermatologic Conditions. Microorganisms, 2019. 7(11).
32. Deuring JJ, Fuhler GM, Konstantinov SR, Peppelenbosch MP, Kuipers EJ, de Haar C *et al*. Genomic ATG16L1 risk allele-restricted Paneth cell ER stress in quiescent Crohn's disease. Gut, 2014. 63(7): p. 1081-91.
33. WHO IAfRoCIS, liver flukes and Helicobacter pylori. IARC working group on the evaluation of carcinogenic risks to humans. Lyon, 7-14 June 1994. IARC Monogr Eval Carcinog risks Hum 1994;61:177e240.
34. de Martel C, Georges D, Bray F, Ferlay J and Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob Health, 2020. 8(2): p. e180-e190.

2

CHAPTER 2

The gastrointestinal microbiota and its role in oncogenesis

S.Y. Lam[1], J. Yu[2], S.H. Wong[2], G.M. Fuhler[1], M.P. Peppelenbosch[1]

[1] Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands; [2] Department of Medicine and Therapeutics, Institute of Digestive Disease, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences and CUHK-Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong.

Best Pract Res Clin Gastroenterol. 2017 Dec;31(6):607-618.

Abstract

Advances in research techniques have made it possible to map the microbial communities in the gastrointestinal (GI) tract, where the majority of bacteria in the human body reside. Disturbances in these communities are referred to as dysbiosis and have been associated with GI cancers. Although dysbiosis is observed in several GI malignancies, the specific role of these changes has not been understood to the extent of *Helicobacter pylori* (HP) in gastric cancer. This review will address the bacterial communities along the GI tract, from the oral cavity to the anal canal, particularly focusing on bacterial dysbiosis and carcinogenesis. Just as non-HP bacteria in the stomach may interact with HP in gastric carcinogenesis, the same may hold true for other GI tract malignancies, where an interplay between microbes in carcinogenesis seems conceivable, especially in colorectal cancer (CRC). In the last part of this review, we will discuss the potential mechanisms of bacterial dysbiosis in GI carcinogenesis.

Keywords:

Microbiota; microbiome; cancer; gastrointestinal; oral; esophageal; gastric; colorectal.

1 Introduction

One of the first people to directly observe bacteria was Antoni van Leeuwenhoek, who in 1683 described the presence of single cell organisms living in the human oral cavity, which he called 'animalcules'. Nevertheless, it was not until the 19th century that the theory that bacteria could cause disease was commonly accepted. Today, we are starting to recognize the complex relationships between microbes and their hosts in health and disease. Historically, the identification and classification of microorganisms was based on microscopic and culture-based methods, and the advent of molecular technologies has contributed significantly to the emerging insight into the microbes that collectively reside in a given ecosystem (the microbiota) and their genomes (the microbiome) [1, 2]. The human microbiome is mostly found at the interface between our body and the outside world – i.e. our skin, mucosa and in particular, the gastrointestinal (GI) tract, which can also be represented as one large complex microbial ecosystem. This tract, simplified as a hollow tube system from the oral cavity to the anal canal, is openly connected to the outside world and the epithelial layer of the GI tract therefore acts as an important barrier function to keep microorganisms from invading. Since a major part of the 10^{13} bacteria in the human body resides in the alimentary tract [3], continuous exposure to these microbiota is inevitable. Owing to the uneven distribution of the bacterial load, different parts of the GI tract are exposed to different amounts and types of microorganisms. Bacterial counts drop from the oral cavity to the acidic stomach and then markedly increase from the small intestines to reach a maximum in the colon. As the latter harbors a quantity that exceeds other organs by at least two orders of magnitude, the bacterial content in the colon has been a preferred focus for study [3, 4]. Nevertheless, there has been an exponential interest in unravelling the microbial content of the entire GI tract. Disturbances in this highly complex system (termed dysbiosis) can affect human health [5-7] and have been associated with different GI diseases, including infectious diarrhea, inflammatory bowel disease (IBD) and cancer.

Given the extensive research and data availability, this review will focus mainly on bacteria as a major component of the GI tract microbiota. The healthy human microbiome, sampled across different body parts, is mainly (>90%) represented by the phyla Actinobacteria, Firmicutes, Proteobacteria and Bacteroidetes [8]. Since each body habitat harbors its own characteristic microbiota, the relative abundances of these phyla and their sublevels (class, order, family, genus and species) are likely to vary across different GI sites in healthy and diseased states (**Figure 1**). In this review, we will focus on the interaction between the microbiota and GI carcinogenesis, focusing on site-specific microbial changes and summarizing what is known about the molecular pathways involved.

2 The microbiome of the upper GI tract

Much effort has been directed towards the identification of causative microbial agents in carcinogenesis. The global burden of new cancer cases attributed to infectious agents was estimated to be 15.4% (2.2 million cases) in 2012 [9]. The leading contributor to this list was *Helicobacter pylori* (HP), with a total of 770,000 cancer cases that year. HP was involved in 29% of the gastric carcinomas of the cardia, and was accountable for 89% of the non-cardia gastric carcinomas and 74% of the non-Hodgkin lymphomas of the stomach. The involvement of HP in gastric cancer (GC) is the most notorious example of microbial infection related cancer within the human GI tract. Already in 1994, this Gram-negative flagellated bacterium was classified as a group I carcinogen by the International Agency of Research on Cancer (IARC) [10]. In contrast, the microbial contribution to other GI tract malignancies has not been fully understood. Since the compartments of the alimentary tract are all virtually interconnected to one another, it is important to uncover site specific microbes involved in carcinogenesis.

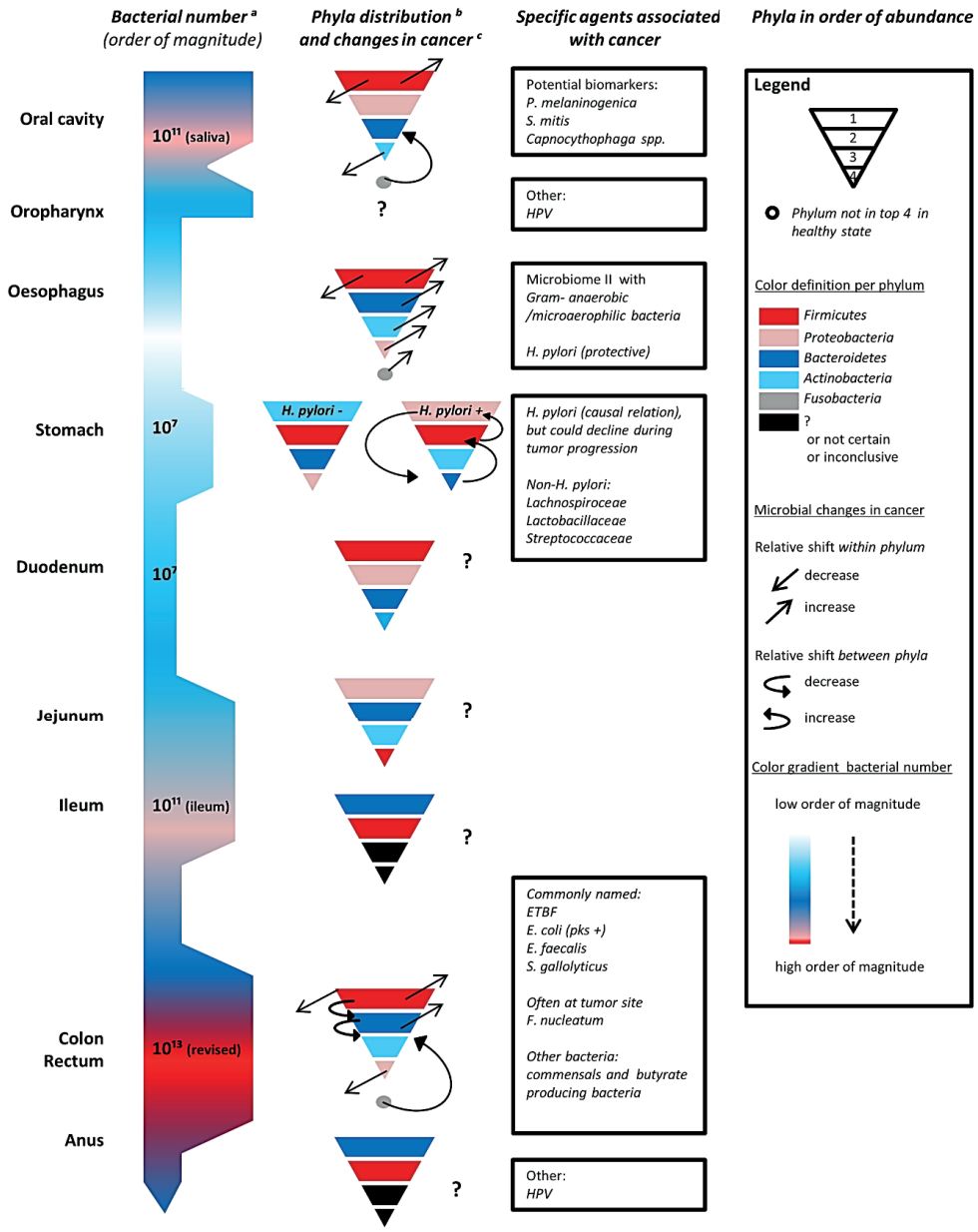
2.1 Oral cavity

The oral cavity is the starting point of the GI tract and is lined by mucosa that covers the lips and the mouth. Despite the possibility of direct visualization, cancer affecting this site is often recognized at late stages [11]. Oral squamous cell carcinoma (OSCC) is the most common type and a major cause of morbidity and mortality [11, 12]. Tobacco smoking remains an important contributor besides other factors such as alcohol and beta nut exposure. Despite these, the etiological picture is not fully elucidated, and recently the attention has been shifted to the identification of potential microbial agents [30]. A minor role (4.3%) for *human papillomavirus* (HPV) infection in oral cancer was acknowledged earlier [9], but studies have not agreed on the role of specific bacteria, individually or collectively, in OSCC [31].

►Figure 1 | The human gastrointestinal (GI) microbiome: composition in healthy and GI cancer.

The bacterial numbers in the different parts of the GI tract; adopted from [3].
The phylum composition in healthy state; adopted from [5, 13-17].
Changes within and between phyla in GI tract cancers; adopted from [18-29].

General abbreviations: +, positive; -, negative. Microbial abbreviations*: E. coli, Escherichia coli; E. faecalis, Enterococcus faecalis; F. nucleatum, Fusobacterium nucleatum; ETBF, enterotoxigenic Bacteroides fragilis; H. pylori, Helicobacter pylori; HPV, human papillomavirus; P. melaninogenica, Prevotella melaninogenica; S. gallolyticus, Streptococcus gallolyticus; S. mitis, Streptococcus mitis.*


Bacterial number [a]
(order of magnitude)
Phyla distribution [b]
and changes in cancer [c]
Specific agents associated
with cancer
Phyla in order of abundance
Oral cavity
Oropharynx
Oesophagus
Stomach
Duodenum
Jejunum
Ileum
Colon
Rectum
Anus
10^11 (saliva)
10^7
10^7
10^11 (ileum)
10^13 (revised)
H. pylori -
H. pylori +
?
Potential biomarkers:
P. melaninogenica
S. mitis
Capnocythophaga spp.
Other:
HPV
Microbiome II with
Gram- anaerobic
/microaerophilic bacteria
H. pylori (protective)
H. pylori (causal relation),
but could decline during
tumor progression
Non-H. pylori:
Lachnospiroceae
Lactobacillaceae
Streptococcaceae
Commonly named:
ETBF
E. coli (pks +)
E. faecalis
S. gallolyticus
Often at tumor site
F. nucleatum
Other bacteria:
commensals and butyrate
producing bacteria
Other:
HPV
Legend
1
2
3
4
Phylum not in top 4 in
healthy state
Color definition per phylum
Firmicutes
Proteobacteria
Bacteroidetes
Actinobacteria
Fusobacteria
?
or not certain
or inconclusive
Microbial changes in cancer
Relative shift within phylum
decrease
increase
Relative shift between phyla
decrease
increase
Color gradient bacterial number
low order of magnitude
high order of magnitude

Two consecutive papers addressed the presence of living bacteria in OSCC tissue, and showed an increased presence of saccharolytic and aciduric bacteria in OSCC as compared to normal tissue of the same patient. This suggests the presence of specialized microbes attracted by the acidic and hypoxic tumor environment [32, 33], and could point to a consequence of tumorigenesis rather than a driving factor for these bacteria. Differences in bacterial composition of tumor versus normal adjacent tissue included an increased abundance of phylum Firmicutes (85%), and a relative shift of Gram-negative to Gram-positive microbiota, including saccharolytic *Streptococcus* [34]. Using next-generation sequencing (NGS), a later study managed to classify almost all reads (99.6%) from three OSCC biopsies to species level (228 species). Thirty-five species were shared among the OSCC subjects studied, including potential pathogens *Fusobacterium spp.*, *Aggregatibacter segnis* and *Prevotella oris* (*P. oris*). Interestingly, a small group of non-oral taxa (5%), including *Bacteroides fragilis (B. fragilis)*, was found [19].

Meanwhile, non-invasive microbial profiling of saliva and oral swabs has been of great interest for the purpose of OSCC diagnostics. One study explored forty common oral microorganisms in saliva and suggested *Prevotella melaninogenica* (*P. melaninogenica*), *Capnocytophaga gingivalis* (*C. gingivalis*) and *Streptococcus mitis* (*S. mitis*) as potential biomarkers, with a promising sensitivity and specificity of ≥80% [35]. To pursue a more complete picture of bacterial saliva, Pulshalkar *et al.* conducted a 454-sequencing study which revealed Firmicutes as the most prevalent of 8 phyla in OSCC subjects. In total, 15 phylotypes were found to be unique for OSCC which included the above-named *P. melaninogenica* and also *Capnocytophaga spp.* [20]. Similarly, salivary changes were also noticed in a high throughput study investigating oral leukoplakia (OLK), OSCC and controls. While Firmicutes was again the dominant phylum, unlike biopsies, it showed a lower prevalence in saliva from OSCC patients [36]. A significant decrease of Firmicutes (genus *Streptococcus*), as well as Actinobacteria (genus *Rothia*), was also found in oral swabs from both cancerous and precursor oral lesions when compared to the non-affected contralateral sites, although not in comparison to healthy controls [37]. However, a recent 16S rRNA gene sequencing study did show significant differences between the oral microbiota from swabs taken from patients presenting with different stages of malignancy and normal controls. Interestingly, the microbial community of potentially malignant disorders overlapped and was positioned between those of oral cancer and healthy controls, suggesting a gradual shift of microbiome during carcinogenesis. One of the microbes isolated from specifically from precancerous lesions was *Megasphaera micronuciformis*, which together with other bacteria that were in higher abundance in

these lesions compared to either the normal (*P. melaninogenica*, *Prevotella veroralis*) or the cancerous sites (*Rothia mucilaginosa*) might serve as potential biomarker [18].

Taken together, while relatively few studies have examined the microbiome associated with oral cancer, they have indicated shifts in microbial diversity. However, methodological differences between these studies make it difficult to draw definite conclusions about the bacterial association in OSCC [31, 37], as the sampling sites (i.e. mucosa vs fluid) can affect outcomes. While Firmicutes is undoubtedly one of the most abundant phyla present in the mouth, both up- and down-regulation of this phylum in OSCC samples have been observed. It remains undetermined whether the shift within the bacterial community contributes to OSCC carcinogenesis or reflects the changed microenvironment.

2.2 Oropharyngeal cavity and esophagus

The next section of the GI tract is the oropharyngeal cavity, which includes the tonsils. As for other head and neck cancers, tobacco and alcohol are important risk factors in oropharyngeal cancers. Nevertheless, HPV infections are highly associated with this malignancy [38, 39] and have been estimated to contribute to 30% of the oropharyngeal cancer cases worldwide [9]. On the other hand, bacterial associations with oropharyngeal malignancies are so far not evident. HP has shown the ability to colonize the oropharyngeal tissue, but its role in carcinogenesis could not be confirmed [40].

Another understudied area is the microbial structure of the esophagus, especially the bacterial community [41]. While HPV has been linked to Barrett's esophagus (BE) and esophageal cancer [42, 43], bacterial involvement in esophageal disease has not been fully determined. Nevertheless, interesting findings have been observed in a limited number of non-culture based studies that have been performed [43, 44]. The normal esophagus harbors a complex bacterial community and has shown to be predominated by the phylum Firmicutes, followed by Bacteroidetes, Actinobacteria, Proteobacteria and Fusobacteria which appeared in decreasing order [45]. Supervised and phenotype directed analyses have demonstrated two types of microbiomes present in the distal esophagus [21]. Type I was more closely associated with normal esophagus and was predominated by genus *Streptococcus,* while type II, in which Gram-negative anaerobic/ microaerophilic bacteria were more abundant, corresponded to esophagitis and Barrett's esophagitis, suggesting the presence of microbiota shifts in disease states. Highly abundant genera of the type II microbiome included *Veillonella* and *Granulicatella* (Firmicutes), *Prevotella* and *Porphyromonas* (Bacteroidetes), *Haemophilus, Neisseria*

and *Campylobacter* (Proteobacteria), *Rothia* and *Actinomyces* (Actinobacteria) and *Fusobacterium* (Fusobacteria).

When focusing on cancer of the esophagus, approximately 90% of the cases worldwide consist of esophageal squamous-cell carcinoma (ESSC), although in North America and Europe, esophageal adenocarcinomas (EAC) is the predominant type [46]. The role of local bacteria in esophageal carcinogenesis is disputed [44, 47], but it has been hypothesized that the type II microbiome contributes to gastro-esophageal reflux disease (GERD) and also EAC, potentially by stimulating the immune system and triggering inflammation-associated carcinogenesis [21]. However, while some characteristics of the type II microbiome have been confirmed in other studies, the exact microbial signature in GERD and EAC remains to be reproduced. One relatively large study combined a cultured based approach with molecular analysis in GERD, BE, EAC and controls and found that the most prominent shift was reflected by increased abundance of the genus *Campylobacter* (species *Campylobacter concisus*) in GERD and BE [48]. A more recent prospective BE cohort study showed that the majority of sequencing reads obtained from the esophageal site belonged to the phyla Firmicutes (genus *Streptococcus*) and Bacteroidetes (genus *Prevotella*), with a small proportion of reads (5%) representing HP [22].

While BE poses a substantial risk factor for esophageal cancer, it has also been speculated that HP infection is protective against both BE [49, 50] and esophageal cancer [50]. This would be consistent with the fact that the presence of HP in BE is associated with lower aneuploidy rates [22]. A meta-analysis of 27 studies confirmed this inverse association between HP and EAC risk, although not for ESCC [51]. Since heterogeneous findings were previously found for BE, this association has been re-evaluated in case-control studies [52, 53]. HP infection was inversely related to BE, but particularly in case of corpus atrophy or regular anti-secretory medication use [52] and cytotoxin-associated gene A (cagA) positive strains [53]. In addition to BE, an inverse correlation with HP was also seen for erosive esophagitis although not for GERD symptoms [53], and HP eradication does not increase GERD risk [54]. Thus, while a role for HP in GERD is less clear, most evidence points towards a protective role for HP in BE and esophageal cancer, potentially by neuro-immunological anti-inflammatory mechanisms [55].

2.3 Stomach

2.3.1 The gastric microbiota composition

Despite its potential protective role in the esophagus, HP is undoubtedly a carcinogen in

gastric oncogenesis and, since its discovery in 1982, has been the most studied bacterium isolated from the stomach [56]. While HP may thrive in the stomach, the relatively hostile conditions in the stomach generally correspond to a lower colonization as compared to other GI sites; with only an estimated abundance of 10^7 bacteria, the stomach is several orders lower in bacterial load than more distal parts of the GI tract. Over time, new molecular techniques have made it possible to identify the remaining members of the gastric community and have overcome the restrictions of conventional culture based methods [57-60]. Emerging studies have pointed out that HP is by no means the only predominant species in the stomach, and the role of these non-Helicobacter species in gastric diseases is under investigation [61-63]. Using traditional culture combined with 16S rRNA pyrosequencing, examination of the gastric microbiota of healthy HP negative individuals showed the presence of 69 genera, 59 families and 9 phyla of which Firmicutes (genera *Streptococcus*, *Lactobacillus* and *Enterococcus*), Proteobacteria and Actinobacteria (genus *Propionibacterium*) were the most abundant in descending order. Other phyla (<3%) included Deinococcus-Thermus, Bacteroidetes and Gemmatimonadetes. Cultivation of gastric juice and mucosa did not only pick up members of these common genera, but also indicated that these bacteria were alive. However, positive cultures were less frequent in gastric juice as compared to gastric mucosa [59], and it has been suggested that the microbial composition at the mucosa would be more representative of the physiology at the epithelium [64]. A high prevalence (65%) of non-Helicobacter microorganisms was also found in a HP positive Chinese cohort with dyspeptic complaints, identifying 18 additional genera and 43 species of which *Neisseria, Streptococcus, Rothia* and *Staphylococcus* were most prevalent [65].

While it is becoming clear that in addition to HP, many other bacteria are able to thrive in the gastric environment, the question remains whether the presence of HP directly shapes the gastric microbial composition. Bik *et al*. was one of the first to confirm the microbial diversity of the stomach (with main phyla Firmicutes [genus *Streptococcus*], Proteobacteria [main genus *Helicobacter*] and Actinobacteria in descending abundance) and showed that the presence of HP did not influence the gastric community distribution [66]. Similarly, a Malaysian study showed that HP did not influence the microbial diversity in patients with divergent gastric diseases [67]. In contrast, a study in healthy persons identified the same major phyla, but did find that HP status was important in making a difference in microbial diversity. In this study, HP negative samples contained a more diverse community with 262 phylotypes belonging to 13 phyla whereas HP positive samples harbored only 33 phylotypes and were dominated by HP [68]. Similarly, Maldonado-Contreras *et al.*

demonstrated a change in microbial composition related to HP status. While similar phyla were detected among HP positive and negative subjects, with phyla Proteobacteria, Firmicutes, Actinobacteria in descending order, HP positivity was associated with a relative increase of abundance within the phyla Proteobacteria, Spirochetes and Acidobacteria and a decline in Actinobacteria, Bacteroidetes and Firmicutes. However, this study also pointed out that although the presence of HP attributed to the composition of the gastric microbiota, origins or ethnicity might contribute even more [69].

Thus, in addition to HP, the four main GI phyla are well represented in the stomach with *Streptococcus* being one of the most often identified genera. Whether or not HP infection affects the composition of the remaining microbiota remains unclear, but may be linked to more to disease state and ethnicity than the presence of HP *per se*.

2.3.2 Microbiota in gastric cancer (GC)

GC is globally the fifth most common cancer and was the third leading cause of cancer death in 2012 [70]. HP has been acknowledged as an important risk factor in the development of GC, in particular non-cardia cancer [71]. Although only a small percentage of HP infected subjects develop GC [72], HP eradication has been shown to reduce the odds of GC incidence [73]. While a direct causative role for HP in gastric carcinogenesis is undisputed [74], the role of the remaining microbiota in the oncogenic process is unclear. Interestingly, spontaneous gastritis and gastrointestinal intra-epithelial neoplasia (GIN) development was significantly retarded in an insulin-gastrin (INS-GAS) mouse model under germ free conditions, demonstrating that commensal bacteria do contribute to gastric carcinogenesis. The authors further suggested that enteric bacterial overgrowth during gastritis contributed to HP-induced carcinogenesis [75]. Indeed, co-infection of INS-GAS mice with restricted Altered Schaedler Flora (rASF; consisting of ASF356 *Clostridium* species, ASF361 *Lactobacillus murinus* and ASF519 Bacteroidetes species) was sufficient to enhance GIN development in an inflammation-dependent manner [76].

Whether non-HP members in the gastric microbiota play a role in gastric carcinogenesis in human remains uncertain. Dicksved *et al*., when comparing GC patients to dyspeptic controls, did not find significant differences in gastric microbiota between these groups using terminal restriction fragment length polymorphism (T-RFLP). Subsequent *16S rRNA* gene analysis of GC samples demonstrated a highly complex community of 102 phylotypes, which was clustered into five major phyla consisting of Firmicutes (most represented), Bacteroidetes, Actinobacteria, Proteobacteria and Fusobacteria. HP was only seen in a low abundance in this study [23]. However, HP incidence is higher in Asia compared to

Western countries, and a Korean study using 16S rRNA pyrosequencing detected HP DNA in 92% of paraffin-embedded tissues from resected gastric adenocarcinomas. In non-HP GC patients, *Helicobacter cinaedi, Helicobacter mustelae* and *Campylobacter hyointestinalis* were detected [77]. Nevertheless, no significant functional role for bacteria other than HP seems to be apparent in GC [78].

However, some studies do suggest that bacterial shifts may occur during progression of gastric disease. In 1992, the Correa model of gastric carcinogenesis illustrated the sequential stages of chronic gastritis, atrophy, intestinal metaplasia (IM) and dysplasia prior to development GC of the intestinal type [79]. During this progression, a gradual shift of gastric microbial community from precancerous lesions to GC appears to take place, although different studies disagree on the nature of these shifts. Both decreases [80] and increases [81, 82] in bacterial diversity have been reported when comparing samples from chronic gastritis and/or IM towards GC. The most consistent findings between these studies were an increase in Lachnospiraceae, and members of the Lactobacillaceae and Streptococcacaea families, during carcinogenesis. The presence of HP appears to enhance microbial differences between chronic gastritis, IM and GC [81] and bacterial load [82]. Furthermore, a recent study showed an inversed correlation between HP abundance and bacterial diversity in chronic gastritis and IM patients, which was corrected upon eradication of HP [83]. We have recently identified differences in bacterial interactions across stages of gastric carcinogenesis. The significant enrichments and network centralities suggest the potentially important roles of *Peptostreptococcus stomatis (P. stomatis), Dialister pneumosintes, Slackia exigua, Parvimonas micra (P. micra) and Streptococcus anginosuss* in GC progression. Moreover, stronger interactions among gastric microbes were observed in HP-negative samples compared to HP-positive samples in superficial gastritis and IM [84].

In addition to comparing the gastric microbiomes from GC and non-GC patients, the microbial profile of the cancerous stomach and the adjacent normal mucosa within the same patient has also been addressed. Paradoxically, HP levels in HP positive GC patients were significantly reduced in malignant tissue compared to normal mucosa [85]. This might seem counterintuitive, as physical interaction of the gastric epithelium with HP is one of the driving factors in gastric carcinogenesis. However, it is conceivable that the altered local microenvironment upon oncogenesis is more permissive of other bacteria rather than HP (as detailed later) [86]. At genus level, a decline was also seen for *Propionibacterium sp., Staphylococcus sp.* and *Corynebacterium sp.*, whereas *Clostridium sp*. and *Prevotella sp*. were increased in GC mucosa [85].

Overall, there is emerging evidence of the presence of non-Helicobacter bacteria in the human stomach besides HP, but the specific role of individual microorganisms in human gastric carcinogenesis remains unclear [82, 87-89]. The gastric microbiota composition of GC patients seems to be affected in some studies, but contradictory outcomes have been reported by others. Again, the different methodological approaches make it difficult to directly compare studies, and more homogeneous studies are required to validate the results obtained so far [90].

2.4 Duodenum

The bacterial residents of the duodenum have not been thoroughly investigated despite some evidence for their involvement in upper GI disorders (e.g. irritable bowel syndrome (IBS) and coeliac disease) [60]. Molecular studies are especially limited, but evidence so far indicates that phyla Firmicutes, Proteobacteria and Actinobacteria are predominately part of the healthy microbiome [13, 91, 92]. When descending to lower taxonomy levels, distinct compositions were noticed for duodenal biopsies (genera *Acinetobacter, Bacteroides, Prevotella*) versus luminal contents (genera *Prevotella, Stenotrophomonas, Streptococcus*) [13], with mucosa-associated microbes appearing less variable and more conserved.

Investigation of bacterial profile shifts might lead to the better understanding of the pathophysiology in upper GI disorders. For instance, in duodenal IBS, both bacterial overgrowth and a decline in microbial diversity have been noticed and might be relevant for future treatments [91]. However, duodenal malignancies are relatively rare and the relevance of bacterial involvement in duodenal cancers has not been investigated yet.

3 The microbiome of the middle and lower GI tract

3.1 Jejunum & ileum

Although the jejunum and ileum are the longest GI parts with the largest contact surface area, collection of samples from these sites remains challenging [60]. Nevertheless, molecular analysis of *16S rRNA* genes demonstrated that the jejunum harbors a less complex bacterial community compared to the distal ileum and large intestines [15]. A small study using a single healthy volunteer showed predominance of the genus *Streptococcus*, but a better impression was given by a more recent NGS study which included jejunal biopsies from nineteen healthy subjects [14]. The phyla in descending order of relative abundance consisted of Proteobacteria (*Neisseria, Helicobacter*), Bacteroidetes (*Prevotella*), Actinobacteria (*Brevibacterium*), Firmicutes (*Streptococcus*) and Fusobacteria.

Like the collection of biopsies, selective luminal collections are usually not easy in the small intestinal lumen of subjects with intact anatomy. However, a post-mortem study managed to demonstrate the presence of mainly facultative anaerobes and aerobes in the luminal content of the jejunum and ileum [93]. Although there were inter-individual differences, genera belonging to the phyla Firmicutes (*Streptococcus, Lactobacillus, Enterococcus*), Proteobacteria (*Gammaproteobacteria*) and Bacteroidetes (*Bacteroides*) were observed. The presence of a highly inter-individual bacterial community was also confirmed in ileostomy subjects [16]. Nonetheless, knowledge about the microbial community of the small intestines is still limited in health and disease, and the role of bacterial dysbiosis in cancers of the small intestines remains to be elucidated.

3.2 Large intestines

3.2.1 Colorectal microbiota composition

The colon harbors one of the most dense and diverse bacterial communities of the human body and consist of both luminal and surface-associated populations [94, 95]. This complexity is depicted by the increasing numbers of new phylotypes identified in recent studies [94, 96-98]. There is little variation within the same subject, but high inter-individual diversities exist in both colonic tissue [94, 98] and fecal [8, 94, 96, 99] samples. Although mucosal and fecal microbiomes differ significantly in lower taxonomy classes, characterization at phylum level indicated that members of the mucosa associated and fecal microbiota consist mainly of Firmicutes (cluster Clostridia) and Bacteroidetes [14, 15, 94, 97, 99-101]. Less abundant phyla in the colorectal compartment include Proteobacteria, Actinobacteria, Fusobacteria and Verrucomicrobia.

3.2.2 Microbiota in colorectal cancer (CRC) – drivers and passengers?

According to the multi-hit model of CRC, sequential mutations are needed to drive colorectal carcinogenesis [102, 103]. Like the Correa sequence defined in GC, an adenoma-to-carcinoma sequence with accumulating mutations is apparent for CRC development. Triggers of these events are not defined yet [104], but accumulating clues have linked bacterial agents to CRC development. A well-established example is the association of *Streptococcus gallolyticus* (*S. gallolyticus*; formerly *Streptococcus bovis*) with CRC, which justifies endoscopic bowel examination in case of positive blood cultures [105]. Since emerging insight into CRC pathogenesis has tremendous health benefits, it is not surprising that the residents of the colon are extensively explored. CRC remains a major health burden as the second and third most common cancer in respectively women and

men. With an estimated 1.3 million cases, worldwide and mortality rates over 690.000 in 2012 [70], it is critical to unravel the role of bacterial dysbiosis in sporadic CRC in hope to improve clinical diagnostics and possibly treatment.

Similar to what was observed in the stomach, early signs of dysbiosis were detected in adenoma [28, 101] and CRC, with more pronounced changes in the latter [28]. Despite high inter-individual differences in the microbiota, shifts in mucosal bacterial patterns have been observed in CRC patients compared to healthy individuals, including the increased presence of *Fusobacterium* [25, 106, 107], *Enterococcus faecalis* (*E. faecalis*) [106] and recently *Peptostreptococcus anaerobius* [108] in colorectal tumors. Other commonly reported agents are genotoxic *Escherichia coli* (*E. coli*) and *enterotoxigenic Bacteroides fragilis* (ETBF) [109-111] which are also able to confer pro-inflammatory signaling in the colon. The association of such bacteria with CRC [112, 113] has led to the 'alpha bug' hypothesis. This model states that certain pro-oncogenic bacteria remodel the colonic bacterial community to enhance mucosal immune responses and epithelial changes leading to CRC [104]. The alpha bugs in this case do not act alone, but remodel the colonic microbiota to induce a more oncogenic ecosystem [114].

Other studies have also investigated the microbiome of tumor and normal adjacent mucosa from CRC patients. Although the over- and under-representation of community members between these sites seem to vary among cohorts, the most consisting finding has been the enrichment *Fusobacterium* (in particular *F. nucleatum*, a common oral commensal [115]) at CRC sites [25, 27, 116], which in high relative abundance has been associated with regional lymph node metastases [116] and tumor location (2% in rectum towards 11% in cecum) [117]. However, other studies have shown that in addition to the tumor site, the microbial constitution of normal mucosa from adenoma patients is also altered [118, 119]. The fact that commensal and anti-inflammatory butyrate-producing bacteria (family Coriobacteriaceae, genera *Fusobacterium, Roseburia, Faecalibacterium, Lactococcus*) are also often found at the tumor site [26] whereas pathogenic members of Proteobacteria (*Pseudomonas*, family Enterobacteriaceae with *Escherichia-Shigella, Citrobacter* and *Salmonella spp.*) are seen in the normal adjacent tissue [25, 26], has led to the postulation of the 'driver-passenger' theory. In this model, 'bacterial drivers' are held responsible for the initiation of carcinogenesis by causing epithelial DNA damage. The subsequent niche changes allow relatively poorly colonizers with competitive advantage in the tumor environment, the 'bacterial passengers', to outgrow the bacterial drivers of CRC, which are consequently enriched in the surrounding mucosa. The 'passengers' themselves can either promote or suppress the carcinogenic process, and therefore may represent a

consequence of disease rather than a causative factor. Temporal changes of both bacterial drivers and passenger could be present due to continuous tumor development [2]. This model could also explain the previously mentioned observation that HP might gradually disappear at later stages of GC [85]. Obviously, temporal changes in bacterial presence 'on' and 'off' the tumor would greatly complicate comparisons of different cohorts, and may explain the fact that despite intense scientific efforts, it has been difficult to get a clear consensus on the 'CRC-associated microbiome'.

In addition to bacteria acting as potential oncogenes, it is also conceivable that the commensal microbiome acts as a tumor-suppressive factor, and that disturbance therefore results in the accumulation of mutagenic events. Feces from patients with malignant lesions show a decrease in beneficial bacteria such as Lachnospiraceae [120] and it was recently demonstrated that fecal bacteria (in particular Actinobacteria; genera *Streptomyces* and *Rhodococcus*) from healthy subjects showed potent *in vitro* and *in vivo* anti-tumor effects [121]. Interestingly, this effect was greater for younger subjects, and it is tempting to speculate that changes in microbiome occurring at older age account for some extent for the increased CRC risk at older age. It is of interest to note that in inflammatory conditions such as ulcerative colitis, associated with altered microbiome characterized by a reduced presence of beneficial bacteria [122-125], is also associated with an increased risk of CRC development. This is commonly attributed to the chronic inflammatory process, but could theoretically also be driven by the loss of commensal anti-tumor effects.

3.2.3 Microbial biomarkers for CRC

In recent years, considerable interest has emerged to develop affordable non-invasive biomarkers for early CRC diagnosis. Both guaiac fecal occult blood tests (gFOBT) and fecal immunochemical tests (FIT) have been used to screen individual stool specimens for the presence of hemoglobin. However, not all pre-malignant lesions bleed, and blood can sometimes also be detected in stool from healthy individuals. Thus, there is clinical need for additional biomarkers for CRC, for which the use of the gut microbiome seems promising [126, 127]. Several metagenome studies have recently been conducted to identify stool based bacterial biomarkers. Zeller *et al.* demonstrated marker species *F. nucleatum subspecies vincentii* and *animalis, P. stomatis,* and *Porphyromonas asaccharolytica,* that together discriminated between CRC and controls with the same accuracy as FOBT [127]. Furthermore, a select microbial panel was also able to discriminate healthy subjects from CRC or adenomas in a study by Zackular *et al* [126]. Among the six CRC-specific operational

taxonomic units (OTUs), genus *Fusobacterium* was enriched in addition to both family/ genus Porphyromonadaceae and *Porphyromonas*.

Two consecutive Chinese studies took the next steps in developing clinical affordable fecal biomarkers. Two potential CRC markers (*F. nucleatum* and *P. micra*) were validated in cohorts of different ethnicity using quantitative polymerase chain reaction (qPCR) [128]. Diagnostic utility was confirmed for a combination of four bacterial markers (*F. nucleatum, Clostridium hathewayi, Bacteroides clarus* and undefined species *'m7'*), and shown to be more accurate than *F. nucleatum* alone [129]. The addition of microbiome profiling to FIT testing seems to lead to the improvement of CRC detection [120, 129, 130], thus validation of these results in other cohorts would be of great use to improve CRC detection rates worldwide. Interestingly, these analyses would be feasible in FIT rest material, as bacterial DNA isolated from FIT tests largely resembles that from feces [131]. However, as it has been shown that the fecal microbiome from adenoma patients differs little from healthy controls (except for the levels of *Ruminococcus*), alternative markers for this group of patients is called for [127].

Concluding, it is safe to say that in addition to the role of HP in GC, the CRC associated microbiome is the best studied interaction between bacteria and GI cancers to date. While mucosal bacteria arguably exert larger influences on the epithelial barrier cells than luminal bacteria due to their direct contact with the host cells, the effect of luminal bacterial metabolites on the epithelial barrier should not be disregarded [132]. Regardless of the differences between fecal and mucosal microbiome, a recent systematic review investigating the role of bacteria in CRC agrees on the overexpression of specific bacteria such as Fusobacteria (*F. nucleatum*) and *E. coli* in CRC [24]. Other bacterial changes validated across several studies include increases in *S. gallolyticus, E. faecalis, B. fragilis, Enterobacter, Leptotrichia, Porphyromonas asaccharolytica, Lactococcus* [25, 28, 29, 109-111, 120]. In contrast, commensal (e.g. *Bifidobacteria, Lactobaccillus, Ruminococcus*) and butyrate producing bacteria (e.g. *Lachnospiraceae* and *Faecalibacterium spp.*) [24, 29, 110, 120] appear to be decreased. Thus, despite the fact that the fecal microbiome does not adequately reflect the mucosal microbiome [107], fecal microbiome profiling may be useful for identification of CRC-associated microbiome risk factors. However, it appears that combining classical FIT tests with microbial profiles results in the highest clinical performance of such assays.

3.3 Anus

The last part of the alimentary tract is the anal canal, where malignant lesions are less common. Globally, approximately ninety percent of anal cancer cases are attributed to HPV infections [9] whereas the role of potential bacterial agents has been understudied. Microbial profiles of the anal compartment are scarce, but a recent study in healthy female adolescents showed that the anal microbiome was dominated by genera *Prevotella* and *Bacteroides* of phylum Bacteroidetes [17]. In a different cohort consisting of MSM (men who have sex with men) subjects, the microbes were mainly distributed in phyla Firmicutes and Bacteroidetes whereas the composition altered during uncontrolled advanced human immunodeficiency virus (HIV) infections with reduced and increased members in respectively phylum Firmicutes and Fusobacteria [133]. However, whether relevant bacterial shifts are indicated in anal malignancies is undetermined so far.

4 Potential mechanism of microbiota in GI tract malignancies

While it is evident that dysbiosis within the niches of the alimentary tract are associated with GI malignancies, the exact identity of potential causative members of these complex bacterial communities and their oncogenic mechanisms are not fully understood. However, lessons learned from the much studied interaction of HP with gastric oncogenesis may provide clues for the oncogenic role of the microbiome in CRC, and the main hypotheses regarding microbial-driven oncogenic routes in GC and CRC will be briefly discussed here.

4.1 Potential microbial mechanism in gastric carcinogenesis

HP exhibits several mechanisms through which it exerts its role in gastric oncogenesis, which have been extensively reviewed elsewhere [74]. While ongoing research shows increasing layers of complexity in the HP-induced oncogenic transformation, the mechanisms through which HP exerts its effect can be broadly divided into virulence factor-induced oncogenic changes, inflammation-mediated carcinogenesis and gastric hormone-driven induction of oncogenic signaling. HP possesses various virulence factors, including vacuolating cytotoxin A (VacA) and cagA, which are injected into host cells through a pilus-like structure. Virulence factors of various strengths are found in different HP strains, which vary globally and may contribute to different global incidences of GC. The main effect of these virulence factors is to activate oncogenic signal transduction pathways (including the mitogenic Ras-extracellular signal regulated kinase (ERK) pathway, the phosphatidyl-inositol 3-kinase (PI3K) cell survival pathway, and CDC42/

Rac-focal adhesion kinase cytoskeletal rearrangement pathways), the combined effects of which are cellular changes in morphology, cell cycle, proliferation and cell death [134-136]. The second common mechanism of oncologic transformation induced by HP is through activation of inflammatory pathways. HP induces the production of reactive oxygen species (ROS) in gastric epithelial cells, which activates inflammatory signaling pathways inside these cells, resulting in cellular stress responses. In addition, massive amounts of reactive oxygen and nitrogen species (NOS) are produced by immune cells attracted by HP infection, and the resulting oxidative stress and nitration of DNA bases drives accumulation of DNA damage and oncogenic mutations. Thirdly, initial HP infection causes atrophic gastritis, characterized by destruction of acid producing parietal cells. A compensatory upregulation of the acid-inducing hormone gastrin ensues, which is directly correlated to gastric inflammation levels. In addition to modulation of acid levels, gastrin is known to directly activate oncogenic signaling, including the earlier mentioned PI3K pathway, in gastric epithelium cells.

In addition to HP, factors involved in GC include host genetic risk factors and environmental factors such as smoking, high salt consumption, and other gastric microbes [113] that interact with HP in a complex manner [112, 137]. In context of the hypothetic models in CRC, HP can be seen as the 'alpha bug' or 'bacterial driver' in the initial phases of GC development. HP is thought to be mainly involved at the stage where chronic atrophic gastritis is induced while subsequent microbial changes could lead to IM and GC progression [88]. This has been supported by the model of crosstalk in which co-colonization of HP with other microbiota could lead to atrophy related pH and nutritional changes and innate immune responses, which then allows bacterial overgrowth of microbes that initially could not thrive in the stomach [137]. Consequently, these microbes might induce further inflammatory responses and epithelial damage and overshadow the role of HP at this point. Although these culprits are not fully identified, general mechanistic routes may be present.

Bacterial overgrowth as a result of reduced gastric acidity has been linked the production of potent carcinogenic N-nitroso-compounds (NOCs) [76, 137, 138]. Endogenous exposure to these substances is associated with an increased risk of non-cardia cancer [139]. The substrate for NOC formation is nitrite, which itself is formed by reduction of nitrate by bacteria. [138, 140]. Elevated nitrite and bacterial levels are measured in hypochlorhydric gastric juice [141, 142] which is in line with the finding of higher nitrite and nitrosamine concentrations in atrophic gastritis compared to normal stomachs and non-atrophic gastritis [143]. Furthermore, GC subjects have significantly elevated gastric nitrite concentration compared to atrophic gastritis subjects with similar pH and HP status [144].

These findings are supported by the doubled numbers of nitrosating and nitrate reducing bacteria in GC subjects, albeit not significant compared to controls [78], and also the enrichment of phylum Nitrospirae in GC subjects in another report [82]. More supportive evidence comes from surgically treated patients in which bacterial gene functions prior to treatment indicated the significant prevalence of N-nitrosation genes, and a shift towards genes for nitric oxide (NO) reductase, nitrous oxide (N_2O) reductase and bile salt hydrolase afterwards [145]. Therefore, the above findings indicate the potential of microbes to influence GC development via NOCs.

Toll-like receptors (TLRs) expressed on immune and epithelial cells play an essential role in the recognition of invading pathogens by activating inflammatory signaling, and inducing pro-inflammatory cytokine as well as ROS and NOS production. As such, they contribute to gastro-intestinal integrity, but possibly also to HP associated malignant transformation when chronically activated [146]. During HP infection, pathogen-associated molecular pattern molecules (PAMPs) derived from this bacterium are recognized by TLR2, TLR4, TLR5, TLR9 [147], which are upregulated in dysplasia/carcinoma as compared to healthy mucosa [148, 149]. An even more striking finding has been the loss of polarized TLR distribution in IM, dysplasia and GC [148-150], suggesting that this diffuse localization could make lesions more susceptible for TLR activation by PAMPS [149] and therefore enhanced exposure to inflammation-mediated oncogenic signaling. It has been suggested that other microbes than HP possibly fulfill a more prominent role during malignant transformation by TLR activation, although other pattern-recognition receptors (PRRs) are also thought to be involved (see [147] for excellent review).

Overall, while the mechanistic role of HP in gastric carcinogenesis is well-described, the interaction between HP and the gastric community in GC seems complex and well-defined mechanisms of non-HP microbes have not been defined yet. The question still remains whether HP facilitates the outgrowth of bacteria with carcinogenic profile, or whether other microbiota could promote a more malignant HP as well [86, 137].

4.2 Potential microbial mechanism in CRC carcinogenesis

Just as virulence factors are of considerable importance in HP induced gastric carcinogenesis, the release of toxic compounds by intestinal microbes also promotes malignant transformation of colonic epithelial cells. A familiar example is the bacterial product from 'alpha bug' ETBF, *B. fragilis* toxin (BFT), which not only triggers a cascade of known CRC oncogenic signaling pathways in colonic epithelial cell, (including the ERK

pathway, but also Wnt, NF-B and STAT3) [104, 151], but possibly also causes ROS production [152] and direct DNA damage [104]. Colibactin is another potent genotoxin that is synthetized by pathogenic *E. coli* strains which harbor the genomic island *pks* (*pks+ E. coli*). Epithelial cells that encounter this strain not only suffer from DNA double-strand breaks by colibactin [153], but may also go into cellular senescence, which characterized by growth factor production that might promote proliferation of other non-infected cells [154].

Another microbe-specific way to provoke malignant transformation is by direct interactions with the colonic epithelial cells. This has been described for *F. nucleatum* which was not only enriched at the tumor site, but also showed signs of invasive behavior [27, 116]. Attachment and invasion of this bacterium into epithelial cells requires surface molecule Fusobacterium adhesion A (FadA) to bind the cellular adhesion molecule E-cadherin. Binding to E-cadherin is sufficient to activate Wnt and other oncogenes, but internalization of the bacterium is essential for the activation of inflammatory genes [155]. The elevated *FadA* gene expression was confirmed in tissues from adenomas and CRC, which is in line with the enrichment of FadA in fecal metagenomes of CRC subjects [127].

As suggested before for colitis associated cancers, inflammation mediated bacterial participation is also conceivable in sporadic cancers. Functional analysis of the fecal microbiome has demonstrated a CRC-associated increase in lipopolysaccharides (LPS) metabolism [127], implying that increased LPS-induced TLR-mediated signaling pathways in epithelial cells may contribute to disease. Indeed, both colon adenomas and cancers express higher levels of TLRs (TLR2, 4, 5) that are diffusely and homogenously spread throughout the cell compared to normal mucosa, including ectopic cytoplasmic expression. An intensified activation of TLR is therefore imaginable and has been reflected by higher expression levels of pro-inflammatory mediator cyclooxygenase (COX) 2 [156] and increased TLR mediated pro-inflammatory NF-B pathway activation [157]. Substrates for TLR activation can derive from various bacterial products of a wide range of bacteria (LPS, PAMPs) as well as damage associated molecular patterns (DAMPs). The contribution of TLR signaling to CRC development is underscored by the fact that CRC risk is significantly affected by functional single nucleotide polymorphisms (SNPs) in TLR2 and TLR4 genes [158].

While the focus has been on potential pathogenic members of the CRC microbiome, it has been indicated that the depletion of protective bacteria might similarly promote oncogenesis [126]. Short chain fatty acids (SCFAs), which have been suggested to have potent anti-inflammatory and anti-tumor effects, are the end product of anaerobic bacterial fermentation of dietary fibers [132, 159]. These metabolites consist of acetate,

propionate and butyrate of which the latter is the major energy source of colonic cells [160]. A higher abundance of butyrate-producing bacteria was found in stools of native Africans with low CRC risk as compared to Afro-American subjects with a higher risk. The microbe *Faecalibacterium prausnitzii* is one of the major butyrate producers together with others of the *Clostridium* cluster IV and XIVa [161], and depletion of the *Faecalibacterium* genus increases CRC risk [29]. Similarly, butyrate producer Lachnospiraceae was reduced in feces of CRC subjects [120]. Therefore, one can imagine that a decline of butyrate and other SCFA producing bacteria could contribute to tumor progression. Interestingly, some potential pathogens (e.g. *Fusobacterium*) are also capable of producing butyrate via different butyrogenic pathways. However, these bacteria use amino acids as substrates for butyrate production (glutarate, 4-aminobutyrate, Lysine), which is accompanied by the formation of ammonia which is harmful to the gut [160].

Since the colonic lumen contains the substrates for bacterial metabolic activity, dietary intake could influence CRC risk. This has been illustrated by higher concentrations of both proteolytic fermentation products and secondary bile acids in African Americans with a relative protein rich and high fat diet in comparison to native Africans that have lower CRC risk [161]. Bacterial products from protein fermentation are thought to have cancer promoting effects and include NOCs and ammonia, hydrogen sulfide and also polyamines [132]. The latter toxic product is one of the degradation products of the putrefaction pathway which was shown to be enriched in CRC associated microbiota [127]. Secondary bile acids are also considered carcinogenic and are formed from primary bile acids that escapes the enterohepatic circulation [132]. Thus, it has been suggested that the metabolic profile induced by the entire microbiome is potentially more relevant to the cancer process than individual pathogenic agents.

In short, there are several ways for microbes to participate in CRC oncogenesis. Beside the release of carcinogenic toxic factors and direct host interaction by specific microbes, other microbes might exert their effects via inflammation and metabolic mediated mechanisms. However, polymicrobial biofilms have recently also been associated with CRC initiation and development, and potentially lead to severe inflammation and a more an aggressive tumor. These 'higher-order spatial structures of bacteria' are able to impair epithelial barrier function, affect cellular proliferation, enhance pro-inflammatory and pro-oncogenic responses and enhance intestinal dysbiosis [162] and as such may present the next focus of the scientific community in the search for microbial causative agents in CRC.

5 Practice points

- Culture-independent microbial research techniques have made it possible to map the residents along the GI tract (microbiota) and their genomes (the microbiome).
- In upper, middle and lower gastrointestinal (GI) tract malignancies, a shift in microbial community (dysbiosis) seems be involved in oncogenesis and reflects this disease state.
- Helicobacter pylori (HP) appears to protect against esophageal cancer, but increases gastric cancer (GC) risk, where non-HP residents may also engage in gastric carcinogenesis.
- In addition to genetic and environmental risk factors, microbial content may contribute to colorectal cancer (CRC), since polymicrobial interplay with 'driver' and 'passenger' bacteria has been indicated in tumor progression.
- The use of the GI microbiome as non-invasive diagnostic biomarkers seems promising, especially in oral (swabs) and colorectal cancers (feces).
- The bacterial mechanisms in GI tract malignancies have not been fully understood, but inflammation mediated pathways and bacterial metabolite associated shifts seems important.

6 Research agenda

- The different methodological approaches make it difficult to compare microbiome studies, suggesting the need for more homogeneous approaches.
- Potential microbial biomarkers in GI malignancies seem promising, but should be validated in different ethnic cohorts to test for general applicability.
- Future microbiome studies are indicated to identify subjects with increased GI cancer risk, to predict disease outcome and to develop potential treatments based on microbial profiles.

7 Summary

The advent of culture-independent methodologies in microbial research has led to increased insight into the members (microbiota) of the gastrointestinal (GI) tract and their genomes (microbiome) in both health and disease. The healthy GI microbiome varies considerably among subjects, but the common bacterial members predominately belong to four major phyla (Firmicutes, Bacteroidetes, Proteobacteria, Actinobacteria). However, disruption of the microbial community (dysbiosis) is seen in GI tract cancers, and are best

described for oral squamous cell carcinomas (OSCC), gastric cancer (GC) and colorectal cancer (CRC) in comparison to other parts of the GI tract (oropharynx, small intestines, anus). In the esophagus, microbiome type II is thought to be involved in carcinogenesis, whereas *Helicobacter pylori* (HP) may play a protective role. In contrast, a causative function of HP in GC is evident, whereas the relative contribution of non-HP in this disease is still under investigation. A polymicrobial interplay has also been suggested in CRC cancer, illustrated by two models, the 'alpha bug' and 'driver-passenger' hypotheses. Both theories center around the involvement of bacteria in early stages of CRC oncogenesis, where the 'alpha bug' theory suggests a cooperation with other bacteria to create a more hostile tumor environment whereas the bacterial 'drivers' are thought to be gradually replaced by 'passengers' with advantages in the new niche. Whether dysbiosis in GI cancer is truly a cause of disease or mainly the effect of an altered microenvironment remains in many cases disputed, as the specific role of these microbes is not fully understood. Nevertheless, inflammation and metabolite mediated pathways are likely to affect the epithelial lining and may contribute to initiation and progression of cancer. Future studies might help to explore the applicability of the microbiome in diagnostics, prognostics and potentially prevention and treatment.

References

1. Lynch SV and Pedersen O. The Human Intestinal Microbiome in Health and Disease. N Engl J Med, 2016. 375(24): p. 2369-2379.
2. Tjalsma H, Boleij A, Marchesi JR and Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol, 2012. 10(8): p. 575-82.
3. Sender R, Fuchs S and Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol, 2016. 14(8): p. e1002533.
4. Sender R, Fuchs S and Milo R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. Cell, 2016. 164(3): p. 337-40.
5. Cho I and Blaser MJ. The human microbiome: at the interface of health and disease. Nat Rev Genet, 2012. 13(4): p. 260-70.
6. Sheflin AM, Whitney AK and Weir TL. Cancer-promoting effects of microbial dysbiosis. Curr Oncol Rep, 2014. 16(10): p. 406.
7. Vogtmann E and Goedert JJ. Epidemiologic studies of the human microbiome and cancer. Br J Cancer, 2016. 114(3): p. 237-42.
8. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, and Knight R. Bacterial community variation in human body habitats across space and time. Science, 2009. 326(5960): p. 1694-7.
9. Plummer M, de Martel C, Vignat J, Ferlay J, Bray F, and Franceschi S. Global burden of cancers attributable to infections in 2012: a synthetic analysis. Lancet Glob Health, 2016. 4(9): p. e609-16.
10. WHO, international agency for research of cancer (IARC). Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. IARC Monogr Eval Carcinog Risks Hum, 1994. 61: p. 177-240.
11. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol, 2009. 45(4-5): p. 309-16.
12. Chen YK, Huang HC, Lin LM and Lin CC. Primary oral squamous cell carcinoma: an analysis of 703 cases in southern Taiwan. Oral Oncol, 1999. 35(2): p. 173-9.
13. Li G, Yang M, Zhou K, Zhang L, Tian L, Lv S *et al*. Diversity of Duodenal and Rectal Microbiota in Biopsy Tissues and Luminal Contents in Healthy Volunteers. J Microbiol Biotechnol, 2015. 25(7): p. 1136-45.
14. Chung CS, Chang PF, Liao CH, Lee TH, Chen Y, Lee YC *et al*. Differences of microbiota in small bowel and faeces between irritable bowel syndrome patients and healthy subjects. Scand J Gastroenterol, 2016. 51(4): p. 410-9.
15. Wang M, Ahrne S, Jeppsson B and Molin G. Comparison of bacterial diversity along the human intestinal tract by direct cloning and sequencing of 16S rRNA genes. FEMS Microbiol Ecol, 2005. 54(2): p. 219-31.
16. Booijink CC, El-Aidy S, Rajilic-Stojanovic M, Heilig HG, Troost FJ, Smidt H *et al*. High temporal and inter-individual variation detected in the human ileal microbiota. Environ Microbiol, 2010. 12(12): p. 3213-27.
17. Smith BC, Zolnik CP, Usyk M, Chen Z, Kaiser K, Nucci-Sack A *et al*. Distinct Ecological Niche of Anal, Oral, and Cervical Mucosal Microbiomes in Adolescent Women. Yale J Biol Med, 2016. 89(3): p. 277-284.
18. 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): p. 1-15.
19. Al-Hebshi NN, Nasher AT, Idris AM and Chen T. Robust species taxonomy assignment algorithm for 16S rRNA NGS reads: application to oral carcinoma samples. J Oral Microbiol, 2015. 7: p. 28934.

20. Pushalkar S, Mane SP, Ji X, Li Y, Evans C, Crasta OR *et al*. Microbial diversity in saliva of oral squamous cell carcinoma. FEMS Immunol Med Microbiol, 2011. 61(3): p. 269-77.
21. Yang L, Lu X, Nossa CW, Francois F, Peek RM, and Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. Gastroenterology, 2009. 137(2): p. 588-97.
22. Gall A, Fero J, McCoy C, Claywell BC, Sanchez CA, Blount PL *et al*. Bacterial Composition of the Human Upper Gastrointestinal Tract Microbiome Is Dynamic and Associated with Genomic Instability in a Barrett's Esophagus Cohort. PLoS One, 2015. 10(6): p. e0129055.
23. Dicksved J, Lindberg M, Rosenquist M, Enroth H, Jansson JK, and Engstrand L. Molecular characterization of the stomach microbiota in patients with gastric cancer and in controls. J Med Microbiol, 2009. 58(Pt 4): p. 509-16.
24. Borges-Canha M, Portela-Cidade JP, Dinis-Ribeiro M, Leite-Moreira AF and Pimentel-Nunes P. Role of colonic microbiota in colorectal carcinogenesis: a systematic review. Rev Esp Enferm Dig, 2015. 107(11): p. 659-71.
25. Gao Z, Guo B, Gao R, Zhu Q and Qin H. Microbiota disbiosis is associated with colorectal cancer. Front Microbiol, 2015. 6: p. 20.
26. Marchesi JR, Dutilh BE, Hall N, Peters WH, Roelofs R, Boleij A *et al*. Towards the human colorectal cancer microbiome. PLoS One, 2011. 6(5): p. e20447.
27. Kostic AD, Gevers D, Pedamallu CS, Michaud M, Duke F, Earl AM *et al*. Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. Genome Res, 2012. 22(2): p. 292-8.
28. Nakatsu G, Li X, Zhou H, Sheng J, Wong SH, Wu WK *et al*. Gut mucosal microbiome across stages of colorectal carcinogenesis. Nat Commun, 2015. 6: p. 8727.
29. Chen W, Liu F, Ling Z, Tong X and Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One, 2012. 7(6): p. e39743.
30. Chocolatewala N, Chaturvedi P and Desale R. The role of bacteria in oral cancer. Indian J Med Paediatr Oncol, 2010. 31(4): p. 126-31.
31. Perera M, Al-Hebshi NN, Speicher DJ, Perera I and Johnson NW. Emerging role of bacteria in oral carcinogenesis: a review with special reference to perio-pathogenic bacteria. J Oral Microbiol, 2016. 8: p. 32762.
32. Hooper SJ, Crean SJ, Fardy MJ, Lewis MA, Spratt DA, Wade WG *et al*. A molecular analysis of the bacteria present within oral squamous cell carcinoma. J Med Microbiol, 2007. 56(Pt 12): p. 1651-9.
33. Hooper SJ, Crean SJ, Lewis MA, Spratt DA, Wade WG, and Wilson MJ. Viable bacteria present within oral squamous cell carcinoma tissue. J Clin Microbiol, 2006. 44(5): p. 1719-25.
34. Pushalkar S, Ji X, Li Y, Estilo C, Yegnanarayana R, Singh B *et al*. Comparison of oral microbiota in tumor and non-tumor tissues of patients with oral squamous cell carcinoma. BMC Microbiol, 2012. 12: p. 144.
35. Mager DL, Haffajee AD, Devlin PM, Norris CM, Posner MR, and Goodson JM. The salivary microbiota as a diagnostic indicator of oral cancer: a descriptive, non-randomized study of cancer-free and oral squamous cell carcinoma subjects. J Transl Med, 2005. 3: p. 27.
36. Hu X, Zhang Q, Hua H and Chen F. Changes in the salivary microbiota of oral leukoplakia and oral cancer. Oral Oncol, 2016. 56: p. e6-8.
37. Schmidt BL, Kuczynski J, Bhattacharya A, Huey B, Corby PM, Queiroz EL *et al*. Changes in abundance of oral microbiota associated with oral cancer. PLoS One, 2014. 9(6): p. e98741.
38. D'Souza G, Kreimer AR, Viscidi R, Pawlita M, Fakhry C, Koch WM *et al*. Case-control study of human papillomavirus and oropharyngeal cancer. N Engl J Med, 2007. 356(19): p. 1944-56.

39. Walvik L, Svensson AB, Friborg J and Lajer CB. The association between human papillomavirus and oropharyngeal squamous cell Carcinoma: Reviewed according to the Bradford Hill criteria for causality. Oral Oncol, 2016. 63: p. 61-65.
40. Lukes P, Pavlik E, Potuznikova B, Nartova E, Foltynova E, Plzak J *et al*. Detection of Helicobacter pylori in oropharyngeal lymphatic tissue with real-time PCR and assessment of its carcinogenic potential. Eur Arch Otorhinolaryngol, 2014. 271(2): p. 399-405.
41. van Baal JW, Diks SH, Wanders RJ, Rygiel AM, Milano F, Joore J *et al*. Comparison of kinome profiles of Barrett's esophagus with normal squamous esophagus and normal gastric cardia. Cancer Res, 2006. 66(24): p. 11605-12.
42. Ding GC, Ren JL, Chang FB, Li JL, Yuan L, Song X *et al*. Human papillomavirus DNA and P16(INK4A) expression in concurrent esophageal and gastric cardia cancers. World J Gastroenterol, 2010. 16(46): p. 5901-6.
43. Rajendra S, Wang B, Snow ET, Sharma P, Pavey D, Merrett N *et al*. Transcriptionally active human papillomavirus is strongly associated with Barrett's dysplasia and esophageal adenocarcinoma. Am J Gastroenterol, 2013. 108(7): p. 1082-93.
44. Yang L, Chaudhary N, Baghdadi J and Pei Z. Microbiome in reflux disorders and esophageal adenocarcinoma. Cancer J, 2014. 20(3): p. 207-10.
45. Pei Z, Bini EJ, Yang L, Zhou M, Francois F, and Blaser MJ. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci U S A, 2004. 101(12): p. 4250-5.
46. Rustgi AK and El-Serag HB. Esophageal carcinoma. N Engl J Med, 2014. 371(26): p. 2499-509.
47. Baghdadi J, Chaudhary N, Pei Z and Yang L. Microbiome, innate immunity, and esophageal adenocarcinoma. Clin Lab Med, 2014. 34(4): p. 721-32.
48. Blackett KL, Siddhi SS, Cleary S, Steed H, Miller MH, Macfarlane S *et al*. Oesophageal bacterial biofilm changes in gastro-oesophageal reflux disease, Barrett's and oesophageal carcinoma: association or causality? Aliment Pharmacol Ther, 2013. 37(11): p. 1084-92.
49. Gashi Z, Sherifi F and Shabani R. The prevalence of helicobacter pylori infection in patients with reflux esophagitis - our experience. Med Arch, 2013. 67(6): p. 402-4.
50. Anderson LA, Murphy SJ, Johnston BT, Watson RG, Ferguson HR, Bamford KB *et al*. Relationship between Helicobacter pylori infection and gastric atrophy and the stages of the oesophageal inflammation, metaplasia, adenocarcinoma sequence: results from the FINBAR case-control study. Gut, 2008. 57(6): p. 734-9.
51. Xie FJ, Zhang YP, Zheng QQ, Jin HC, Wang FL, Chen M *et al*. Helicobacter pylori infection and esophageal cancer risk: an updated meta-analysis. World J Gastroenterol, 2013. 19(36): p. 6098-107.
52. Fischbach LA, Graham DY, Kramer JR, Rugge M, Verstovsek G, Parente P *et al*. Association between Helicobacter pylori and Barrett's esophagus: a case-control study. Am J Gastroenterol, 2014. 109(3): p. 357-68.
53. Rubenstein JH, Inadomi JM, Scheiman J, Schoenfeld P, Appelman H, Zhang M *et al*. Association between Helicobacter pylori and Barrett's esophagus, erosive esophagitis, and gastroesophageal reflux symptoms. Clin Gastroenterol Hepatol, 2014. 12(2): p. 239-45.
54. Yaghoobi M, Farrokhyar F, Yuan Y and Hunt RH. Is there an increased risk of GERD after Helicobacter pylori eradication?: a meta-analysis. Am J Gastroenterol, 2010. 105(5): p. 1007-13; quiz 1006, 1014.
55. Shahabi S, Rasmi Y, Jazani NH and Hassan ZM. Protective effects of Helicobacter pylori against gastroesophageal reflux disease may be due to a neuroimmunological anti-inflammatory mechanism. Immunol Cell Biol, 2008. 86(2): p. 175-8.

56. Warren J. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet, 1983. 1(8336): p. 1273-5.
57. Sanduleanu S, Jonkers D, De Bruine A, Hameeteman W and Stockbrügger RW. Non-Helicobacter pylori bacterial flora during acid-suppressive therapy: differential findings in gastric juice and gastric mucosa. Alimentary Pharmacology & Therapeutics, 2001. 15(3): p. 379-388.
58. Vartoukian SR, Palmer RM and Wade WG. Strategies for culture of 'unculturable' bacteria. FEMS Microbiol Lett, 2010. 309(1): p. 1-7.
59. Delgado S, Cabrera-Rubio R, Mira A, Suarez A and Mayo B. Microbiological survey of the human gastric ecosystem using culturing and pyrosequencing methods. Microb Ecol, 2013. 65(3): p. 763-72.
60. Wang ZK and Yang YS. Upper gastrointestinal microbiota and digestive diseases. World J Gastroenterol, 2013. 19(10): p. 1541-50.
61. Engstrand L and Lindberg M. Helicobacter pylori and the gastric microbiota. Best Pract Res Clin Gastroenterol, 2013. 27(1): p. 39-45.
62. Ianiro G, Molina-Infante J and Gasbarrini A. Gastric Microbiota. Helicobacter, 2015. 20 Suppl 1: p. 68-71.
63. Walker MM and Talley NJ. Review article: bacteria and pathogenesis of disease in the upper gastrointestinal tract--beyond the era of Helicobacter pylori. Aliment Pharmacol Ther, 2014. 39(8): p. 767-79.
64. Sung J, Kim N, Kim J, Jo HJ, Park JH, Nam RH *et al*. Comparison of Gastric Microbiota Between Gastric Juice and Mucosa by Next Generation Sequencing Method. J Cancer Prev, 2016. 21(1): p. 60-5.
65. Hu Y, He LH, Xiao D, Liu GD, Gu YX, Tao XX *et al*. Bacterial flora concurrent with Helicobacter pylori in the stomach of patients with upper gastrointestinal diseases. World J Gastroenterol, 2012. 18(11): p. 1257-61.
66. Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F *et al*. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A, 2006. 103(3): p. 732-7.
67. Khosravi Y, Dieye Y, Poh BH, Ng CG, Loke MF, Goh KL *et al*. Culturable bacterial microbiota of the stomach of Helicobacter pylori positive and negative gastric disease patients. ScientificWorldJournal, 2014. 2014: p. 610421.
68. Andersson AF, Lindberg M, Jakobsson H, Backhed F, Nyren P, and Engstrand L. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One, 2008. 3(7): p. e2836.
69. Maldonado-Contreras A, Goldfarb KC, Godoy-Vitorino F, Karaoz U, Contreras M, Blaser MJ *et al*. Structure of the human gastric bacterial community in relation to Helicobacter pylori status. Isme J, 2011. 5(4): p. 574-9.
70. Ferlay J; Soerjomataram I; Ervik M; Dikshit R; Eser S; Mathers C; Rebelo M; Parkin DM; Forman D; Bray F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. 2013; Available from: http://globocan.iarc.fr.
71. Helicobacter and Cancer Collaborative Group. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut, 2001. 49(3): p. 347-53.
72. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M *et al*. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med, 2001. 345(11): p. 784-9.
73. Ma JL, Zhang L, Brown LM, Li JY, Shen L, Pan KF *et al*. Fifteen-year effects of Helicobacter pylori, garlic, and vitamin treatments on gastric cancer incidence and mortality. J Natl Cancer Inst, 2012. 104(6): p. 488-92.
74. Polk DB and Peek RM, Jr. Helicobacter pylori: gastric cancer and beyond. Nat Rev Cancer, 2010. 10(6): p. 403-14.

75. Lofgren JL, Whary MT, Ge Z, Muthupalani S, Taylor NS, Mobley M *et al*. Lack of commensal flora in Helicobacter pylori-infected INS-GAS mice reduces gastritis and delays intraepithelial neoplasia. Gastroenterology, 2011. 140(1): p. 210-20.

76. Lertpiriyapong K, Whary MT, Muthupalani S, Lofgren JL, Gamazon ER, Feng Y *et al*. Gastric colonisation with a restricted commensal microbiota replicates the promotion of neoplastic lesions by diverse intestinal microbiota in the Helicobacter pylori INS-GAS mouse model of gastric carcinogenesis. Gut, 2014. 63(1): p. 54-63.

77. Han HS, Lee KY, Lim SD, Kim WS and Hwang TS. Molecular identification of Helicobacter DNA in human gastric adenocarcinoma tissues using Helicobacter species-specific 16S rRNA PCR amplification and pyrosequencing analysis. Oncol Lett, 2010. 1(3): p. 555-558.

78. Jo HJ, Kim J, Kim N, Park JH, Nam RH, Seok YJ *et al*. Analysis of Gastric Microbiota by Pyrosequencing: Minor Role of Bacteria Other Than Helicobacter pylori in the Gastric Carcinogenesis. Helicobacter, 2016. 21(5): p. 364-74.

79. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res, 1992. 52(24): p. 6735-40.

80. Aviles-Jimenez F, Vazquez-Jimenez F, Medrano-Guzman R, Mantilla A and Torres J. Stomach microbiota composition varies between patients with non-atrophic gastritis and patients with intestinal type of gastric cancer. Sci Rep, 2014. 4: p. 4202.

81. Eun CS, Kim BK, Han DS, Kim SY, Kim KM, Choi BY *et al*. Differences in gastric mucosal microbiota profiling in patients with chronic gastritis, intestinal metaplasia, and gastric cancer using pyrosequencing methods. Helicobacter, 2014. 19(6): p. 407-16.

82. 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(3): p. 261-6.

83. Li TH, Qin Y, Sham PC, Lau KS, Chu KM, and Leung WK. Alterations in Gastric Microbiota After H. Pylori Eradication and in Different Histological Stages of Gastric Carcinogenesis. Sci Rep, 2017. 7: p. 44935.

84. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G *et al*. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut, 2018. 67(6): p. 1024-1032.

85. Seo IJ, 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(2): p. 162-169.

86. He C, Yang Z and Lu N. Imbalance of Gastrointestinal Microbiota in the Pathogenesis of Helicobacter pylori-Associated Diseases. Helicobacter, 2016. 21(5): p. 337-48.

87. Cao L and Yu J. Effect of Helicobacter pylori Infection on the Composition of Gastric Microbiota in the Development of Gastric Cancer. Gastrointest Tumors, 2015. 2(1): p. 14-25.

88. Dias-Jacome E, Libanio D, Borges-Canha M, Galaghar A and Pimentel-Nunes P. Gastric microbiota and carcinogenesis: the role of non-Helicobacter pylori bacteria - A systematic review. Rev Esp Enferm Dig, 2016. 108(9): p. 530-40.

89. Nardone G and Compare D. The human gastric microbiota: Is it time to rethink the pathogenesis of stomach diseases? United European Gastroenterol J, 2015. 3(3): p. 255-60.

90. Sheh A and Fox JG. The role of the gastrointestinal microbiome in Helicobacter pylori pathogenesis. Gut Microbes, 2013. 4(6): p. 505-31.

91. Giamarellos-Bourboulis E, Tang J, Pyleris E, Pistiki A, Barbatzas C, Brown J *et al*. Molecular assessment of differences in the duodenal microbiome in subjects with irritable bowel syndrome. Scand J Gastroenterol, 2015. 50(9): p. 1076-87.

92. Angelakis E, Armougom F, Carriere F, Bachar D, Laugier R, Lagier JC *et al*. A Metagenomic Investigation of the Duodenal Microbiota Reveals Links with Obesity. PLoS One, 2015. 10(9): p. e0137784.

93. Hayashi H, Takahashi R, Nishi T, Sakamoto M and Benno Y. Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. J Med Microbiol, 2005. 54(Pt 11): p. 1093-101.

94. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al*. Diversity of the human intestinal microbial flora. Science, 2005. 308(5728): p. 1635-8.

95. Wang X, Heazlewood SP, Krause DO and Florin TH. Molecular characterization of the microbial species that colonize human ileal and colonic mucosa by using 16S rDNA sequence analysis. J Appl Microbiol, 2003. 95(3): p. 508-20.

96. Hayashi H, Sakamoto M and Benno Y. Phylogenetic analysis of the human gut microbiota using 16S rDNA clone libraries and strictly anaerobic culture-based methods. Microbiol Immunol, 2002. 46(8): p. 535-48.

97. Hayashi H, Sakamoto M, Kitahara M and Benno Y. Molecular analysis of fecal microbiota in elderly individuals using 16S rDNA library and T-RFLP. Microbiol Immunol, 2003. 47(8): p. 557-70.

98. Green GL, Brostoff J, Hudspith B, Michael M, Mylonaki M, Rayment N *et al*. Molecular characterization of the bacteria adherent to human colorectal mucosa. J Appl Microbiol, 2006. 100(3): p. 460-9.

99. Delgado S, Suarez A and Mayo B. Identification of dominant bacteria in feces and colonic mucosa from healthy Spanish adults by culturing and by 16S rDNA sequence analysis. Dig Dis Sci, 2006. 51(4): p. 744-51.

100. Hold GL, Pryde SE, Russell VJ, Furrie E and Flint HJ. Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. FEMS Microbiol Ecol, 2002. 39(1): p. 33-9.

101. 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, 2016. 6: p. 26337.

102. Vogelstein B and Kinzler KW. The multistep nature of cancer. Trends Genet, 1993. 9(4): p. 138-41.

103. Fearon ER and Vogelstein B. A genetic model for colorectal tumorigenesis. Cell, 1990. 61(5): p. 759-67.

104. Sears CL and Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis, 2011. 203(3): p. 306-11.

105. Kuipers EJ and de Jong A. [Gastrointestinal disorders and Streptococcus bovis bacteremia]. Gastro-intestinale aandoeningen en Streptococcus bovis-bacteriemie. Ned Tijdschr Geneeskd, 1990. 134(28): p. 1337-9.

106. Zhou Y, He H, Xu H, Li Y, Li Z, Du Y *et al*. Association of oncogenic bacteria with colorectal cancer in South China. Oncotarget, 2016. 7(49): p. 80794-80802.

107. 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(4): p. 633-643.

108. 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(6): p. 1419-1433 e5.

109. Gagniere J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E *et al*. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol, 2016. 22(2): p. 501-18.

110. Hold GL. Gastrointestinal Microbiota and Colon Cancer. Dig Dis, 2016. 34(3): p. 244-50.

111. Sears CL and Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe, 2014. 15(3): p. 317-28.

112. Abreu MT and Peek RM, Jr. Gastrointestinal malignancy and the microbiome. Gastroenterology, 2014. 146(6): p. 1534-1546 e3.

113. Wroblewski LE, Peek RM, Jr. and Coburn LA. The Role of the Microbiome in Gastrointestinal Cancer. Gastroenterol Clin North Am, 2016. 45(3): p. 543-56.

114. Hajishengallis G, Darveau RP and Curtis MA. The keystone-pathogen hypothesis. Nat Rev Microbiol, 2012. 10(10): p. 717-25.

115. Han YW. Fusobacterium nucleatum: a commensal-turned pathogen. Curr Opin Microbiol, 2015. 23: p. 141-7.

116. Castellarin M, Warren RL, Freeman JD, Dreolini L, Krzywinski M, Strauss J *et al*. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. Genome Res, 2012. 22(2): p. 299-306.

117. Mima K, Cao Y, Chan AT, Qian ZR, Nowak JA, Masugi Y *et al*. Fusobacterium nucleatum in Colorectal Carcinoma Tissue According to Tumor Location. Clin Transl Gastroenterol, 2016. 7(11): p. e200.

118. Shen XJ, Rawls JF, Randall T, Burcal L, Mpande CN, Jenkins N *et al*. Molecular characterization of mucosal adherent bacteria and associations with colorectal adenomas. Gut Microbes, 2010. 1(3): p. 138-47.

119. Sanapareddy N, Legge RM, Jovov B, McCoy A, Burcal L, Araujo-Perez F *et al*. Increased rectal microbial richness is associated with the presence of colorectal adenomas in humans. Isme J, 2012. 6(10): p. 1858-68.

120. Baxter NT, Ruffin MTt, Rogers MA and Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med, 2016. 8(1): p. 37.

121. Zhou YJ, Zhao DD, Liu H, Chen HT, Li JJ, Mu XQ *et al*. Cancer killers in the human gut microbiota: diverse phylogeny and broad spectra. Oncotarget, 2017. 8(30): p. 49574-49591.

122. Eppinga H, Fuhler GM, Peppelenbosch MP and Hecht GA. Gut Microbiota Developments With Emphasis on Inflammatory Bowel Disease: Report From the Gut Microbiota for Health World Summit 2016. Gastroenterology, 2016. 151(2): p. e1-4.

123. Eppinga H, Poortinga S, Thio HB, Nijsten TEC, Nuij V, van der Woude CJ *et al*. Prevalence and Phenotype of Concurrent Psoriasis and Inflammatory Bowel Disease. Inflamm Bowel Dis, 2017. 23(10): p. 1783-1789.

124. Deuring JJ, Fuhler GM, Konstantinov SR, Peppelenbosch MP, Kuipers EJ, de Haar C *et al*. Genomic ATG16L1 risk allele-restricted Paneth cell ER stress in quiescent Crohn's disease. Gut, 2014. 63(7): p. 1081-91.

125. Eppinga H, Thio HB, Schreurs MWJ, Blakaj B, Tahitu RI, Konstantinov SR *et al*. Depletion of Saccharomyces cerevisiae in psoriasis patients, restored by Dimethylfumarate therapy (DMF). PLoS One, 2017. 12(5): p. e0176955.

126. Zackular JP, Rogers MA, Ruffin MTt and Schloss PD. The human gut microbiome as a screening tool for colorectal cancer. Cancer Prev Res (Phila), 2014. 7(11): p. 1112-21.

127. 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: p. 766.

128. 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(1): p. 70-78.

129. Liang Q, Chiu J, Chen Y, Huang Y, Higashimori A, Fang J *et al*. Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. Clin Cancer Res, 2017. 23(8): p. 2061-2070.

130. Wong SH, Kwong TNY, Chow TC, Luk AKC, Dai RZW, Nakatsu G *et al*. Quantitation of faecal Fusobacterium improves faecal immunochemical test in detecting advanced colorectal neoplasia. Gut, 2017. 66(8): p. 1441-1448.

131. Baxter NT, Koumpouras CC, Rogers MA, Ruffin MTt and Schloss PD. DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. Microbiome, 2016. 4(1): p. 59.

132. Louis P, Hold GL and Flint HJ. The gut microbiota, bacterial metabolites and colorectal cancer. Nat Rev Microbiol, 2014. 12(10): p. 661-72.

133. Yu G, Fadrosh D, Ma B, Ravel J and Goedert JJ. Anal microbiota profiles in HIV-positive and HIV-negative MSM. Aids, 2014. 28(5): p. 753-60.

134. Hoekstra E, Das AM, Swets M, Cao W, van der Woude CJ, Bruno MJ *et al*. Increased PTP1B expression and phosphatase activity in colorectal cancer results in a more invasive phenotype and worse patient outcome. Oncotarget, 2016. 7(16): p. 21922-38.

135. Fuhler GM, Tyl MR, Olthof SG, Lyndsay Drayer A, Blom N, and Vellenga E. Distinct roles of the mTOR components Rictor and Raptor in MO7e megakaryocytic cells. Eur J Haematol, 2009. 83(3): p. 235-45.

136. Queiroz KC, Milani R, Ruela-de-Sousa RR, Fuhler GM, Justo GZ, Zambuzzi WF *et al*. Violacein induces death of resistant leukaemia cells via kinome reprogramming, endoplasmic reticulum stress and Golgi apparatus collapse. PLoS One, 2012. 7(10): p. e45362.

137. Brawner KM, Morrow CD and Smith PD. Gastric microbiome and gastric cancer. Cancer J, 2014. 20(3): p. 211-6.

138. Naylor G and Axon A. Role of bacterial overgrowth in the stomach as an additional risk factor for gastritis. Can J Gastroenterol, 2003. 17 Suppl B: p. 13B-17B.

139. Jakszyn P, Bingham S, Pera G, Agudo A, Luben R, Welch A *et al*. Endogenous versus exogenous exposure to N-nitroso compounds and gastric cancer risk in the European Prospective Investigation into Cancer and Nutrition (EPIC-EURGAST) study. Carcinogenesis, 2006. 27(7): p. 1497-501.

140. Lundberg JO, Weitzberg E, Lundberg JM and Alving K. Intragastric nitric oxide production in humans: measurements in expelled air. Gut, 1994. 35(11): p. 1543-6.

141. Reed PI, Smith PL, Haines K, House FR and Walters CL. Gastric juice N-nitrosamines in health and gastroduodenal disease. Lancet, 1981. 2(8246): p. 550-2.

142. Ruddell WS, Bone ES, Hill MJ, Blendis LM and Walters CL. Gastric-juice nitrite. A risk factor for cancer in the hypochlorhydric stomach? Lancet, 1976. 2(7994): p. 1037-9.

143. Schlag P, Bockler R and Peter M. Nitrite and nitrosamines in gastric juice: risk factors for gastric cancer? Scand J Gastroenterol, 1982. 17(1): p. 145-50.

144. Kodama K, Sumii K, Kawano M, Kido T, Nojima K, Sumii M *et al*. Gastric juice nitrite and vitamin C in patients with gastric cancer and atrophic gastritis: is low acidity solely responsible for cancer risk? Eur J Gastroenterol Hepatol, 2003. 15(9): p. 987-93.

145. Tseng CH, Lin JT, Ho HJ, Lai ZL, Wang CB, Tang SL *et al*. Gastric microbiota and predicted gene functions are altered after subtotal gastrectomy in patients with gastric cancer. Sci Rep, 2016. 6: p. 20701.

146. Fukata M and Abreu MT. Role of Toll-like receptors in gastrointestinal malignancies. Oncogene, 2008. 27(2): p. 234-43.

147. Castano-Rodriguez N, Kaakoush NO and Mitchell HM. Pattern-recognition receptors and gastric cancer. Front Immunol, 2014. 5: p. 336.

148. Pimentel-Nunes P, Afonso L, Lopes P, Roncon-Albuquerque R, Jr., Goncalves N, Henrique R *et al*. Increased expression of toll-like receptors (TLR) 2, 4 and 5 in gastric dysplasia. Pathol Oncol Res, 2011. 17(3): p. 677-83.

149. Pimentel-Nunes P, Goncalves N, Boal-Carvalho I, Afonso L, Lopes P, Roncon-Albuquerque R, Jr. *et al*. Helicobacter pylori induces increased expression of Toll-like receptors and decreased Toll-interacting protein in gastric mucosa that persists throughout gastric carcinogenesis. Helicobacter, 2013. 18(1): p. 22-32.

150. Schmausser B, Andrulis M, Endrich S, Muller-Hermelink HK and Eck M. Toll-like receptors TLR4, TLR5 and TLR9 on gastric carcinoma cells: an implication for interaction with Helicobacter pylori. Int J Med Microbiol, 2005. 295(3): p. 179-85.

151. Wu S, Powell J, Mathioudakis N, Kane S, Fernandez E, and Sears CL. Bacteroides fragilis enterotoxin induces intestinal epithelial cell secretion of interleukin-8 through mitogen-activated protein kinases and a tyrosine kinase-regulated nuclear factor-kappaB pathway. Infect Immun, 2004. 72(10): p. 5832-9.

152. 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. 108(37): p. 15354-9.

153. Nougayrede JP, Homburg S, Taieb F, Boury M, Brzuszkiewicz E, Gottschalk G *et al*. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. Science, 2006. 313(5788): p. 848-51.

154. Dalmasso G, Cougnoux A, Delmas J, Darfeuille-Michaud A and Bonnet R. The bacterial genotoxin colibactin promotes colon tumor growth by modifying the tumor microenvironment. Gut Microbes, 2014. 5(5): p. 675-80.

155. Rubinstein MR, Wang X, Liu W, Hao Y, Cai G, and Han YW. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/beta-catenin signaling via its FadA adhesin. Cell Host Microbe, 2013. 14(2): p. 195-206.

156. Pimentel-Nunes P, Goncalves N, Boal-Carvalho I, Afonso L, Lopes P, Roncon-Albuquerque R, Jr. *et al*. Decreased Toll-interacting protein and peroxisome proliferator-activated receptor gamma are associated with increased expression of Toll-like receptors in colon carcinogenesis. J Clin Pathol, 2012. 65(4): p. 302-8.

157. Li TT, Ogino S and Qian ZR. Toll-like receptor signaling in colorectal cancer: carcinogenesis to cancer therapy. World J Gastroenterol, 2014. 20(47): p. 17699-708.

158. Pimentel-Nunes P, Teixeira AL, Pereira C, Gomes M, Brandao C, Rodrigues C *et al*. Functional polymorphisms of Toll-like receptors 2 and 4 alter the risk for colorectal carcinoma in Europeans. Dig Liver Dis, 2013. 45(1): p. 63-9.

159. Nagao-Kitamoto H, Kitamoto S, Kuffa P and Kamada N. Pathogenic role of the gut microbiota in gastrointestinal diseases. Intest Res, 2016. 14(2): p. 127-38.

160. Anand S, Kaur H and Mande SS. Comparative In silico Analysis of Butyrate Production Pathways in Gut Commensals and Pathogens. Front Microbiol, 2016. 7: p. 1945.

161. Ou J, Carbonero F, Zoetendal EG, DeLany JP, Wang M, Newton K *et al*. Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans. Am J Clin Nutr, 2013. 98(1): p. 111-20.

162. Li S, Konstantinov SR, Smits R and Peppelenbosch MP. Bacterial Biofilms in Colorectal Cancer Initiation and Progression. Trends Mol Med, 2017. 23(1): p. 18-30.

3

CHAPTER 3

Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals

F.E.R. Vuik[1*], J. Dicksved[2*], S.Y. Lam[1], G.M. Fuhler[1], L.J.W. van der Laan[3], A. van de Winkel[1], S.R. Konstantinov[1], M.C.W. Spaander[1], M.P. Peppelenbosch[1], L. Engstrand[4], E.J. Kuipers[1]
* These authors contributed equally

[1] Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands; [2] Department of Animal Nutrition and Management, Swedish University of Agricultural Sciences, Uppsala, Sweden; [3] Department of Surgery, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands; [4] Department of Microbiology, Tumor and Cell Biology, Karolinska institute, Stockholm, Sweden.

United European Gastroenterol J. 2019 Aug;7(7):897-907.

Abstract

Background: Homeostasis of the gastrointestinal tract depends on healthy bacterial microbiota, with alterations in microbiota composition suggested to contribute to diseases. To unravel bacterial contribution to disease pathology, a thorough understanding of the microbiota of the complete gastrointestinal tract is essential. To date, most microbial analyses have either focused on fecal samples, or on the microbial constitution of one gastrointestinal location instead of different locations within one individual.

Objective: We aimed to analyse the mucosal microbiome along the entire gastrointestinal tract within the same individuals.

Methods: Mucosal biopsies were taken from nine different sites in 14 individuals undergoing antegrade and subsequent retrograde double-balloon enteroscopy. The bacterial composition was characterized using 16S rRNA amplicon sequencing with Illumina Miseq.

Results: At double-balloon enteroscopy, one individual had a cecal adenocarcinoma and one individual had Peutz-Jeghers polyps. The composition of the microbiota distinctively changed along the gastrointestinal tract with larger bacterial load, diversity and abundance of Firmicutes and Bacteroidetes in the lower gastrointestinal tract than the upper gastrointestinal tract, which was predominated by Proteobacteria and Firmicutes.

Conclusion: We show that gastrointestinal location is a larger determinant of mucosal microbial diversity than inter-person differences. These data provide a baseline for further studies investigating gastrointestinal microbiota-related disease.

Keywords:

Colonic microflora; colon; gastrointestinal tract; intestinal microbiology; small bowel.

1 Introduction

In recent years, an increasing level of knowledge on the interaction between host and bacteria has made us come to regard the gut microbiota as a separate entity [1]. The microbiota have important immunological, structural, metabolic and defence functions in the gut. Alterations in microbiota composition have been linked to intestinal disease, including colorectal cancer and inflammatory bowel disease (IBD). Unravelling the microbiota composition and its distribution along the gastrointestinal (GI) lining in healthy individuals is important to understand the role of the microbiota in disease [2].

Characterization of the microbiota in the entire GI tract is hampered by the fact that some locations are more difficult to access than others and most research has focused on the colonic fecal microbiota [1]. The mucosal microbiome is arguably the more relevant compartment, as such mucosa-associated flora lives in close contact with the GI tract lining. The microbial composition of the colonic mucosa has been most often investigated. While it is clear that the composition and abundance of mucosal microbiota of the esophagus and stomach in healthy individuals differ from that in the colon [3-5], information about the microbial composition in the jejunum and ileum is scarce because of the inaccessibility of these sites. Nevertheless, differences in the physiological functions of GI sites logically predict regional bacterial differences. The colonic microbiota for example, are driven by complex carbohydrates whereas simple carbohydrates fuel the microbiota in the small intestine [2, 6]. Furthermore, the composition of the mucus layer protecting the epithelial barrier from excessive bacterial contact differs along the intestinal tract [7, 8]. Given the limited information about mucosal microbiota in the entire GI tract, we aimed to characterize the mucosal microbiota along the length of the entire GI tract within the same subjects.

2 Materials and Methods

2.1 Subject recruitment

Subjects, all inhabitants of The Netherlands, had abdominal symptoms of unknown cause requiring diagnostic antegrade and subsequent retrograde double-balloon enteroscopy (DBE). Exclusion criteria were: patients younger than 18 years, use of antibiotics three months before DBE, IBD, and failure to understand written Dutch. The study was conducted in accordance with the Declaration of Helsinki Principles and approved by the ethical committee of the Erasmus MC University Medical Center, Rotterdam (MEC-2017-151) on 3 April 2017. Written informed consent was obtained from each patient included in the study.

2.2 Sampling

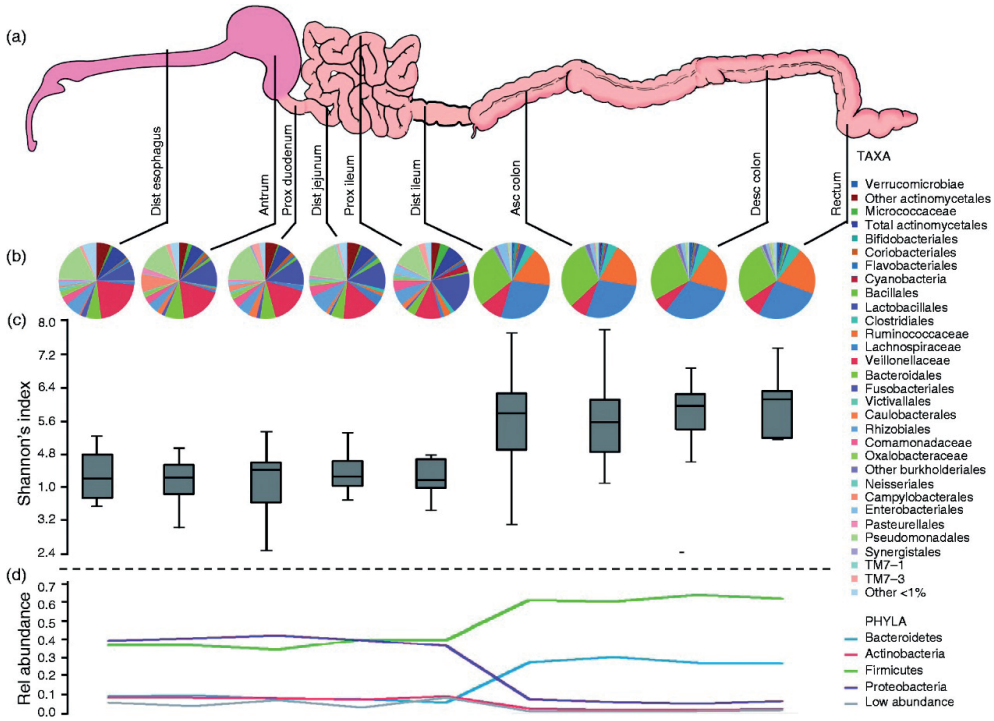
Mucosal samples were obtained endoscopically using antegrade and subsequent retrograde DBE at the Erasmus MC University Medical Center Rotterdam using Fujinon EN-450P5 and EN-450T5 endoscopes (Fujinon Inc., Saitama, Japan). Endoscopes were disinfected before use. Mucosal biopsies using standard biopsy forceps were taken at nine different sites of the GI tract (**Figure 1**). Upper GI biopsies (esophagus to proximal ileum) were collected using antegrade endoscopy and lower GI biopsies (distal ileum to rectum) with retrograde endoscopy. Between the antegrade and retrograde endoscopy, the canal of the endoscope was cleaned with sterile water. All patients used bowel preparation before DBE consisting of macrogol and electrolytes (Klean-Prep, Norgine BV, Amsterdam, The Netherlands).

Samples were stored in Eppendorf tubes (0.2 ml) with a stabilising reagent Allprotect (Qiagen Gmbh, Hilden, Germany). The samples were homogenised using the MagNA Lyser machine (Roche Diagnostics, Mannheim, Germany), stored in Trizol tubes (Invitrogen, Groningen, The Netherlands) and immediately frozen and stored at −80˚C for subsequent analyses. Deoxyribonucleic acid (DNA) was isolated from the samples using QIAamp DNA mini kit (Qiagen) with an initial bead beating step added to the protocol, as described previously [9].

2.3 Generation of 16S rRNA gene amplicons

Sequencing libraries were prepared by amplifying the V3–V4 region of the *16S rRNA* gene using the 341F-805R primers, as described earlier [10]. After the initial amplification, polymerase chain reaction (PCR) products were confirmed with gel electrophoresis and purified using Agencourt AMPure XP magnetic beads (Beckham Coulter Inc., Bromma, Sweden).

A second PCR was performed to attach Illumina adapters and barcodes that allow for multiplexing and the products were purified as above, quantified and pooled into equimolar amounts. Samples were sequenced using the Illumina MiSeq platform at Science for Life Laboratory, Solna, Sweden. From the generated sequence data, primer sequences were trimmed away and the paired-end reads produced by the sequencing instrument were merged using SeqPrep version 1.1 (https://github.com/jstjohn/SeqPrep) with default parameters and thereafter the merged sequences were processed with QIIME 1.8 pipeline (Quantitative Insights into Microbial Ecology) [11]. A de novo operational taxonomic unit (OTU) strategy was used to assign sequences to OTUs. Using the UCLUST algorithm built into the QIIME pipeline, sequences were clustered at 97% identity against the Greengenes reference database [12, 13].



▲**Figure 1 | Study overview. (a)** Location of retrieved mucosal biopsies of the gastrointestinal (GI) tract. **(b)** Marked differences in bacterial taxa are present between different GI locations as indicated by boxplot of the median Shannon index of different locations. **(c)** Diversity as measured by the Shannon index is higher in the distal (dist) ileum, ascending (asc) colon, descending (desc) colon and rectum as compared to distal (dist) esophagus, antrum, proximal (prox) duodenum, distal (dist) jejunum and proximal (prox) ileum. **(d)** Relative abundance of the major phyla fluctuates along the GI tract.

2.4 Polymerase chain reaction (PCR) analysis

Conventional PCR was performed to confirm bacterial and human DNA isolation of biopsies. While analysing the results of this study, we noticed that family Helicobacteraceae was in the antrum, but also in other parts of the GI tract. However, sequencing did not allow us to identify this feature on species level. To improve our understanding, additional PCR analyses were performed. DNA amplification was executed with the Applied Biosystems 2720 Thermal Cycler (Thermo Fisher Scientific, Waltham, MA) using primers targeting *16S rRNA*, *Helicobacter pylori* (HP) specific *UreA* and *VacA* and human *ACTB* genes (**Supplementary Table S1**). The reaction mixture contained GoTaq buffer (Promega, Madison, WI), 1.25mM $MgCl_2$ (Promega), 0.167mM (each) deoxynucleotides (Roche Diagnostics), 2.5U GoTaq polymerase (Promega), 333nM of each primer (Sigma-Aldrich, St Louis, MO) and 2µl non-

normalized stock DNA. The PCR cycle program consisted of four minutes of 95˚C, several cycles of 30 seconds of denaturing at 95˚C, 30 seconds of annealing and one minute extension at 72˚C, followed by the final extension for 10 min at 72˚C. Annealing temperature was 60˚C for *16S rRNA, UreA* and *VacA* primers and 60.5˚C for *ACTB primers*. Number of cycles was 40 for HP specific gene primers and 35 for *16S rRNA* and *ACTB* gene primers. Amplicons were analyzed by gel electrophoresis using 2% agarose gel in 1X TBE (Tris-borate-EDTA) buffer and bacterial DNA load quantified by Image J software.

2.5 Statistical analysis

The similarity between two samples was calculated using weighted UniFrac distances. Biodiversity within a sample was measured using the Shannon index. All diversity calculations were also performed for a least detectable relative abundance of 0.1%, corresponding to 1000 sequences in a sample, but this did not alter the results (data not included). Principal coordinate analysis (PCoA) using Bray-Curtis metrics based on abundance data from sequences classified to genus level was performed to determine clustering patterns among the subjects. Differences in diversity and similarity indices were tested with Mann-Whitney or Kruskal-Wallis test using the IBM SPSS statistics 21 software (Chicago, IL). For differences in relative abundance of specific bacterial taxa, Wilcoxon tests and linear regressions were applied using the *r* statistical framework, version 3.0.1.

3 Results

3.1 Subject population

Fourteen subjects undergoing an antegrade and subsequent retrograde DBE were included. In 13 patients, the mucosal samples were also studied by histology. Twelve subjects had no relevant anomalies found with DBE and histology (**Table 1**). One patient had Peutz-Jeghers polyps in the distal jejunum and one patient had a cecum tumor (**Supplementary Table S2**).

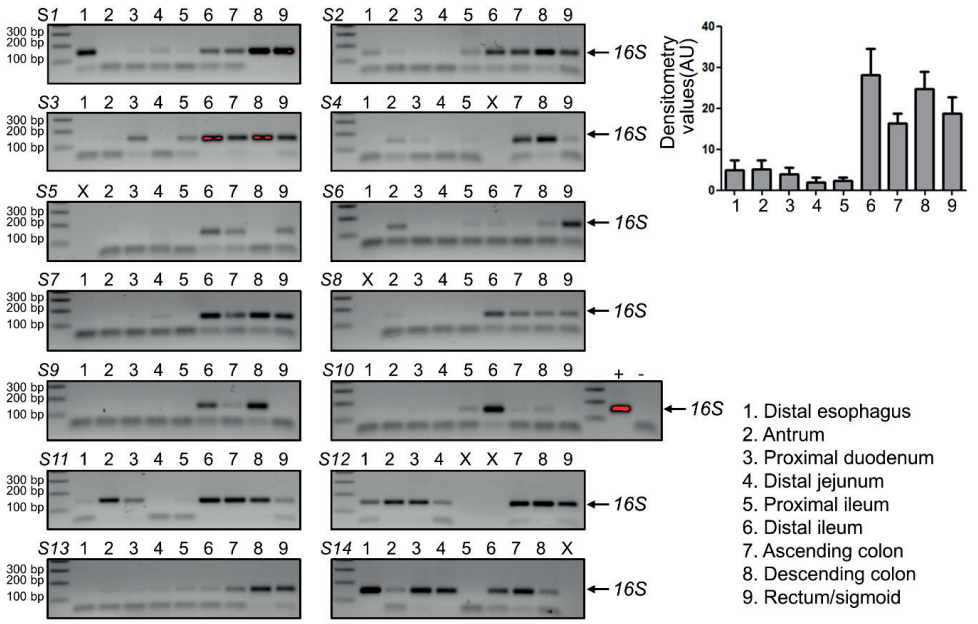
3.2 Overview of sequencing data generated from the samples

A total of 118 mucosal samples were retrieved from nine locations of the GI tract in 14 individuals. Eight samples could not be sequenced due either to inability to analyse the retrieved samples or inability to reach the site. First, we confirmed bacterial DNA isolation from all samples by conventional PCR. While human genomic DNA content was similar in all samples (**Supplementary Figure S1**), the bacterial load decreased from esophagus to proximal ileum, but increased again in the lower GI tract (**Figure 2**). Samples were subsequently subjected

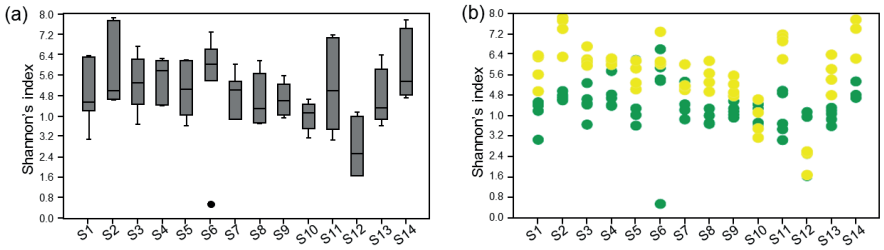
Table 1 | Baseline characteristics of study subjects

Mean age, mean in years (range)	51	(42-60)
Sex, *n* of males (%)	7	(50%)
Race, *n* of Caucasians (%)	10	(71%)
Body mass index (BMI)		
Mean (SD) (kg, m^2)	22.9	(5.4)
Unknown	5	
Current smoker, *n* (%)		
Yes	8	(58%)
No	3	(21%)
Unknown	3	(21%)
Alcohol, *n* (%)		
Yes	6	(43%)
No	5	(36%)
Unknown	3	(21%)
Medication use, *n* (%)		
Yes	11	(79%)
No	3	(21%)
Medical history, *n* (%)		
Hypertension	1	(6%)
Diabetes	2	(14%)
Cardiac disease	1	(6%)
Peripheral arterial disease	2	(14%)
Stroke	1	(6%)
Chronic pulmonary disease	1	(6%)
Liver disease	1	(6%)
Resection part of GI tract	2	(14%)
Other	2	(14%)
No medical history	2	(14%)
Presenting symptoms, *n* (%)		
Iron deficiency anemia	5	(29%)
Diarrhea	4	(24%)
Abdominal complaints	4	(24%)
Weight loss	3	(18%)
Rectal blood loss	1	(5%)
Findings at double-balloon enteroscopy, *n* (%)		
No abnormal findings	10	(71%)
Ulcerative lesions in small bowel	1	(7%)
Polyps in small bowel	2	(14%)
Polyps in colon	1	(7%)
Pathology finding, *n* (%)		
No abnormal findings	9	(64%)
Reflux esophagitis	1	(7%)
Chronic inflammation antrum	1	(7%)
Chronic inflammation small bowel	1	(7%)
Peuthz Jeghers polyps	1	(7%)
Ulcerative changes	1	(7%)

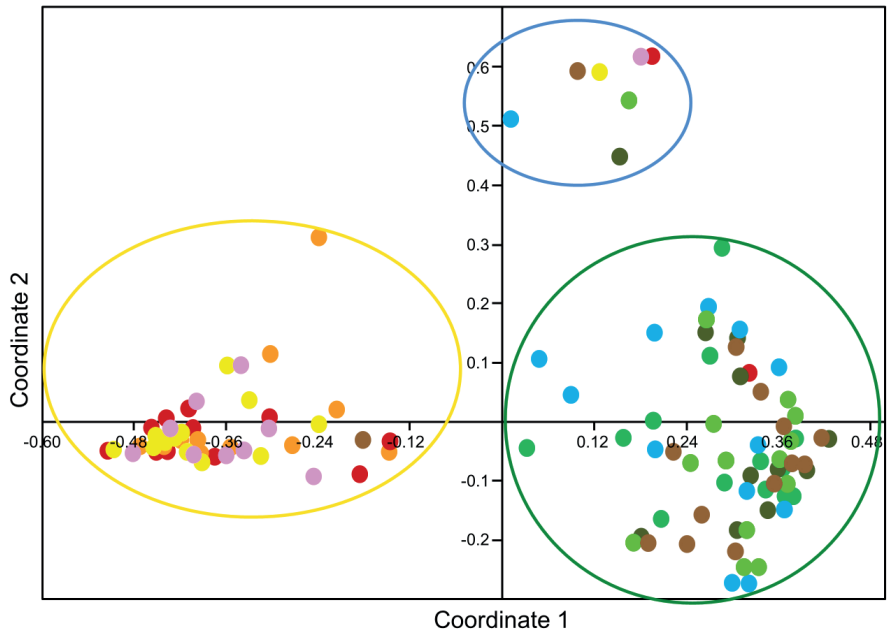
Abbreviations: GI, gastrointestinal; n, number.



▲Figure 2 | Differential bacterial load at the mucosa along the gastrointestinal tract. Bacterial abundance at all locations of the 14 included subjects was determined by 16S polymerase chain reaction (PCR) and electrophoresis results are shown for all samples. Missing samples are indicated by 'X'. +: positive control (DNA isolated from human fecal sample); –: negative control (water). For semi-quantitative analysis, bands were quantified and for each patient, the data was normalized to the total intensity per gel to adjust for differences between gel compositions and staining intensity. Mean ± standard error of the mean (SEM) is shown in bar graph.



▲Figure 3 | The α-diversity of the microbiota of the gastrointestinal (GI) tract. (a) Boxplot of the median Shannon index over all locations within each subject (S1–S14). Subject S12 shows a low α-diversity. The outlier for subject S6 represents the antrum biopsy. **(b)** The same data, but represented in a Jitter plot, with each dot representing a location in the GI tract. Green-coloured dots represent the distal esophagus, antrum, proximal duodenum, distal jejunum and proximal ileum (upper GI tract) and the yellow colored dots represent the distal ileum, ascending colon, descending colon and rectum (lower GI tract samples). All subjects, except S10 show a higher α-diversity in samples obtained from the lower GI tract as compared to the upper GI tract.

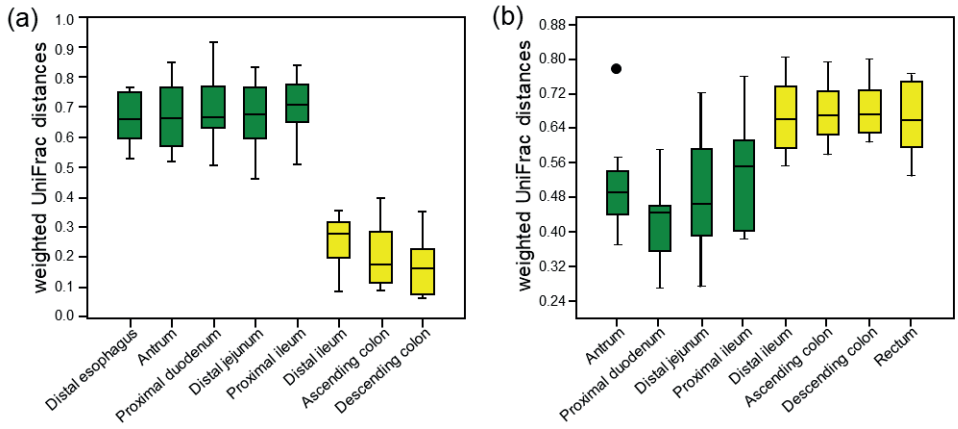


▲Figure 4 | Principal coordinate analysis (PCoA) plot illustrates a clear difference between gut location and microbial composition. Different coloured dots represent different locations of the gastrointestinal (GI) tract. The green circle contains mainly esophagus, antrum, proximal duodenum, distal jejunum and proximal ileum samples (upper GI tract), the yellow circle contains only distal ileum, ascending colon, descending colon and rectum samples (lower GI tract). The blue circle highlights samples dominated by Enterobacteriaceae which derived from one patient with a cecum tumor (S12).

to 16S rRNA amplicon sequencing using a V3-V4 specific primer set, resulting in a total of 4.369.079 high-quality sequences, with 37.026 sequences per sample (range: 17.294–68.696).

3.3 Diversity of the microbiota along the GI tract

To estimate the diversity of microbial communities of biopsies in the entire GI tract, α-diversity (Shannon index) analysis, was performed (**Figure 1**). The location of sampling had a significant influence on the α-diversity of the microbiota, with samples taken from esophagus to proximal ileum harboring a lower level of microbial diversity than samples obtained from terminal ileum to rectum ($P < 0.05$). When comparing the average α-diversity of the individual locations from individual subjects, a wide spread in the mean Shannon index between individuals became apparent with, in particular, subject 12 showing a low diversity in all samples (**Figure 3a**). This patient was diagnosed with a cecum tumor. Nevertheless, all participants, except subject 10, showed a higher α-diversity in lower GI locations (**Figure 3b**) as compared to upper GI locations.



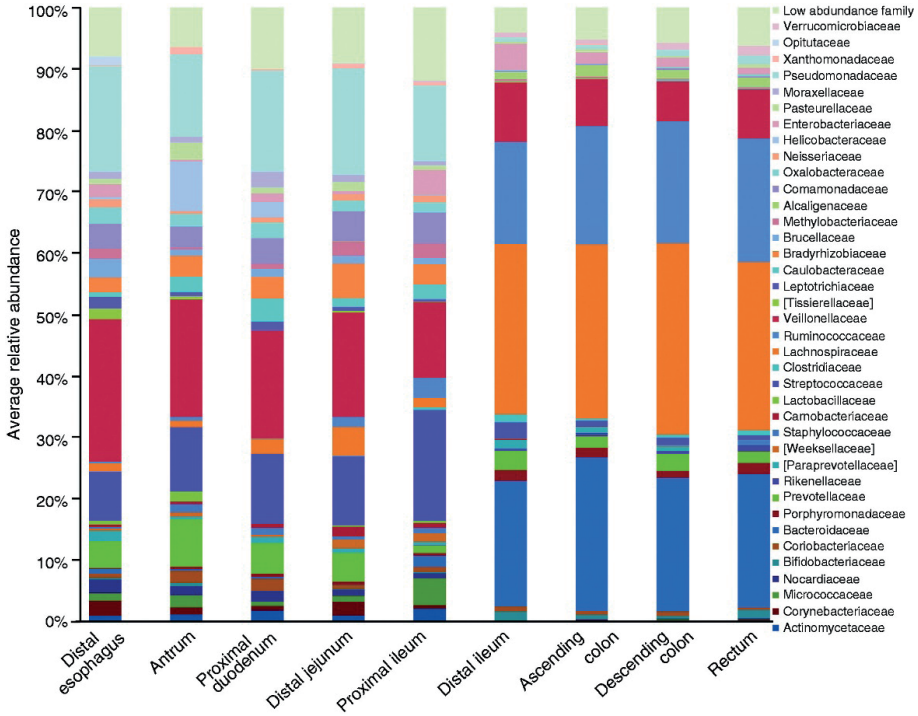
▲Figure 5 | Similarity between gastrointestinal (GI) tract sites analyzed using weighted UniFrac distances. (a) The microbiota in the rectum were compared to eight other locations and serve as a good proxy for other lower GI locations. **(b)** The microbiota of the distal esophagus were compared to eight other locations, and is a less efficient predictor for microbiota of other locations. Green: upper GI tract; Yellow: lower GI tract.

3.4 Differential microbial composition along the GI tract

We further searched for clustering patterns among samples according to their microbial population structure by PCoA based on Bray-Curtis dissimilarity. Again, a distinct separation of bacterial community structure was observed, with samples from the distal esophagus to the proximal ileum clustering together, separately from distal ileum to rectum (**Figure 4**). Several samples clustered neither with the upper nor the lower GI samples, but belonged to the patient with a cecum tumor. These samples appeared to be dominated by Enterobacteriaceae (**Supplementary Figure S2**).

Cluster analysis using Euclidian distance at family level was used to visualise the data in a different way, which again demonstrates the separate clustering of this patient with a cecum tumor and other samples (**Supplementary Figure S3**). Samples from individual patients appear to cluster more closely together in lower GI samples than upper GI samples (**Supplementary Figures S3-4**).

The similarity in microbiota composition between different sites in the GI tract was also visualised using weighted UniFrac, which showed that the microbial composition in the rectum was a good predictor for the microbial composition in the ascending and descending colon and – to a somewhat lesser extent – the distal ileum (**Figure 5a**). The composition of the microbiota in the distal esophagus was also compared to the other locations in the GI tract. However, the microbiota in the distal esophagus was not as good

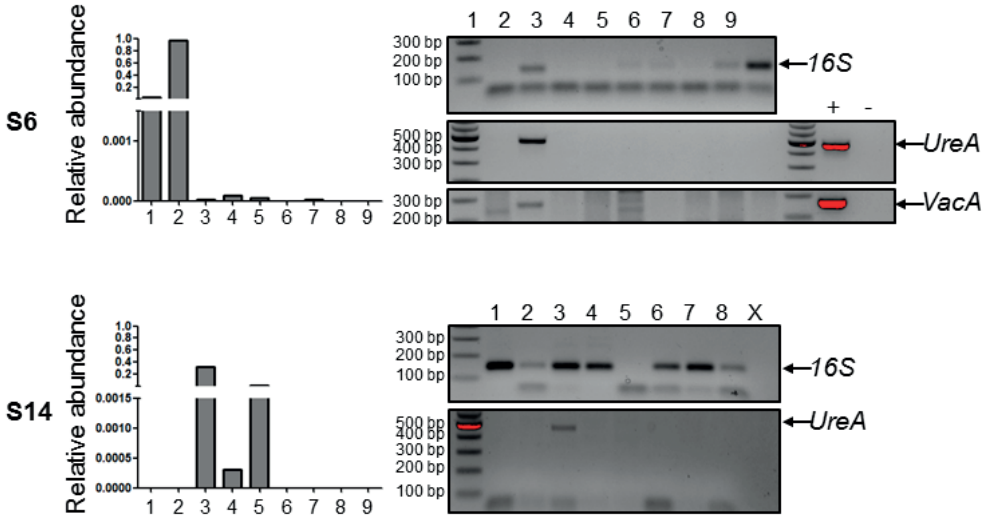


▲Figure 6 | Most important bacteria at family level (>1% abundance) per gastrointestinal (GI) location. Samples from subject 12, which were predominated by Enterobacteriaceae and of low α-diversity, were excluded from this analysis.

a predictor for the other locations in the upper GI tract as the rectum was for the lower GI tract (**Figure 5b**).

3.5 Characterization of mucosa-associated microbiota

All regions in the GI tract were dominated by three major bacterial phyla: Bacteroidetes, Firmicutes and Proteobacteria. Although ubiquitously dominant within the entire GI tract, each of the three phyla revealed distinct profiles along the length of the GI tract (**Figure 1d**). The mucosa-associated microbiota of the upper GI tract were dominated by Proteobacteria (mean abundance of 40 ± 2.1%) and Firmicutes (38 ± 2.3%). However, in the lower GI tract the level of Proteobacteria decreased consistently (distal colon (5.3 ± 0.4%). Firmicutes, already highly abundant in the upper GI tract, dominated the large intestine with the highest level in the distal colon (mean abundance 64 ± 7%). Bacteroidetes was present at low levels in the upper GI tract (8 ± 1.6%), but became a dominant phylum in the lower GI tract (mean abundance in ascending colon 28 ± 1.6%).



▲Figure 7 | *Helicobacter pylori* (HP) predominates in the antrum from one patient, and extends beyond the stomach. The upper and lower panels show the relative abundance of *Helicobacter* species across the nine different gastrointestinal (GI) sites as determined by sequencing in subject S6 and S14, respectively. The 16S PCR, similar to Figure 2, is shown here to allow comparison of total bacterial abundance in these samples. HP was confirmed by polymerase chain reaction in the high *Helicobacter* abundant samples by *UreA* and *VacA* in subject S6. The antrum was dominated by HP, resulting in a low diversity in this sample (see Figure 2). Identification of HP at species level was confirmed by PCR of *UreA* in subject 14. While HP was not detected in the antrum, high levels were present in the proximal duodenum. Abbreviations: X: missing samples.1: distal esophagus; 2: antrum; 3: proximal duodenum; 4: distal jejunum, 5: proximal ileum; 6: distal ileum; 7: ascending colon; 8: descending colon; 9: rectum; +: positive control of pure HP culture strain ATCC®43504 (American Type Culture Collection, Rockville, Maryland); –: negative control (water).

The most prevalent bacterial families in the upper GI tract were Veillonellaceae, Pseudomonadaceae and Streptococcaceae (**Figure 6**). In contrast to other sites in the GI tract, Prevotellaceae (relative abundance of 8%) and Helicobacteraceae (relative abundance of 8%) were dominant in the antrum. Helicobacter species were detected in nine subjects, and predominated the antrum of one subject (S6) to the extent that other species were almost not found (**Supplementary Figure S3**). PCR analysis of the UreA and VacA gene confirmed that the Helicobacteraceae detected by sequencing were indeed HP (**Figure 7**; upper panel). Helicobacter was present across the entire upper GI tract, and some lower GI tract locations in three subjects, which confirms data that this bacterium may spread beyond the stomach (**Supplementary Figure S5**). Interestingly, subject S14 showed high levels of Helicobacteraceae in the proximal duodenum, while not detected in the antrum (**Figure 7**; lower panel).

In the distal jejunum, Bradyrhizobiaceae (relative abundance of 6%) occurred more

often compared to other parts of the GI tract. The same applies to Micrococcaceae (relative abundance of 4%) in the proximal ileum. The lower GI tract was dominated by Lachnospiraceae, Bacteroidaceae, Ruminococcaceae and Veillonellaceae. The highest abundance of the bacterial family Clostridiaceae (relative abundance of 1%) was seen in the distal ileum. Rikenellaceae was only seen with a higher relative abundance than 1% in the rectum.

4 Discussion

This study describes the composition of the microbiota along the entire GI tract in the same individuals without significant pathology. In agreement with earlier reports, the bacterial load decreases from the esophagus to the proximal ileum, but drastically increases again in the lower GI tract, starting from the distal ileum. The composition of the microbiota markedly changes along the GI tract, with the most prevalent bacterial families present in the upper GI tract Veillonellaceae, Pseudomonadaceae and Streptococcaceae, while the lower GI tract is dominated by Lachnospiraceae, Bacteroidaceae and Ruminococcaceae.

Our findings to a large extent reflect data obtained from other studies comparing only partly matched samples, but probing multiple locations within one patient may provide better accuracy. One report comparing only duodenal and rectal content from healthy individuals reported higher Shannon diversity values in both mucosa and luminal content from the duodenum [14], while others support our findings of a less complex luminal microbiota in the small intestine compared to the colonic content [6, 15].

Arguably, the least studied GI sites in the current literature are the jejunum and distal ileum. In the jejunum, Proteobacteria and Firmicutes were the most dominant phyla, and at family level Veillonellaceae, Pseudomonadaceae and Streptococcaceae dominated. A previous study retrieving mucosal biopsies from the proximal jejunum of 19 healthy individuals also observed Proteobacteria, Bacteroidetes and Firmicutes as the predominant phyla, although family level classification indicated Brevibacteriaceae, Barnesiellaceae and Leuconostocaceae [16]. Possible explanations for these discrepancies could be the difference in individual populations (Taiwanese versus Dutch population) as well as alternative methodologies used for sampling, preparation and analysis of the samples.

In terms of the proximal and distal ileum, our samples were found to have large differences in composition. In the proximal ileum, Proteobacteria and Firmicutes dominated, whereas Firmicutes and Bacteroidetes were the dominant phyla in the distal ileum. It is conceivable that the distal ileum was contaminated from the colon, either due to sampling or through

bowel movements. At present the only comparison that can be made in this context comes from animal studies. A study comparing 10 paired GI locations in mice showed that the largest difference between two locations in terms of bacterial diversity was seen between ileum and proximal cecum, with lower GI samples clustering away from upper GI samples [17]. In pigs, a similar clear separation between the upper and lower GI could be seen, although in this case the dividing line appeared to lie between jejunum and ileum [18].

A further notable finding in our study was that a patient who had a cecum tumor showed significant dysbiosis predominated by Enterobacteriaceae in all other GI sites tested. A role for Enterobacteriaceae in carcinogenesis has been suggested before, as several enterobacterial strains are known to produce DNA-damaging genotoxins and may therefore cause mutations [19, 20]. The major strength of this study is that we collected nine mucosal samples along the entire GI tract of 14 different individuals allowing us to study the composition of the microbiota along the length of the gut. Since all individuals underwent an antegrade DBE followed directly by a retrograde DBE, no bias could have occurred based on the timeframe.

There are also a number of limitations. Firstly, the same endoscope was used for anterograde and retrograde DBE. Although the canal of the endoscope was cleaned with sterile water between the antegrade and retrograde DBE, it is impossible to exclude contamination from the upper GI tract to the lower GI tract using this methodology [21]. However, the low level of similarity of the microbial composition in the upper and lower GI tract suggests that this is not a major issue in our study. Secondly, the subjects in our study underwent DBE for unexplained symptoms and therefore may not fully represent healthy individuals. However, ethical considerations preclude performing DBE in individuals without clinical indication and thus we consider our study the best that can be achieved with current technical approaches. Third, neither DBE nor histopathology of the retrieved biopsies showed clinical abnormalities except for one patient with a cecum tumor and one patient with Peutz-Jeghers polyps. Fourth, patients were treated with colonic lavages prior to DBE, which could potentially have diminished the diversity of the mucosa-associated microbiota. Unfortunately, DBE cannot be performed without bowel preparation [22]. Finally, stool samples were not collected of these patients and therefore the fecal microbiota could not be analyzed. Whether stool and mucosal microbiome correlate well is somewhat debated in literature, and having stool samples would have been of value [1, 14]. With the exception of the patient with a cecum tumor, the data represented here could be conceived as representing the 'normal' mucosal microbiome. While it is already well described that education of the immune system depends on the intestinal microbiome,

to what extent local mucosal differences affect local immunological responses is less well elucidated. Diseases like IBD are largely driven by an altered immunological response towards intestinal microbes. Thus, a comparison of disease-location specific mucosal microbial changes to normal microbiome signatures at these sites may be of use [23]. The use of fecal microbiota transplantation for IBD has been advocated, and it is thought that optimal donor selection is important for clinical efficacy, although more research is needed to identify which components of the gut microbiome constitute key members [24].

In conclusion, we have generated a first overview of the composition of the microbiota along the entire GI tract. This study is of particular importance in helping us to understand the interactions between bacterial communities and human cells and takes us to the next step in describing the impact of the microbiota on health and its involvement in diseases.

References

1. Watt E, Gemmell MR, Berry S, Glaire M, Farquharson F, Louis P *et al*. Extending colonic mucosal microbiome analysis-assessment of colonic lavage as a proxy for endoscopic colonic biopsies. Microbiome, 2016. 4(1): p. 61.
2. Marchesi JR, Adams DH, Fava F, Hermes GD, Hirschfield GM, Hold G *et al*. The gut microbiota and host health: a new clinical frontier. Gut, 2016. 65(2): p. 330-9.
3. Pei Z, Bini EJ, Yang L, Zhou M, Francois F, and Blaser MJ. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci U S A, 2004. 101(12): p. 4250-5.
4. Yang L, Lu X, Nossa CW, Francois F, Peek RM, and Pei Z. Inflammation and intestinal metaplasia of the distal esophagus are associated with alterations in the microbiome. Gastroenterology, 2009. 137(2): p. 588-97.
5. Bik EM, Eckburg PB, Gill SR, Nelson KE, Purdom EA, Francois F *et al*. Molecular analysis of the bacterial microbiota in the human stomach. Proc Natl Acad Sci U S A, 2006. 103(3): p. 732-7.
6. Zoetendal EG, Raes J, van den Bogert B, Arumugam M, Booijink CC, Troost FJ *et al*. The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. Isme J, 2012. 6(7): p. 1415-26.
7. Moran C, Sheehan D and Shanahan F. The small bowel microbiota. Curr Opin Gastroenterol, 2015. 31(2): p. 130-6.
8. Johansson ME, Sjövall H and Hansson GC. The gastrointestinal mucus system in health and disease. Nat Rev Gastroenterol Hepatol, 2013. 10(6): p. 352-61.
9. Dicksved J, Lindberg M, Rosenquist M, Enroth H, Jansson JK, and Engstrand L. Molecular characterization of the stomach microbiota in patients with gastric cancer and in controls. J Med Microbiol, 2009. 58(Pt 4): p. 509-516.
10. Hugerth LW, Wefer HA, Lundin S, Jakobsson HE, Lindberg M, Rodin S *et al*. DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. Appl Environ Microbiol, 2014. 80(16): p. 5116-23.
11. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al*. QIIME allows analysis of high-throughput community sequencing data. Nat Methods, 2010. 7(5): p. 335-6.
12. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 2010. 26(19): p. 2460-1.
13. Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, and Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics, 2010. 26(2): p. 266-7.
14. Li G, Yang M, Zhou K, Zhang L, Tian L, Lv S *et al*. Diversity of Duodenal and Rectal Microbiota in Biopsy Tissues and Luminal Contents in Healthy Volunteers. J Microbiol Biotechnol, 2015. 25(7): p. 1136-45.
15. Jandhyala SM, Talukdar R, Subramanyam C, Vuyyuru H, Sasikala M, and Nageshwar Reddy D. Role of the normal gut microbiota. World J Gastroenterol, 2015. 21(29): p. 8787-803.
16. Chung CS, Chang PF, Liao CH, Lee TH, Chen Y, Lee YC *et al*. Differences of microbiota in small bowel and faeces between irritable bowel syndrome patients and healthy subjects. Scand J Gastroenterol, 2016. 51(4): p. 410-9.
17. Suzuki TA and Nachman MW. Spatial Heterogeneity of Gut Microbial Composition along the Gastrointestinal Tract in Natural Populations of House Mice. PLoS One, 2016. 11(9): p. e0163720.
18. Kelly J, Daly K, Moran AW, Ryan S, Bravo D, and Shirazi-Beechey SP. Composition and diversity of mucosa-associated microbiota along the entire length of the pig gastrointestinal tract; dietary influences. Environ Microbiol, 2017. 19(4): p. 1425-1438.

19. Yurdakul D, Yazgan-Karataş A and Şahin F. Enterobacter Strains Might Promote Colon Cancer. Curr Microbiol, 2015. 71(3): p. 403-11.
20. Allen-Vercoe E and Jobin C. Fusobacterium and Enterobacteriaceae: important players for CRC? Immunol Lett, 2014. 162(2 Pt A): p. 54-61.
21. Walker MM and Talley NJ. Review article: bacteria and pathogenesis of disease in the upper gastrointestinal tract--beyond the era of Helicobacter pylori. Aliment Pharmacol Ther, 2014. 39(8): p. 767-79.
22. Hollister EB, Gao C and Versalovic J. Compositional and functional features of the gastrointestinal microbiome and their effects on human health. Gastroenterology, 2014. 146(6): p. 1449-58.
23. Ueno A, Jeffery L, Kobayashi T, Hibi T, Ghosh S, and Jijon H. Th17 plasticity and its relevance to inflammatory bowel disease. J Autoimmun, 2018. 87: p. 38-49.
24. Woodworth MH, Carpentieri C, Sitchenko KL and Kraft CS. Challenges in fecal donor selection and screening for fecal microbiota transplantation: A review. Gut Microbes, 2017. 8(3): p. 225-237.

Supplementary information

Supplementary Table S1 | Primer sequences

Target	Sequence (5'-3')		Ref
Universal 16S	Forward	CGGTGAATACGTTCCCGG	1-3
	Reverse	TACGGCTACCTTGTTACGACTT	
UreA	Forward	ATGAAACTCACCCCAAAAGA	4-5
	Reverse	TTCACTTCAAAGAAATGGAAGTGTGA	
VacA S1/S1	Forward	ATGGAAATACAACAAACACAC	6
	Reverse	CTGCTTGAATGCGCCAAAC	
ACTB	Forward	CTGGAACGGTGAAGGTGACA	7
	Reverse	AAGGGACTTCCTGTAACAATGCA	

Abbreviations: ACTB, actin-beta.

Supplementary references

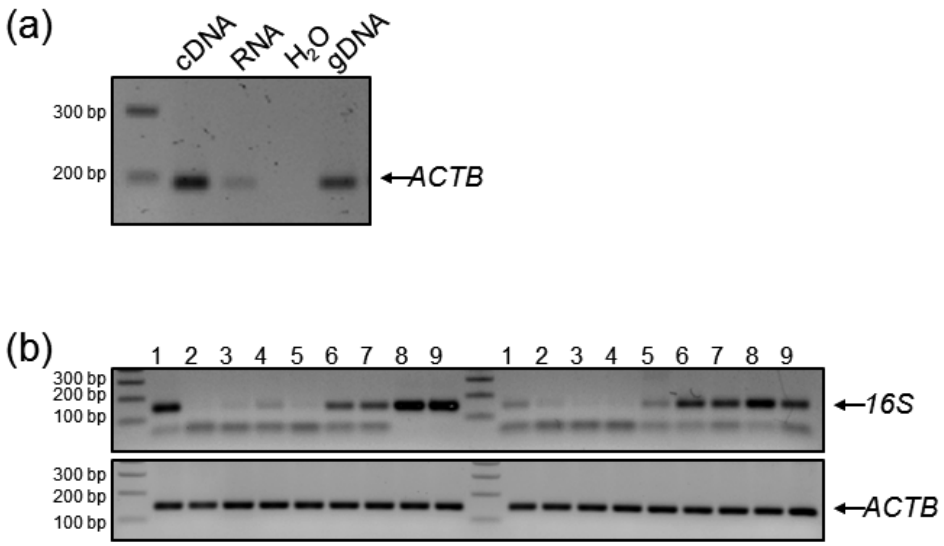
1. Suzuki MT, Taylor LT and DeLong EF. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl Environ Microbiol, 2000. 66(11): p. 4605-14.
2. Rodes L, Saha S, Tomaro-Duchesneau C and Prakash S. Microencapsulated Bifidobacterium longum subsp. infantis ATCC 15697 favorably modulates gut microbiota and reduces circulating endotoxins in F344 rats. Biomed Res Int, 2014. 2014: p. 602832.
3. Furet JP, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S *et al.* Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol, 2009. 68(3): p. 351-62.
4. van Vliet AH, Kuipers EJ, Waidner B, Davies BJ, de Vries N, Penn CW *et al.* Nickel-responsive induction of urease expression in Helicobacter pylori is mediated at the transcriptional level. Infect Immun, 2001. 69(8): p. 4891-7.
5. van Vliet AH, Kuipers EJ, Stoof J, Poppelaars SW and Kusters JG. Acid-responsive gene induction of ammonia-producing enzymes in Helicobacter pylori is mediated via a metal-responsive repressor cascade. Infect Immun, 2004. 72(2): p. 766-73.
6. Kim JW, Kim JG, Chae SL, Cha YJ and Park SM. High prevalence of multiple strain colonization of Helicobacter pylori in Korean patients: DNA diversity among clinical isolates from the gastric corpus, antrum and duodenum. Korean J Intern Med, 2004. 19(1): p. 1-9.
7. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol, 2002. 3(7): p. RESEARCH0034.

Supplementary Table S2 | Baseline characteristics of individual study subjects

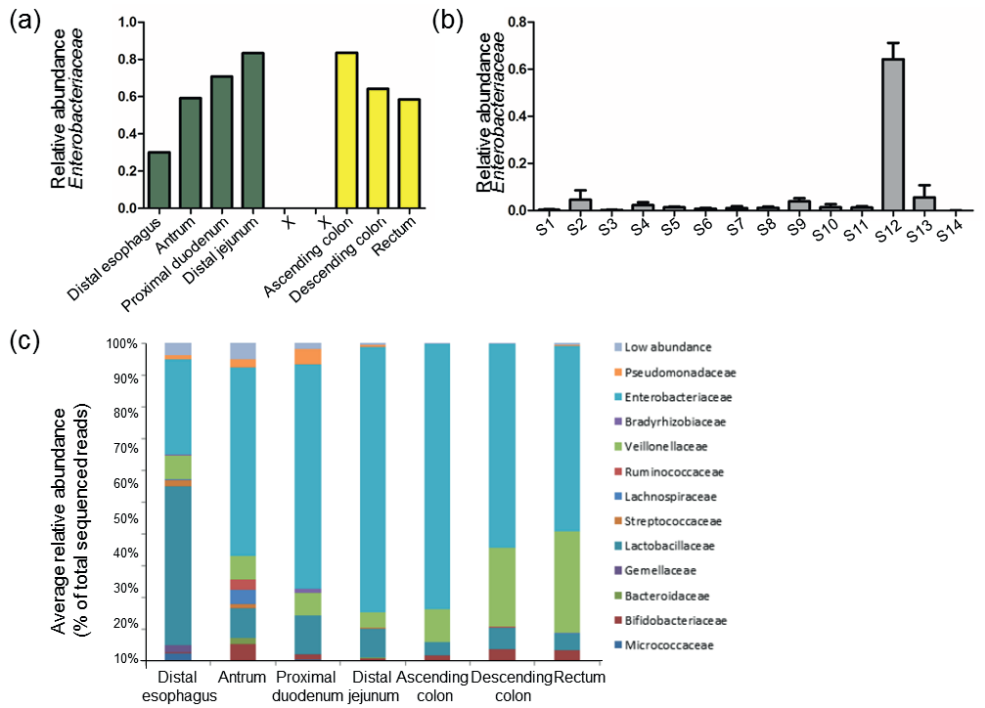
S[1]	Age[2]	Sex[3]	Race[4]	BMI[5]	Alcohol[6]	Smoking[7]	Medical history	Medication	Symptoms	DBE[8]	Pathology
1	50	M	C	20	4-5u / month	-	Resection small bowel (1964)	Methadone	Blood loss	Small bowel (proximal): multiple polyps	Peutz Jeghers polyps
2	38	F	C	21	-	15 /day	*BCRA1* gene mutation	Eye drops | Pancreatic enzymes	Diarrhea	-	-
3	44	F	?	?	?	?	-	-	Abdominal pain	-	-
4	61	F	C	24	3u /day	-	Atrial fibrillation	Antiplatelet drug | Acenocoumarol | Benzodiazepine | Paracetamol | PPI	Diarrhea Weight loss	Duodenum (proximal): polyp	-
5	26	F	C	?	-	1-10 /day	-	-	Abdominal pain	-	-
6	39	M	B	?	?	?	Diabetes type 2	Antiplatelet drug | Metformin | PPI |	Abdominal pain	-	-
7	67	F	C	16	-	-	Hepatitis B Liver transplantation (1997) COPD Osteopenia	Ciclosporin | Calcium/ Vitamin D | PPI	Iron deficiency Anemia	-	-
8	47	F	C	15	-	-	Fibromyalgia Endometriosis	Antispasmodic | Budesonide | H_2 antagonist	Diarrhea Weight loss Abdominal pain	-	-
9	54	M	C	32	6u /week	-	Hypertension Cholecystectomy Peripheral arterial disease	-	Diarrhea	Jejunum: erosive abnormalities	Reflux esophagitis
10	53	M	C	?	2u /day	35 /day	Cerebrovascular accident Stenosis of carotid artery	Antiplatelet drug | Cholesterol inhibitor | iron supplement | PPI	Iron deficiency Anemia	Venectasia	-

S[1]	Age[2]	Sex[3]	Race[4]	BMI[5]	Alcohol[6]	Smoking[7]	Medical history	Medication	Symptoms	DBE[8]	Pathology
11	50	M	C	21	2u /day	23 /day	Barrett's esophagus	Cholesterol inhibitor | PPI	Weight loss	Colon: small polyp	-
12	74	F	C	?	?	?	Obesity for which jejuno-ileal bypass surgery obesity	?	Iron deficiency Anemia	Distal ileum: extensive ulcerative abnormalities Cecum: obstructive mass*	Ulcerative lesion Tumor *
13	55	M	C	29	-	23 /day	Diabetes type 1 Hypertension	ACE inhibitor | Antiplatelet drug | Cholesterol inhibitor | Fludrocortisone | Hydrocortisone | insulin | PPI	Iron deficiency Anemia	-	-
14	62	M	?	27	1u /day	10 /day	Peripheral arterial disease	Antiplatelet drug | Cholesterol inhibitor | Iron supplements | PPI | Tramadol	Iron deficiency Anemia	-	-

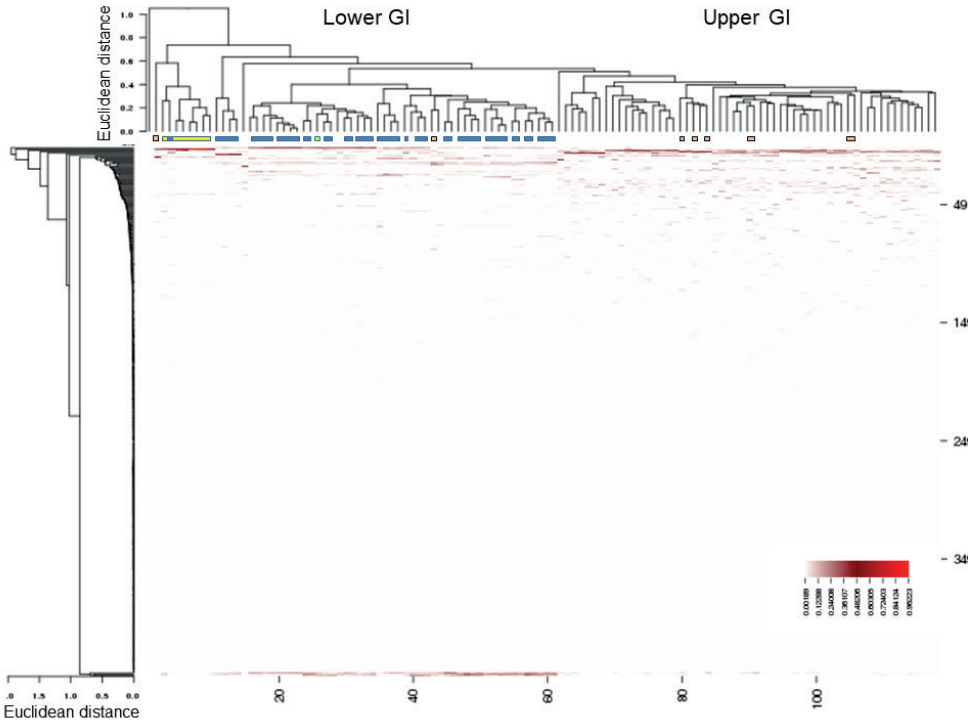
[1] Individual subjects numbered 1-14. [2] Age in years. [3] Gender: male (M) or female (F). [4] Race: Caucasian (C), black (B) or not available (?). [5] Body mass index (BMI). [6] Alcohol in units (u) per time [7] Smoking: number of cigarettes per day. [8] Double-balloon enteroscopy; * Of note: no biopsies were taken from the cecum tumor during the procedure, but instead during follow-up endoscopy. *Other abbreviations: ?, information not available; -. no specialities; COPD, chronic obstructive pulmonary disease; PPI, proton pump inhibitor.*



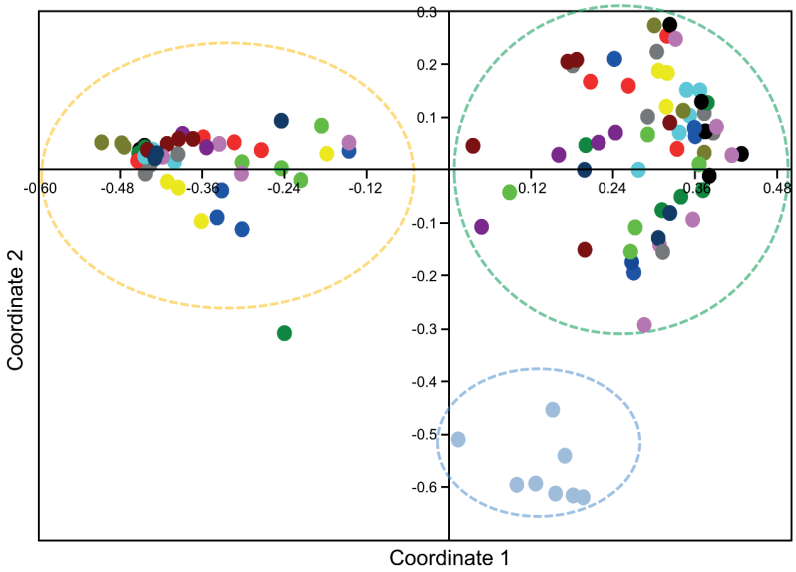
▲Supplementary Figure S1 | Bacterial abundance, unlike human genomic content, fluctuates along the intestinal tract. (a) Human *ACTB* gene primers identify the gene encoding *beta-actin* in both human copyDNA (cDNA) and genomic DNA (gDNA) isolated from human colorectal epithelial cancer cell lines CACO2. **(b)** Two representative examples of comparison of bacterial DNA (16S) and human DNA (ACTB) along the intestinal tract from two subjects (S1 and S2). 1: distal esophagus; 2: antrum; 3: proximal duodenum; 4: distal jejunum, 5: proximal ileum; 6: distal ileum; 7: ascending colon; 8: descending colon; 9: rectum.



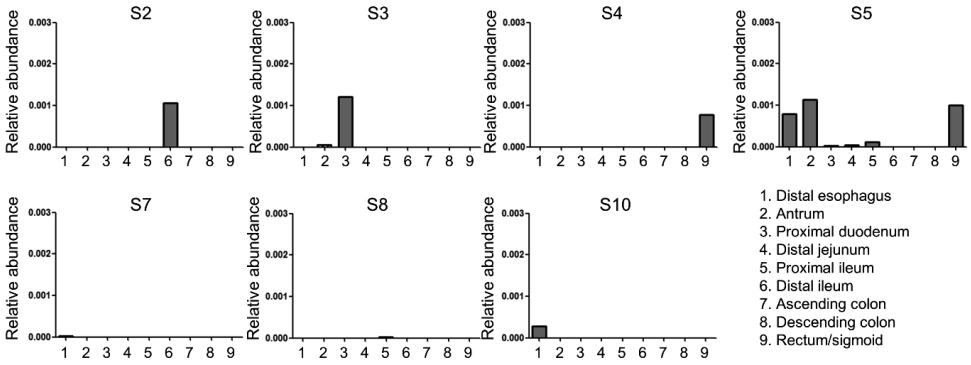
▲**Supplementary Figure S2 | Abundance of Enterobacteriaceae at family level along the gastrointestinal tract. (a)** Relative abundance of enterobacteriaceae in mucosal biopsies from a patient with a cecum tumor (S12) is shown. X: missing sample. Green: upper gastrointestinal locations. Yellow: lower gastrointestinal locations. **(b)** Comparison of abundance of Enterobacteriaceae at family level between patients, mean±SEM of all the GI locations are shown for subjects 1-14. **(c)** Most important bacteria at family level (>1% abundance) per location in patient with a cecum tumor.



▲Supplementary Figure S3 | Cluster analysis of taxonomy at family level demonstrating the clustering per patients and the upper and lower digestive tract. Samples indicated with yellow box were from subject 12, who was characterized by Enterobacteriaceae dominance. The utmost left sample was a *Helicobacter*-dominated sample from subject 6 (indicated in orange). Blue boxes indicate clustering of two or more samples from one individual patient.



▲Supplementary Figure S4 | Principal coordinate analysis (PCoA) plot of Bray-Curtis dissimilarity. Similar to Figure 4, but now the different coloured dots represent different patients. Egg blue dots circled in blue indicate subject S12, dominated by Enterobacteriaceae. In the left cluster (lower GI, circled in yellow), individual patient samples appear to lie closer together than in the right cluster (upper GI, circled in green).



▲Supplementary Figure S5 | Relative abundance of Helicobacter species across the different gastrointestinal (GI) sites. Helicobacter was detected in nine subjects. The relative abundance of Helicobacteraceae as detected by sequencing are shown here for individual GI locations of 7 subjects. Subjects S6 and S14 are shown in Figure 7.

4

CHAPTER 4

Technical challenges regarding the use of formalin-fixed paraffin embedded (FFPE) tissue specimens for the detection of bacterial alterations in colorectal cancer

S.Y. Lam*, A. Ioannou*, P. Konstanti, T. Visseren, M. Doukas, M.P. Peppelenbosch, C. Belzer, G.M. Fuhler
* These authors contributed equally

[1] Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands; [2] Laboratory of Microbiology, Wageningen University & Research Wageningen, The Netherlands; [3] Department of Pathology, Erasmus MC University Medical Center Rotterdam, The Netherlands.

Submitted

Abstract

Background: Formalin-fixed paraffin embedded (FFPE) tissues may provide an exciting resource to study microbial associations in human disease, but the use of these low biomass specimens remains challenging. We aimed to reduce unintentional bacterial interference in molecular analysis of FFPE tissues and investigated the feasibility of conducting quantitative polymerase chain reaction (qPCR) and 16S rRNA amplicon sequencing using 14 colorectal cancer, 14 normal adjacent and 13 healthy control tissues.

Results: Bacterial contaminants from the laboratory environment and the co-extraction of human DNA can affect bacterial analysis. The application of undiluted template improves bacterial DNA amplification, allowing the detection of specific bacterial markers (*Escherichia coli* and *Faecalibacterium prausnitzii*) by qPCR. Nested and non-nested PCR-based 16S rRNA amplicon sequencing approaches were employed, showing that bacterial communities of tissues and paired paraffin controls cluster separately at genus level on weighted UniFrac in both non-nested (R^2=0.045; Pr(>F)=0.053) and nested (R^2=0.299; Pr(>F)=0.001) PCR datasets. Nevertheless, considerable overlap of bacterial genera within tissues was seen with paraffin, DNA extraction negatives (non-nested PCR) or PCR negatives (nested PCR). Following mathematical decontamination, no differences in α- and β diversity were found between tumor, normal adjacent and control tissues.

Conclusion: Bacterial marker analysis by qPCR seems feasible using non-normalized template, but 16S rRNA amplicon sequencing remains challenging. Critical evaluation of laboratory procedures and incorporation of positive and negative controls for bacterial analysis of FFPE tissues are essential for quality control and to account for bacterial contaminants.

Keywords:

DNA contamination; gastrointestinal microbiome; high-throughput nucleotide sequencing; formalin-fixed paraffin embedded; low biomass, colorectal neoplasms.

1 Background

The preservation of formalin-fixed paraffin embedded (FFPE) tissue samples in the archives of health institutes has facilitated the study of human disease worldwide. In contrast to prospectively collected fresh and frozen material, FFPE tissue specimens are readily available to investigate a variety of health-related issues [1, 2]. Pathology archives are also an exciting potential source of information to answer microbe-related health questions. Both bacterial and viral deoxyribonucleic acid (DNA) can be detected in FFPE tissue specimens and have been used to investigate associations between invading pathogens and diseases, for instance to determine the presence of *Helicobacter pylori* (HP) in gastric adenocarcinoma [3] as well as hepatocellular carcinoma [4] and human papilloma virus (HPV) in cervical cancer [5]. Since innovative technologies have enabled the identification of microbiota and their genomes (microbiome) in the different niches of the human body [6, 7], FFPE tissue specimens might serve as an additional source to map these communities. Research questions which require the investigation of specific disease sites, rare diseases or a long follow-up time of patients may in particular benefit from the use of long-term collection and storage of FFPE tissue material. Examples of such studies include the investigation of colorectal cancer (CRC)-specific microbial composition [8] and the exploration of intestinal bacterial communities in neonates with necrotising enterocolitis [9-11].

Nevertheless, the application of FFPE tissues for microbiome analyses is associated with several challenges. First, obtaining sufficient quantities of genomic DNA of good quality remains difficult [12-15]. Neutrally buffered formalin prevents total DNA degradation [1, 15], but DNA cross-linking and fragmentation [15, 16] as well as storage time post-fixation [13, 17] impair the recovery of nucleic acids. Whereas the amplification of large DNA fragments is considered problematic due to DNA integrity deterioration [14, 15], shorter fragments have been used for molecular analyses [13, 16-19], even in samples archived for over twenty years [20]. Secondly, FFPE tissues have relatively high human genomic DNA content and are considered low bacterial biomass samples. Amplification steps such as nested polymerase chain reaction (PCR) may improve specificity and sensitivity of detected bacteria. Thirdly, microbial contaminants were shown to be present in commonly used reagents and can critically influence microbiome results, especially in low bacterial biomass samples [21-26]. Since bacterial contamination affects both 16S ribosomal RNA (rRNA) amplicon sequencing and shotgun metagenomics [21], careful handling of samples is essential during bacterial DNA retrieval and subsequent molecular analysis.

In this study, our aim is to optimise a method to reduce the interference of non-informative microbial contaminants in order to extract biologically relevant information from FFPE tissue specimens. To investigate the feasibility of conducting microbial analyses, we employed a cohort of 41 FFPE specimens to explore microbial associations in CRC using bacterial marker analysis and 16S rRNA amplicon sequencing. We report the difficulties encountered in an effort to optimize the processing of FFPE tissue specimens for future microbial studies.

2 Material and Methods

2.1 FFPE tissue and paraffin collections

FFPE colonic tissue specimens (n=8) were used for our initial 16S rRNA amplicon sequencing pilot. A total of 41 FFPE tissue specimens, containing CRC (n=14), paired normal adjacent tissue (n=14) and healthy colonic tissue (n=13), were collected for bacterial marker analysis and 16S rRNA amplicon sequencing. Microscopic findings were confirmed by an expert gastrointestinal (GI) pathologist. All FFPE tissue blocks were processed in neutral-buffered formalin and embedded with paraffin during routine medical practice and obtained from the department of pathology at the Erasmus MC University Medical Center Rotterdam, the Netherlands. In addition, paraffin was sampled from six sources including two batches of paraffin grains, one tissue processor machine and three paraffin embedding stations and transferred into clean autoclaved bottles until bacterial DNA isolation.

2.2 Microtome sectioning

A regular cleaning protocol was applied for processing samples of our 16S rRNA amplicon sequencing pilot, entailing the use of ethanol to clean the microtome and metal tweezers before sectioning. For each block, 14 consecutive sections of 5μM were obtained after disposal of the superficial layers and transferred in autoclaved 1.5 ml Eppendorf tubes for storage until bacterial DNA isolation. A second more stringent contamination-prevention protocol including DNA-Zap treatment of all surfaces and the use of facemasks and flow cabinets was employed to process following specimens (**Supplementary methods 1**). To control for potential contamination in downstream analysis, paired empty paraffin (0.05 gram) from the same FFPE tissue block was collected using a sterile disposable surgical knife and a clean weighing scale. All specimens were cut within several days of each other, placed in Eppendorf tubes and transferred into a dark box for storage in a cold room to prevent degradation by light/heat until DNA isolation within a few weeks. All specimens were cut within several days of each other, placed in Eppendorf tubes and transferred into a dark box for storage in a cold room to prevent degradation by light/heat until DNA isolation within a few weeks.

2.3 Bacterial DNA extraction

Bacterial DNA isolation of FFPE tissues and paraffin was carried out with the RTP Bacteria DNA Mini Kit (STRATEC Molecular Gm, Berlin, Germany) according to the manufacturer's protocol for FFPE material. The first step was modified using xylene to dissolve paraffin. Melted paraffin (100µl) from six pathology sources served as starting material for DNA isolation with two RTP Bacteria Mini Kits. Autoclaved water (500µl) and/or blank (no template) samples that were processed by the RTP Bacteria DNA Mini Kit served as additional controls.

For bacterial marker analysis and 16S rRNA amplicon sequencing, FFPE tissue and their paired empty paraffin samples were concurrently processed for DNA extraction in a non-specific order using a stringent decontamination protocol (**Supplementary methods 1**). To prevent repetitive pipetting steps, each DNA sample was divided in aliquots of which one was utilized for 16S rRNA amplicon sequencing. DNA purity was measured with NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA) and concentration with Qubit dsDNA BR Assay Kit (Thermo Fisher). Samples were stored at -20˚C until further analysis.

2.4 Polymerase chain reaction

All samples and potential contamination sources, e.g. individual components of three individual RTP Bacteria DNA Mini Kits, extraction additives xylene, ethanol and isopropanol and paraffin were subjected to PCR amplification. PCR assays were executed with an Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Waltham, MA) using primers targeting the *16S rRNA* gene, the human *beta-actin (ACTB)* gene and *Ralstonia* species (**Supplementary Table S1**). Each PCR reaction contained GoTaq® buffer (Promega, Madison, WI), 1.25mM $MgCl_2$ (Promega), 0.167mM (each) deoxynucleotides (Roche Diagnostics, Mannheim, Germany), 2.5U GoTaq®polymerase (Promega), 333nM of each primer (Sigma-Aldrich, St Louis, MO), 2µl of template and water to a final volume of 30µl. After 4 minutes of denaturation at 95˚C, 40 cycles consisting of 30 seconds denaturation at 95˚C, 30 seconds annealing and 1 minute extension at 72˚C were applied, and followed by the final extension of 10 minutes at 72˚C. Template was not normalized (undiluted) or normalized to 10ng/µl where otherwise specified. Water served as negative PCR control and positive controls were fecal bacterial DNA, human genomic DNA from FFPE tissues and known *Ralstonia*-contaminated elution buffer. Amplicons were visualized by gel electrophoresis using 2% agarose gel in 1X TBE (Tris-borate-EDTA) buffer containing Serva DNA stain G (Promega).

2.5 Quantitative polymerase chain reaction assays

Primer details for qPCR assays are described in **Supplementary Table S1**. To determine bacterial versus human genomic DNA concentration within FFPE tissue samples (n=3), a standard curve with equimolar *Escherichia coli* (*E. coli*) and human genomic DNA was prepared (**Supplementary methods 2**). The reaction mixture comprised of SYBR Select Master Mix (Applied Biosystems), 200-500nM of each primer (Sigma-Aldrich), 2µl of normalized template (10ng/µl) and water for a total volume of 20µl and DNA was amplified using the same cycles as described above. The bacterial versus human genomic DNA concentration were calculated using their respective standard curves and illustrated with Graph Pad Prism 5 software (GraphPad, San Diego, CA). Additionally, *16S rRNA* gene copy numbers in paired tissue (n=39) and empty paraffin (n=38) samples were calculated and groups analyzed with the Wilcoxon test.

Bacterial marker analysis was performed with *E. coli* and *ClbA* gene primers to detect *E. coli* and CRC associated genotoxic strains carrying the pathogenicity island *pks* [27, 28], respectively. Gut commensal *Faecalibacterium prausnitzii (F. prausnitzii)*, which has been reported to be negatively associated with CRC [29, 30], was additionally selected. To account for different FFPE tissue sizes, the *ACTB* gene was measured. PCR conditions were similar as described above, except for the use of 4µl non-normalized template to enhance amplification. The $2^{-\Delta\Delta CT}$ method was applied to calculate the fold change. The ΔCT_{sample} (= $CT_{bacterial\ target} - CT_{ACTB\ target}$) was first obtained for each sample by normalization to the amount of total human DNA. The average ΔCT_{sample} of the healthy tissues (control group) was then used to calculate ΔΔCT (= ΔCT_{sample} – average $\Delta CT_{control\ group}$), after which the fold change derived from $2^{-\Delta\Delta CT}$. The Kruskal-Wallis with the Dunn's Multiple Comparison test for post-hoc analysis were performed in Graph Pad Prism 5.

2.6 Library preparation and 16S rRNA amplicon sequencing

16S rRNA amplicon sequencing was performed at the Macrogen Institute, Seoul, Korea, using amplification of the 16S rRNA hypervariable region V3-V4 by 341F/805R primers. Libraries included paired empty paraffin controls (n=6) and DNA extraction negatives (n=6) for sequencing on the Illumina MiSeq platform (2x300 bp) (Illumina, San Diego, CA). Secondly, a nested PCR approach was applied in house using 27F/1369R and 515F/806R primers for respectively the initial and a subsequent PCR targeting the V4 region. Paired empty paraffin (n=41), synthetic bacterial mock communities (n=2) and PCR negative controls (=2) were concurrently processed with the tissues. DNA extraction negatives from the aforementioned RTP Bacteria Mini Kits (n=6) were additionally prepared to allow

in depth comparison of these controls with the non-nested data set. Sequencing was conducted on the Illumina NovaSeq 6000 platform (2x150 bp) at GATC Biotech (Konstanz, Germany). More details about library preparation and primer sequences are described in **Supplementary methods 3** and **Supplementary Table S2**.

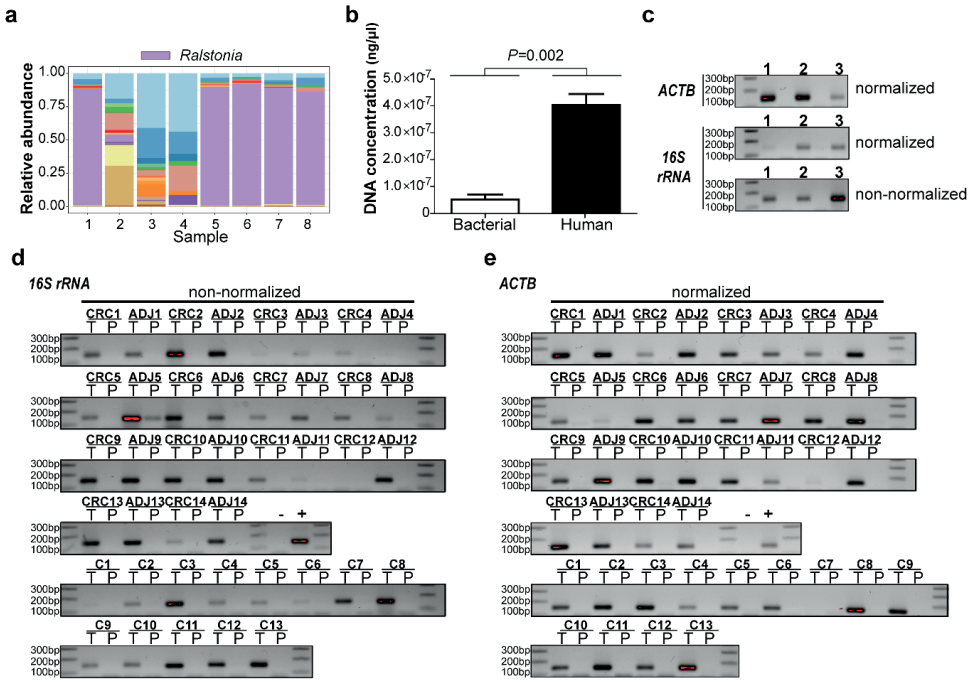
2.7 16S rRNA amplicon sequencing data processing

For both the 16S rRNA amplicon sequencing pilot and the two larger data sets (non-nested and nested PCR approaches), quality control of the reads was performed with FASTQC [31] in Java Runtime Environment and Rqc package [32] in R version 3.5.0 [33]. The NG-Tax pipeline with default settings was applied [34, 35]. The operational taxonomic unit (OTU) table was constructed at 0.1% abundance threshold, unassigned reads with one mismatch included and chimeras removed. Taxonomic assignment was conducted with the USEARCH algorithm [36] against the Silva SSU 128 database [37]. Further analysis was performed in R with the 'phyloseq' [38], 'microbiome' [39] and 'vegan' [40] packages. Group comparison was conducted based on sample type (tissue, paraffin, controls). Alpha-diversity was computed with Shannon and Chao1 Indices while β-diversity was assessed with Principal Coordinates Analysis (PCoA) based on weighted UniFrac and Bray-Curtis dissimilarity at genus and OTU level after relative abundance transformation of the data. The 'adonis' permutational multivariate analyses of variance (PERMANOVA) was applied to determine statistical significance between groups. The 'betadisper' function was used to test for multivariate homogeneity of groups dispersions [41]. To ensure reproducibility, the seed was set to 995 for both permutations tests. The biological significance of the data sets was reassessed following removal of contaminants identified by the prevalence method of the 'decontam' package [42]. A 0.5 threshold was set and negative controls consisted of DNA extraction negatives for the non-nested PCR dataset and PCR negatives for the nested PCR approach.

3 Results

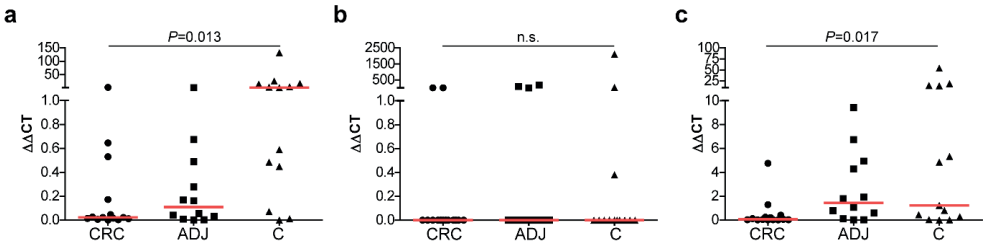
3.1 Bacterial and human DNA interference in microbial analyses of FFPE tissue specimens

Interference from contaminants is a common problem for samples of low microbial biomass. The 16S rRNA amplicon sequencing pilot results showed a predominance of *Ralstonia* in five out of eight samples (**Figure 1a**). Retrospective analysis demonstrated the elution buffer as the contaminating source while other extraction reagents were excluded (**Supplementary Figure S1a**). Differences in bacrial DNA detection were observed for the



▲Figure 1 | Unintentional bacterial and human genomic DNA interference in molecular analyses of formalin-fixed paraffin embedded (FFPE) tissues. (a) 16S rRNA amplicon sequencing pilot results showing the relative abundance of the top 20 genera in FFPE tissue samples (n=8). A predominance of genus *Ralstonia* is observed in a majority of samples. **(b)** Quantitative polymerase chain reaction (qPCR) results demonstrating the amount of bacterial versus human genomic DNA in normalized template (n=3). Statistical significance was calculated using the unpaired t-test. **(c)** Gel electrophoresis results following *actin beta* (*ACTB*) and *16S rRNA* gene amplification, emphasizing the use of non-normalized (undiluted) template to improve bacterial DNA detection in FFPE tissue samples (n=3). **(d-e)** Gel electrophoresis results showing bacterial and human genomic DNA presence in 41 paired FFPE tissue (T) samples and their paired empty paraffin (P) controls. Tissues consist of colorectal cancer (CRC; n=14), normal adjacent (ADJ; n=14) and healthy control (C; n=13) tissues. Full length gel electrophoresis results are shown in Additional Figure A1.

same set of pathology paraffin collections using two separate kits (**Supplementary Figure S1b**), indicating that the extent of contamination varies per newly opened kit and its components (**Supplementary Figure S1c**), while paraffin itself is not a contaminating source. Nevertheless, bacterial DNA presence was confirmed in the majority of 41 FFPE tissues and was minor in paired paraffin controls (**Figure 1d**), with quantification of the bacterial biomass showing significantly higher *16S rRNA* gene copies numbers (<10^5 per µl of DNA) in tissues (n=39) compared to matched paraffin samples (n=38) (**Supplementary Figure S1d**). However, human genomic DNA is co-extracted with bacterial DNA and present in higher concentrations in FFPE tissues (*P*=0.002) (**Figure 1b, e**), and therefore normalization of samples for DNA concentration may hamper bacterial DNA detection



▲Figure 2 | Quantitative polymerase chain reaction (qPCR) analysis of specific bacterial markers in formalin-fixed paraffin embedded (FFPE) tissues. qPCR results of bacterial markers in colorectal cancer (CRC; n=14), normal adjacent (ADJ; n=12) and healthy colonic control (C; n=13) using non-normalized template. The relative amount of each sample is normalized to human genomic DNA and expressed in terms of fold change ($2^{-\Delta\Delta CT}$). **(a-b)** Detection of *Escherichia coli (E. coli)* (a) and strains harboring the *pks*+ island (b) using *E. coli* and *ClbA* gene specific primers, respectively. **(c)** Detection of *Faecalibacterium prausnitzii*.

4

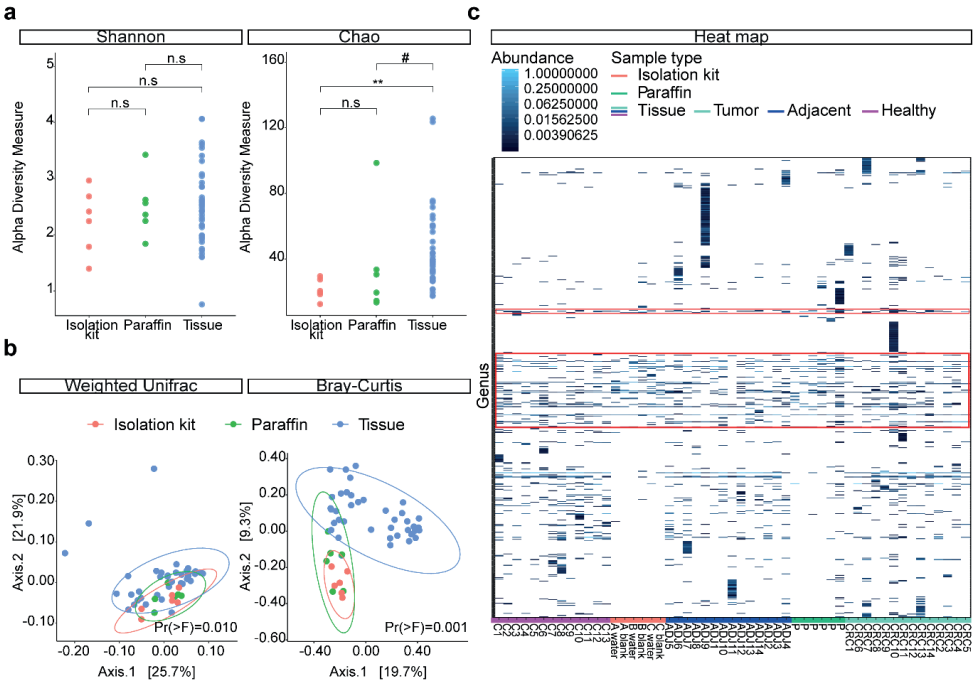
(**Figure 1c**). Thus, human and microbial contaminants may interfere in microbial analysis of low biomass samples and should be accounted for.

3.2 Bacterial marker analysis of FFPE tissue specimens by qPCR

The use of non-normalized template allowed the comparison of bacterial markers in CRC (n=14), normal adjacent tissues (n=12) and healthy tissues (n=13). The fold change ($2^{-\Delta\Delta CT}$) levels of *E. coli* were significantly different among groups (*P*=0.013), in particular CRC compared to healthy controls (**Figure 2a**). A minority of tissue samples harbored *pks* positive strains, but no differences were detected among groups (**Figure 2b**). The levels of *F. prausnitzii* differed between tissue types (*P*=0.017) with post-hoc analysis indicating significant higher levels in healthy controls compared to CRC (**Figure 2c**).

3.3. Exploring bacterial communities in FFPE tissues, paraffin and controls with two sequencing approaches

First, we employed the most commonly used sequencing approach, with amplification of the bacterial 16S rRNA hypervariable region V3-V4. A total of 41 FFPE tissue samples, six paraffin controls and six DNA extraction negatives were included for sequencing. No differences in Shannon diversity were found at genus level, but the Chao1 diversity index was significantly higher in tissues than DNA extraction negatives (*P*<0.010) (**Figure 3a**). The bacterial communities clustered separately based on weighted UniFrac (R2=0.088; Pr(>F)=0.010) and Bray-Curtis dissimilarity (R2=0.116; Pr(>F)=0.001), as shown by the PCoA plots at genus level (**Figure 3b**). Pairwise comparisons indicated that tissues were distinct from DNA extraction negatives on weighted UniFrac (R2=0.063; Pr(>F)=0.017) and Bray-



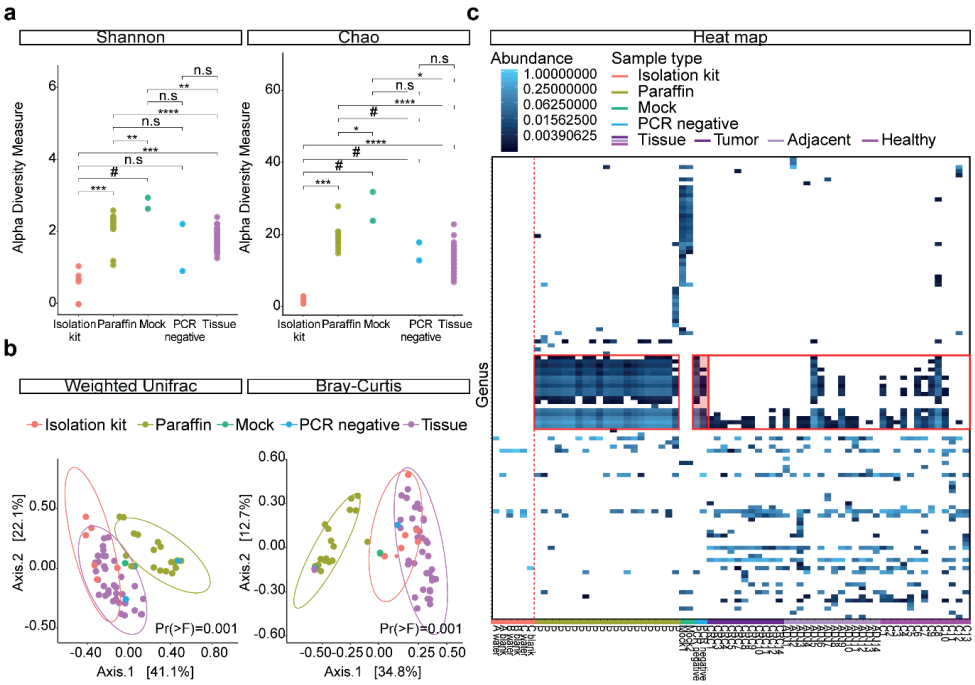
▲Figure 3 | Bacterial communities of tissue, paraffin and DNA extraction negatives following a targeted *16S rRNA* gene amplification and sequencing approach. The 16S rRNA amplicon sequencing results of formalin-fixed paraffin embedded (FFPE) tissue (n=41), empty paraffin controls (n=6) and DNA extraction controls (n=6) at genus level using a non-nested polymerase chain reaction (PCR) approach. **(a)** Dot plots show the bacterial biodiversity measured by Shannon and Chao1 diversity indices. **(b)** Principal Coordinates Analysis (PCoA) plots illustrate the bacterial composition of tissue, paraffin and DNA extraction controls using weighted UniFrac and Bray-Curtis dissimilarity. **(c)** Heat map of the relative abundance of genera across samples. Tissue specimens consist of tumor (n=14), normal adjacent tissue (n=14) and healthy colonic tissue (n=13). The red box indicates genera present in DNA extraction negatives and other samples. Abbreviations for level of significance: n.s. not significant; #, $P<0.1$; **, $P<0.01$.

Curtis dissimilarity (R2=0.081; Pr(>F)=0.001). Tissues also differed from paraffin when computed with Bray-Curtis dissimilarity (R2=0.066; Pr(>F)=0.001) while a trend was observed for weighted (R2=0.045; Pr(>F)=0.053). Homogeneity conditions were met and findings were in agreement at OTU level (**Supplementary Table S3**). The heat map demonstrated a considerable overlap of genera among tissues, paraffin and DNA extraction negatives (**Figure 3c**). A total of 53 bacterial families were detected in these latter controls, indicating significant interference from contaminants/artefacts derived during DNA isolation, library preparation and sequencing procedures (**Supplementary Figure S2a**).

A second approach using nested PCR for bacterial DNA amplification was performed on 38 FFPE tissues, 21 paraffin controls, two mock communities, six DNA extraction negatives (samples undergoing DNA isolation procedure) and two PCR negatives (samples not undergoing DNA isolation, i.e. PCR/sequencing controls). Shannon and Chao1 diversity indices were higher in tissues than DNA extraction negatives (P<0.001 and P<0.0001), but lower compared to paraffin (both P<0.0001) (**Figure 4a**). No differences in α-diversity were observed between tissues and PCR negatives. Moreover, the community structure among the five groups were significantly different based on weighted UniFrac (R2=0.374; Pr(>F)=0.001) (**Figure 4b**). Tissues were distinct from paraffin (R2=0.299; Pr(>F)=0.001) and DNA extraction negatives (R2=0.155; Pr(>F)=0.001), but not from PCR negatives (R2=0.055; Pr(>F)=0.056). Also paraffin did not differ from these PCR controls (R2=0.061; Pr(>F)=0.234). When using Bray-Curtis dissimilarity for overall group comparison, bacterial communities clustered separately (R2=0.361; Pr(>F)=0.001), albeit heterogeneous dispersion (Pr(>F)=0.004) was noticed (**Figure 4b**). Tissues differed from DNA extraction negatives (R2=0.089; Pr(>F)=0.001) and PCR negatives (R2=0.054; Pr(>F)=0.020), but pairwise comparison with paraffin did not meet the condition of homogenous dispersion. No significant differences were found between paraffin and PCR negatives (R2=0.072; Pr(>F)=0.219). Analyses were also performed at OTU level, which showed similar results (**Supplementary Table S3**). Multiple bacterial genera were shared between tissues, paraffin and PCR negatives, but not with DNA extraction negatives and mock controls (**Figure 4c**). Six bacterial families were found in DNA extraction negatives (**Supplementary Figure S2a**). To estimate the accuracy of the experimental procedure and the pipeline, Pearson correlation for theoretical and experimental mock communities was calculated. The correlation was r=0.923 (P=0.064) and r=0.952 (P=0.064) for the mock controls, respectively (**Supplementary Figure S2b**).

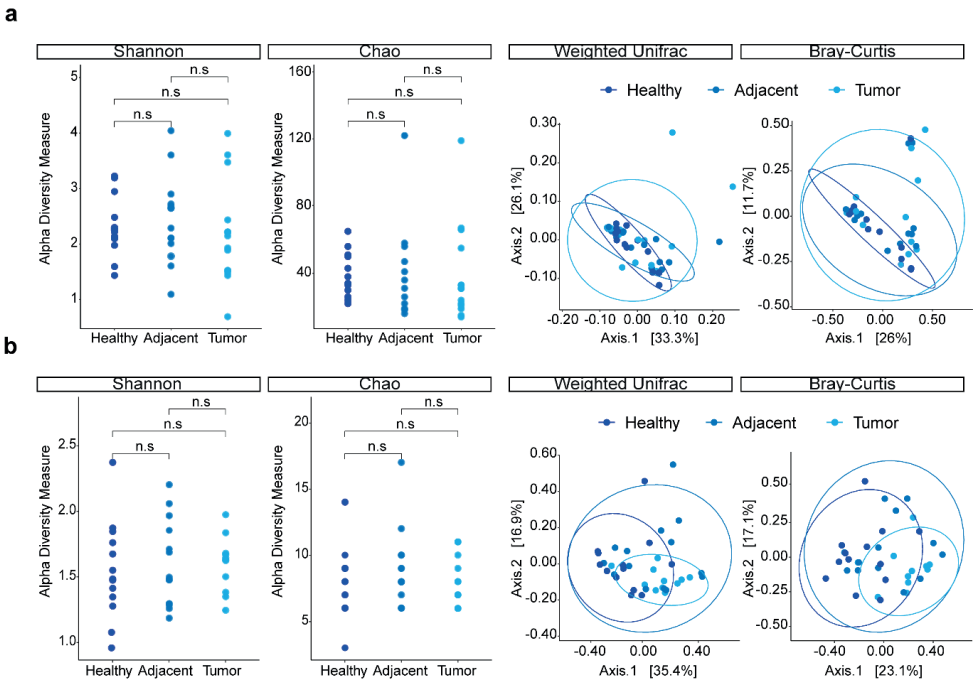
3.4 Retrieving biologically relevant information from FFPE samples remains challenging

The library size of both datasets comprised of low read numbers overall for FFPE tissues which overlapped with DNA extraction negatives in the non-nested PCR data set and PCR negatives in the nested PCR dataset (**Supplementary Figure S3a,b**). These controls were set for the identification of contaminants using the prevalence method at a threshold of 0.5 that was selected based on its discriminative ability (**Supplementary Figure S3c,d)** [42]. This resulted in the removal of 1.006.017 (13.8%) and 5.464.548 (22.4%) reads corresponding to the biopsy samples in respect tively the non-nested and nested PCR datasets. The remaining 1684 OTUs within the non-nested PCR dataset following decontamination



▲Figure 4 | Bacterial communities of tissue, paraffin and controls following a nested *16S rRNA* gene amplification and sequencing approach. The 16S rRNA amplicon sequencing results of formalin-fixed paraffin embedded (FFPE) tissue (n=38), empty paraffin controls (n=21), DNA extraction controls (n=6), polymerase chain reaction (PCR) negatives (n=2) and synthetic mock communities (n=2) at genus level using a nested PCR approach. **(a)** Dot plots show the bacterial biodiversity measured by Shannon and Chao1 diversity indices. **(b)** Principal Coordinates Analysis (PCoA) plots illustrate the bacterial composition of tissue, paraffin and the different controls using weighted UniFrac and Bray-Curtis dissimilarity. **(c)** Heat map analysis of the relative abundance of genera across samples. Tissue specimens consist of tumor (n=11), normal adjacent tissue (n=14) and healthy colonic tissue (n=13). The red box indicates the genera present in PCR negatives and other samples. The red dotted line separates the DNA extraction negatives that were processed separately during library preparation. Abbreviations for level of significance: n.s. not significant; #, $P<0.1$; *, $P<0.05$; **, $P<0.01$; ***, $P<0.001$; ****, $P<0.0001$.

belonged to 42 phyla including Proteobacteria (30.6%), Firmicutes (29.2%) and Bacteroidetes (14.4%) while the nested PCR data set comprised of 440 OTUs from six phyla including Proteobacteria (45.2%), Firmicutes (41.8%) and Actinobacteria (8.2%) in descending order (**Supplementary Figure S4a,b**). Deinococcus-Thermus from both data sets and many spurious others from the non-nested PCR data set (e.g. Planctomycetes) are generally not seen in fecal samples [43], which might hamper the retrieval of biologically relevant information. Nevertheless, the detection of *Faecalibacterium* and *Escherichia-Shigella* species within tissue samples was indeed possible by sequencing. After decontamination, there were no differences in α- and β diversity between tumor, normal adjacent and healthy



▲Figure 5 | Bacterial diversity of tumor, normal adjacent and healthy control tissues following decontamination. (a-b) The α- and β diversity results of formalin-fixed paraffin embedded (FFPE) tissues at genus level following decontamination are shown for the datasets obtained by a non-nested (a) and a nested (b) polymerase chain reaction (PCR) approach. The dot plots show the bacterial biodiversity measured by Shannon and Chao1 diversity indices. Principal Coordinates Analysis (PCoA) plots illustrate the bacterial composition of tissue using weighted UniFrac and Bray-Curtis dissimilarity. Abbreviations for level of significance: n.s. not significant.

tissues in the non-nested dataset at genus level (**Figure 5a**). The nested dataset failed to meet the homogeneity condition of PERMANOVA, thus rendering dispersion as the possible reason for significant differences in β-diversity (**Figure 5b**).

4 Discussion

FFPE tissue specimens may provide an important source to study the microbiota in health and disease, but their use is associated with several technical challenges. Our study demonstrates that despite the use of specialised bacterial DNA isolation kits, normalization of low biomass FFPE tissue samples is primarily driven by human DNA rather than bacterial DNA presence. Our study also shows that the use of non-normalized (undiluted) template may improve bacterial detection in downstream bacterial analyses. Although bacterial marker analysis by qPCR was feasible for two selected markers (**Figure**

2), high throughput analyses would require hundreds of individual qPCR assays and sequencing efforts may thus prove to be more efficient. Nevertheless, we show that the extraction of biological relevant information from 16S rRNA amplicon sequencing data remains difficult. Our sequencing pilot was distorted by a previously reported contaminant, *Ralstonia* [21, 23, 24, 44], underscoring the critical impact of bacterial contamination on low biomass samples as described by others [21-26]. We applied and recommend stringent measures for processing FFPE tissues to reduce the chance of contamination (**Supplementary Figure S5**), but it should be noted that PCR negatives not undergoing DNA extraction also showed significant presence of bacterial DNA, which were either introduced during PCR and sequencing efforts, or the result of sequencing artefacts arising as a result of low biomass (**Figure 4c**). In particular the detection of biologically relevant taxa such as the Bifidobacteriaceae, Lactobacillaceae and Enterobacteriaceae families in negative controls (**Supplementary Figure S2a**) can reduce the robustness concerning the presence of such taxa in samples. Since negative controls demonstrate both the nature and the source of contamination, quality control at different steps of sequencing efforts is particularly recommended. The inclusion of paraffin controls allowed us to demonstrate that bacterial communities of tissues and paraffin tend to cluster separately at genus level in both datasets, and therefore paraffin seems less informative regarding to contaminants. However, the incorporation of DNA extraction negatives and PCR negatives showed that these low biomass samples (tissue and paraffin) were highly affected by bacterial contaminants and/or sequencing artefacts, in line with recent literature [25, 45]. Synthetic mock controls with high bacterial biomass were less affected, but future studies should consider mock controls with both high and low concentration [46]. Controls with similar microbial biomass as the experimental samples would be representative for the actual effect of contamination and the loss of specific signal due to sequencing efforts.

Knowledge of the characteristics of the investigated ecosystem and the introduction of controls is important for the critical appraisal of results [23]. The DNA quality obtained from FFPE tissue specimens allowed bacterial marker analysis by qPCR. In addition to the detection of *E. coli* and *ClbA* gene positive strains, the finding of lower *F. prausnitzii* levels in CRC compared to healthy controls is in accordance to previous findings [29, 30]. Our sequencing efforts indeed indicated the presence of *Escherichia* spp. and *Faecalibaterium* spp. in a part of the FFPE tissues. The finding of Proteobacteria as the major phylum in both datasets is not fully understood, but in line with a recent study suggesting that the paraffin embedding process might influence the microbial profile [47]. Multiple aspects in different processes involving pre-processing of tissues, storage conditions, DNA isolation,

library preparation and actual sequencing are known to influence the outcome [21, 24-26]. While FFPE tissue specimens have previously been used for 16S rRNA amplicon sequencing [8, 11] and also for shotgun metagenomics study recently [48], data retrieval is limited compared to frozen tissues [47]. Nevertheless, FFPE samples are sometimes the only available source to answer research questions, allowing complementary taxonomic and functional exploration of the microbiome despite relatively low read counts [48]. Thus, when using this material, we highly recommend the identification of prominent contaminating sources to increase the robustness of the dataset's biological information.

Our study has several limitations. First, our bacterial amplification approaches were not directly comparable due to different library preparations and sequencing platforms. Although PCR negatives and mock controls were not available in both sequencing efforts, their findings together emphasized the importance to include both positive and negative controls to account for bacterial contamination. Secondly, the retrieval of biological relevant information from low abundance bacterial DNA of questionable quality and quantity remains challenging. The maximum DNA fragment size detectable was not determined for each of our DNA samples, but it might be helpful to guide the experimental set-up in future studies using FFPE tissues. Although the exploration of alternative tissue fixation and isolation protocols was also out of our scope, bacterial marker analysis by qPCR with FFPE tissue specimens was possible here. Thirdly, the decontamination process leads to loss of both data in general and possibly rare bacterial taxa in low bacterial biomass samples, which should be taken into consideration by researchers when extrapolating such results. Lastly, individual laboratory reagents for library preparation were not tested, but the inclusion PCR negatives are essential to account for procedural related contamination and to interpret results. The application of enzymatic treatment of PCR master mixes has been suggested [24], and might be considered in future efforts.

5 Conclusion

Our study with FFPE tissue specimens has stressed the importance to implement measures against bacterial contamination in microbiome research with low bacterial biomass samples. Since human genomic DNA is being co-extracted from FFPE tissues, the use of non-normalized (undiluted) template is recommended for bacterial detection in downstream molecular analyses. While the execution of 16S rRNA amplicon sequencing on FFPE tissue specimens remains difficult, the inclusion of negative controls (e.g. DNA extraction negatives and PCR negatives) and positive controls (e.g. synthetic mock communities) is important for quality control, and the use of only one source of

contamination control may not be sufficient. Future microbiome studies with low biomass specimens should critically evaluate laboratory procedures to account for bacterial contamination.

References

1. Nagahashi M, Shimada Y, Ichikawa H, Nakagawa S, Sato N, Kaneko K *et al*. Formalin-fixed paraffin-embedded sample conditions for deep next generation sequencing. J Surg Res, 2017. 220: p. 125-132.
2. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA and Baloch ZW. Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. Biopreserv Biobank, 2013. 11(2): p. 101-6.
3. Han HS, Lee KY, Lim SD, Kim WS and Hwang TS. Molecular identification of Helicobacter DNA in human gastric adenocarcinoma tissues using Helicobacter species-specific 16S rRNA PCR amplification and pyrosequencing analysis. Oncol Lett, 2010. 1(3): p. 555-558.
4. Rabelo-Goncalves E, Roesler B, Guardia AC, Milan A, Hara N, Escanhoela C *et al*. Evaluation of five DNA extraction methods for detection of H. pylori in formalin-fixed paraffin-embedded (FFPE) liver tissue from patients with hepatocellular carcinoma. Pathol Res Pract, 2014. 210(3): p. 142-6.
5. Lagheden C, Eklund C, Kleppe SN, Unger ER, Dillner J, and Sundstrom K. Validation of a standardized extraction method for formalin-fixed paraffin-embedded tissue samples. J Clin Virol, 2016. 80: p. 36-9.
6. Ursell LK, Metcalf JL, Parfrey LW and Knight R. Defining the human microbiome. Nutr Rev, 2012. 70 Suppl 1: p. S38-44.
7. Lynch SV and Pedersen O. The Human Intestinal Microbiome in Health and Disease. N Engl J Med, 2016. 375(24): p. 2369-2379.
8. Bundgaard-Nielsen C, Baandrup UT, Nielsen LP and Sorensen S. The presence of bacteria varies between colorectal adenocarcinomas, precursor lesions and non-malignant tissue. BMC Cancer, 2019. 19(1): p. 399.
9. Smith B, Bodé S, Petersen BL, Jensen TK, Pipper C, Kloppenborg J *et al*. Community analysis of bacteria colonizing intestinal tissue of neonates with necrotizing enterocolitis. BMC microbiology, 2011. 11: p. 73-73.
10. Ullrich T, Tang YW, Correa H, Garzon SA, Maheshwari A, Hill M *et al*. Absence of gastrointestinal pathogens in ileum tissue resected for necrotizing enterocolitis. Pediatr Infect Dis J, 2012. 31(4): p. 413-4.
11. Stewart CJ, Fatemizadeh R, Parsons P, Lamb CA, Shady DA, Petrosino JF *et al*. Using formalin fixed paraffin embedded tissue to characterize the preterm gut microbiota in necrotising enterocolitis and spontaneous isolated perforation using marginal and diseased tissue. BMC microbiology, 2019. 19(1): p. 52-52.
12. Gilbert MT, Haselkorn T, Bunce M, Sanchez JJ, Lucas SB, Jewell LD *et al*. The isolation of nucleic acids from fixed, paraffin-embedded tissues-which methods are useful when? PLoS One, 2007. 2(6): p. e537.
13. Lin J, Kennedy SH, Svarovsky T, Rogers J, Kemnitz JW, Xu A *et al*. High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. Anal Biochem, 2009. 395(2): p. 265-7.
14. Farrugia A, Keyser C and Ludes B. Efficiency evaluation of a DNA extraction and purification protocol on archival formalin-fixed and paraffin-embedded tissue. Forensic Sci Int, 2010. 194(1-3): p. e25-8.
15. Klopfleisch R, Weiss AT and Gruber AD. Excavation of a buried treasure--DNA, mRNA, miRNA and protein analysis in formalin fixed, paraffin embedded tissues. Histol Histopathol, 2011. 26(6): p. 797-810.
16. Bonin S and Stanta G. Nucleic acid extraction methods from fixed and paraffin-embedded tissues in cancer diagnostics. Expert Rev Mol Diagn, 2013. 13(3): p. 271-82.
17. Kotorashvili A, Ramnauth A, Liu C, Lin J, Ye K, Kim R *et al*. Effective DNA/RNA co-extraction for analysis of microRNAs, mRNAs, and genomic DNA from formalin-fixed paraffin-embedded specimens. PLoS One, 2012. 7(4): p. e34683.

18. Impraim CC, Saiki RK, Erlich HA and Teplitz RL. Analysis of DNA extracted from formalin-fixed, paraffin-embedded tissues by enzymatic amplification and hybridization with sequence-specific oligonucleotides. Biochem Biophys Res Commun, 1987. 142(3): p. 710-6.
19. Dedhia P, Tarale S, Dhongde G, Khadapkar R and Das B. Evaluation of DNA extraction methods and real time PCR optimization on formalin-fixed paraffin-embedded tissues. Asian Pac J Cancer Prev, 2007. 8(1): p. 55-9.
20. Gillio-Tos A, De Marco L, Fiano V, Garcia-Bragado F, Dikshit R, Boffetta P *et al*. Efficient DNA extraction from 25-year-old paraffin-embedded tissues: study of 365 samples. Pathology, 2007. 39(3): p. 345-8.
21. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF *et al*. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol, 2014. 12: p. 87.
22. Glassing A, Dowd SE, Galandiuk S, Davis B and Chiodini RJ. Inherent bacterial DNA contamination of extraction and sequencing reagents may affect interpretation of microbiota in low bacterial biomass samples. Gut Pathog, 2016. 8: p. 24.
23. de Goffau MC, Lager S, Salter SJ, Wagner J, Kronbichler A, Charnock-Jones DS *et al*. Recognizing the reagent microbiome. Nat Microbiol, 2018. 3(8): p. 851-853.
24. Stinson LF, Keelan JA and Payne MS. Identification and removal of contaminating microbial DNA from PCR reagents: impact on low-biomass microbiome analyses. Lett Appl Microbiol, 2019. 68(1): p. 2-8.
25. Karstens L, Asquith M, Davin S, Fair D, Gregory WT, Wolfe AJ *et al*. Controlling for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. mSystems, 2019. 4(4).
26. Dahlberg J, Sun L, Persson Waller K, Östensson K, McGuire M, Agenäs S *et al*. Microbiota data from low biomass milk samples is markedly affected by laboratory and reagent contamination. PLoS One, 2019. 14(6): p. e0218257.
27. Pleguezuelos-Manzano C, Puschhof J, Rosendahl Huber A, van Hoeck A, Wood HM, Nomburg J *et al*. Mutational signature in colorectal cancer caused by genotoxic pks(+) E. coli. Nature, 2020. 580(7802): p. 269-273.
28. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ *et al*. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science, 2012. 338(6103): p. 120-3.
29. Balamurugan R, Rajendiran E, George S, Samuel GV and Ramakrishna BS. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, Desulfovibrio and Enterococcus faecalis in the feces of patients with colorectal cancer. J Gastroenterol Hepatol, 2008. 23(8 Pt 1): p. 1298-303.
30. Lopez-Siles M, Martinez-Medina M, Surís-Valls R, Aldeguer X, Sabat-Mir M, Duncan SH *et al*. Changes in the Abundance of Faecalibacterium prausnitzii Phylogroups I and II in the Intestinal Mucosa of Inflammatory Bowel Disease and Patients with Colorectal Cancer. Inflamm Bowel Dis, 2016. 22(1): p. 28-41.
31. Andrews S. FastQC: a quality control tool for high trhougput sequence data 2010.
32. de Souza W, Carvalho BdS and Lopes-Cendes I. Rqc: A Bioconductor Package for Quality Control of High-Throughput Sequencing Data. 2018, 2018. 87(Code Snippet 2): p. 14.
33. R Core Team. R: A language and environment for statistical computing. 2008.
34. Ramiro-Garcia J, Hermes GDA, Giatsis C, Sipkema D, Zoetendal EG, Schaap PJ *et al*. NG-Tax, a highly accurate and validated pipeline for analysis of 16S rRNA amplicons from complex biomes. F1000Res, 2016. 5: p. 1791.
35. Poncheewin W, Hermes GDA, van Dam JCJ, Koehorst JJ, Smidt H, and Schaap PJ. NG-Tax 2.0: A Semantic Framework for High-Throughput Amplicon Analysis. Frontiers in Genetics, 2020. 10(1366).

36. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 2010. 26(19): p. 2460-1.
37. 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. 41(Database issue): p. D590-6.
38. McMurdie PJ and Holmes S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One, 2013. 8(4): p. e61217.
39. Lahti L and Shetty S. Tools for microbiome analysis in R. Bioconductor, 2017.
40. Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D *et al*. Vegan: community ecology package. 2019.
41. Anderson MJ. Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 2006. 62(1): p. 245-53.
42. Davis NM, Proctor DM, Holmes SP, Relman DA and Callahan BJ. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome, 2018. 6(1): p. 226.
43. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR *et al*. Enterotypes of the human gut microbiome. Nature, 2011. 473(7346): p. 174-80.
44. Barton HA, Taylor NM, Lubbers BR and Pemberton AC. DNA extraction from low-biomass carbonate rock: an improved method with reduced contamination and the low-biomass contaminant database. J Microbiol Methods, 2006. 66(1): p. 21-31.
45. Erb-Downward JR, Falkowski NR, D'Souza JC, McCloskey LM, McDonald RA, Brown CA *et al*. Critical Relevance of Stochastic Effects on Low-Bacterial-Biomass 16S rRNA Gene Analysis. mBio, 2020. 11(3).
46. Caruso V, Song X, Asquith M and Karstens L. Performance of Microbiome Sequence Inference Methods in Environments with Varying Biomass. mSystems, 2019. 4(1).
47. Pinto-Ribeiro I, Ferreira RM, Pereira-Marques J, Pinto V, Macedo G, Carneiro F *et al*. Evaluation of the Use of Formalin-Fixed and Paraffin-Embedded Archive Gastric Tissues for Microbiota Characterization Using Next-Generation Sequencing. Int J Mol Sci, 2020. 21(3).
48. Debesa-Tur G, Pérez-Brocal V, Ruiz-Ruiz S, Castillejo A, Latorre A, Soto JL *et al*. Metagenomic analysis of formalin-fixed paraffin-embedded tumor and normal mucosa reveals differences in the microbiome of colorectal cancer patients. Sci Rep, 2021. 11(1): p. 391.

Supplementary information

Supplementary methods

1 Protocol to process formalin-fixed paraffin embedded (FFPE) tissue specimens for bacterial analysis

Microtome sectioning of FFPE tissue specimens entailed the use of disposable surgical gowns, gloves and face masks. Prior to and between sectioning of each sample, the microtome was cleaned with ethanol and DNA *Zap*™ PCR DNA Degradation Solutions (Thermo Fisher Scientific Inc., Waltham, MA), rinsed with autoclaved demineralized water and wiped dry with surgical gauze. Blades and gloves were consistently changed between samples to prevent cross-contamination and sterile disposable surgical tweezers were used to place serial sections into autoclaved 1.5 ml Eppendorf tubes. Bacterial DNA isolation was carried out in a laminar flow cabinet to minimize environmental and aerogenic contamination, except for the first step which required a fume hood for xylene. All pipetting work was performed using Filter Tips after dissembling and thorough cleaning of the pipettes. Similar to other laboratory equipment, treatment with ethanol and DNA *Zap* solutions was followed by ultraviolet (UV) radiation in the laminar flow cabinet. Disposable surgical gowns and face masks were used and gloves were regularly changed when handling samples.

2 Standard curve for quantitative polymerase chain reaction (qPCR) analyses

A kanamycin-resistance gene containing *Escherichia coli* (*E. coli*) strain (DH5α) was cultured at 37˚C in Luria Broth (LB) medium containing Kanamycin (50µg/ml) and DNA was isolated according to the protocol of the RTP Bacteria DNA Mini Kit (STRATEC Molecular Gm, Berlin, Germany) for bacterial cell pellets. Human DNA was extracted from a gastric epithelial cell (GES) line using the Wizard DNA Purification kit (Promega, Madison, WI). Both *E. coli* and human samples were subjected to the StepOnePlus Real-Time PCR system (Applied Biosystems, Waltham, MA) for PCR amplification with subsequent gel electrophoresis. The *E. coli* and human amplification products were purified from the agarose gel with the Monarch DNA Gel Extraction Kit (New England Biolabs Inc., MA), normalized to a concentration of 10ng/µl and combined in equivalent volumes to create a subsequent five 1:10 serial dilution with autoclaved water.

3 Library preparation and 16S ribosomal RNA (rRNA) gene sequencing

The V3-V4 region of the 16S rRNA gene was amplified at the Macrogen Institute using 341F-805R primers (**Supplementary Table S2**), which were attached to Illumina overhang

adapters (forward: 5'-TCGTCGGCAGCGT CAGATGTGTATAAGAGACAG-3'; reverse 5'-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-3'). The reaction mixture consisted of 12.5µl 2x KAPA HiFi HotStart Ready Mix (Kapa Biosystems, Cape Town, South Africa), 200nM of each primer and 2.5µl of non-normalized template for a total volume of 25µl. After 3 minutes of initial denaturation at 95˚C, 25 cycles of 30 seconds denaturation at 95˚C, 30 seconds of extension at 55˚C and 30 seconds of elongation at 72˚C took place, and ended with 5 minutes of final extension at 72˚C. Following PCR clean-up with AMPure XP beads (Agencourt Bioscience Corporation, Beverly, MA), multiplexing indices and Illumina sequencing adapters were attached in an index PCR using the Nextera XT Index Kit (Illumina, San Diego, CA). After 8 cycles under the above described PCR conditions, the reaction mixture was purified with the AMPure XP beads. The final library was normalized and sequenced on the MiSeq system (Illumina) to produce 2x300bp end reads.

A nested PCR approach was conducted in house using consecutive 27F/1369R and 515F/806R primer pairs (**Supplementary Table S2**). The reaction mixture of the first PCR contained 10µl 5x Phusion HF buffer (ThermoFisher Scientific, Waltham, MA), 0.5µl Phusion Hot Start II DNA polymerase (ThermoFisher Scientific), 200µM deoxynucleotides, 100nM of each primer, 1µl non-normalized template and 36.5µl water for a total volume of 50µl. After 30 seconds of initial denaturation at 98˚C, 25 cycles of 10 seconds denaturation at 98˚C, 10 seconds annealing at 50˚C and 30 seconds extension at 72˚C were applied, and followed by 10 minutes of final extension at 72˚C. After amplicon purification with the DNA Clean & Concentrator Kit (Zymo Research, Irvine, CA), 1µl template was subjected to a second PCR targeting the V4 region of the 16S rRNA gene using primers which were tagged with a barcode at the 5' end for recognition. PCR amplification was performed in triplicate under the following conditions: 30 seconds of denaturation at 98˚C, 25 cycles of 10 seconds denaturation at 98˚C, 10 seconds annealing at 50˚C and 10 seconds of extension at 72˚C, and final extension of 10 minutes at 72˚C. Amplicons were visualized on 1% agarose gel for verification, pooled and purified using the CleanPCR kit (CleanNA, Waddinxveen, the Netherlands). Only amplicons of FFPE tissue and paraffin controls samples that were verified on agarose gel were included for sequencing on the Illumina NovaSeq 6000 platform (Illumina) to produce 2x150bp paired-end reads at GATC Biotech (Konstanz, Germany).

Supplementary Table S1 | Overview of primers used for (quantitative) polymerase chain reaction analysis

Target	**Sequence (5'-3')**		**T_a** (°C)	**Size** (bp)	**Ref**
Ralstonia species[a]	Forward	CTGGGGTCGATGACGGTA	56	546	1
	Reverse	ATCTCTGCTTCGTTAGTGGC			
ACTB gene[a,b,c]	Forward	CTGGAACGGTGAAGGTGACA	60.5	140	2
	Reverse	AAGGGACTTCCTGTAACAATGCA			
16S rRNA gene[a,b]	Forward	CGGTGAATACGTTCCCGG	60	145	3-4
	Reverse	TACGGCTACCTTGTTACGACTT			
16S rRNA gene[d]	Forward	GTGSTGCAYGGYYGTCGTCA	52	147	5
	Reverse	ACGTCRTCCMCNCCTTCCTC			
E.coli[c]	Forward	CATGCCGCGTGTATGAAGAA	62	96	6
	Reverse	CGGGTAACGTCAATGAGCAAA			
E.coli (ClbA gene)[c]	Forward	ATGAGGATTGATATATTAATTGGACA	58	233	7-8
	Reverse	GGTTTGCCATATTTGCACGTAC			
F. prausnitzii[c]	Forward	GATGGCCTCGCGTCCGATTAG	58	198	9
	Reverse	CCGAAGACCTTCTTCCTCC			

[a] End-point PCR [b] qPCR to determine human versus bacterial DNA concentrations [c] qPCR for bacterial marker analysis [d] qPCR to determine 16S rRNA copy numbers.

Abbreviations: ACTB, beta-actin; E. coli, Escherichia coli; F. prausnitzii, Faecalibacterium prausnitzii; Ref, reference; T_a, annealing temperature.

Supplementary references

1. Coenye T, Spilker T, Reik R, Vandamme P and Lipuma JJ. Use of PCR analyses to define the distribution of Ralstonia species recovered from patients with cystic fibrosis. J Clin Microbiol, 2005. 43(7): p. 3463-6.
2. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol, 2002. 3(7): p. RESEARCH0034.
3. Suzuki MT, Taylor LT and DeLong EF. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl Environ Microbiol, 2000. 66(11): p. 4605-14.
4. Furet JP, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S *et al*. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol, 2009. 68(3): p. 351-62.
5. Maeda H, Fujimoto C, Haruki Y, Maeda T, Kokeguchi S, Petelin M *et al*. Quantitative real-time PCR using TaqMan and SYBR Green for Actinobacillus actinomycetemcomitans, Porphyromonas gingivalis, Prevotella intermedia, tetQ gene and total bacteria. FEMS Immunol Med Microbiol, 2003. 39(1): p. 81-6.
6. Huijsdens XW, Linskens RK, Mak M, Meuwissen SG, Vandenbroucke-Grauls CM, and Savelkoul PH. Quantification of bacteria adherent to gastrointestinal mucosa by real-time PCR. J Clin Microbiol, 2002. 40(12): p. 4423-7.
7. Prorok-Hamon M, Friswell MK, Alswied A, Roberts CL, Song F, Flanagan PK *et al*. Colonic mucosa-associated diffusely adherent afaC+ Escherichia coli expressing lpfA and pks are increased in inflammatory bowel disease and colon cancer. Gut, 2014. 63(5): p. 761-70.

8. Eklöf V, Löfgren-Burström A, Zingmark C, Edin S, Larsson P, Karling P *et al*. Cancer-associated fecal microbial markers in colorectal cancer detection. Int J Cancer, 2017. 141(12): p. 2528-2536.
9. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, and Macfarlane S. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. Appl Environ Microbiol, 2007. 73(22): p. 7435-42.

Supplementary Table S2 | Overview of primer sequences to prepare 16S ribosomal RNA amplicon sequencing libraries

Target	Sequence (5'-3')		Ref
341F-805R[a]	Forward	CCTACGGGNGGCWGCAG	1 - 2
	Reverse	GACTACHVGGGTATCTAATCC	
27F-1369R[b]	Forward	AGAGTTTGATCMTGGCTCAG	3 - 5
	Reverse	GCCCGGGAACGTATTCACCG	
515F-806R[b]	Forward	GTGCCAGCMGCCGCGGTAA	6 - 8
	Reverse	GGACTACHVHHHTWTCTAAT	

[a] Primers used at the Macrogen Institute [b] Primers used in house.

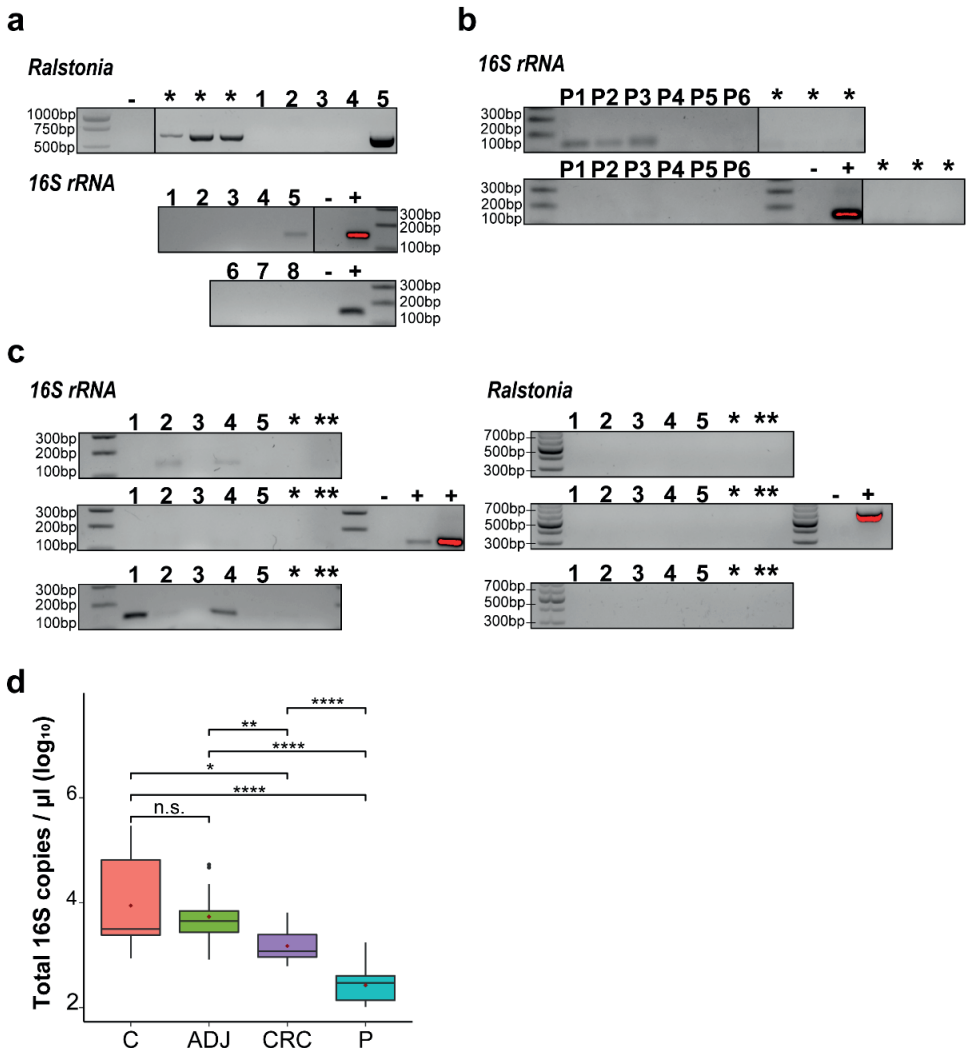
Abbreviations: Ref, reference.

Supplementary references

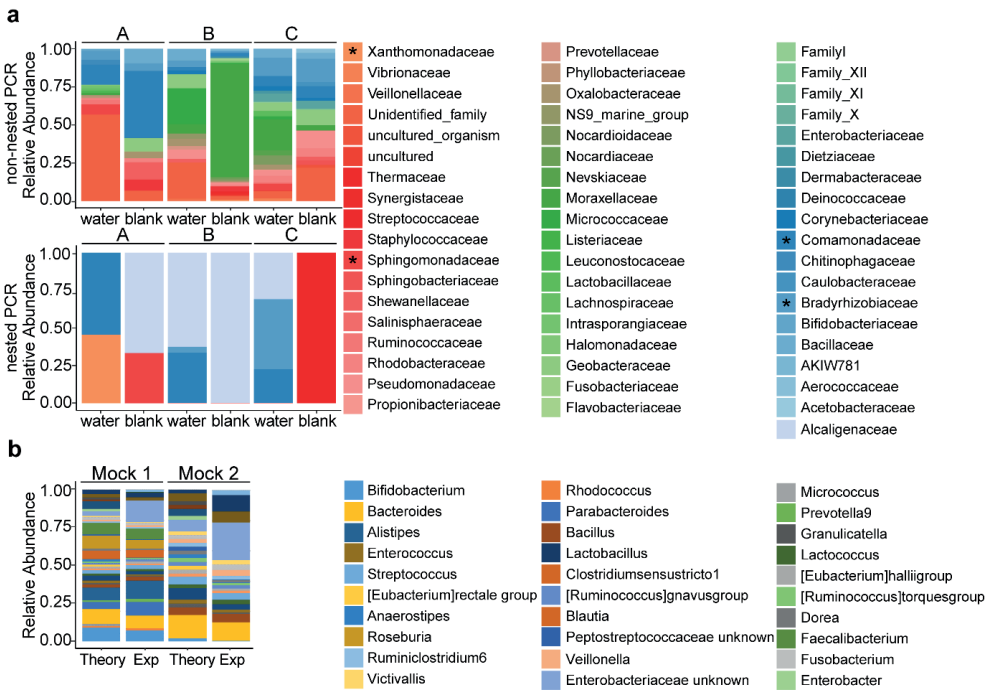
1. Herlemann DP, Labrenz M, Jurgens K, Bertilsson S, Waniek JJ, and Andersson AF. Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. Isme J, 2011. 5(10): p. 1571-9.
2. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M *et al*. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research, 2012. 41(1): p. e1-e1.
3. Weisburg WG, Barns SM, Pelletier DA and Lane DJ. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol, 1991. 173(2): p. 697-703.
4. Iwamoto T, Tani K, Nakamura K, Suzuki Y, Kitagawa M, Eguchi M *et al*. Monitoring impact of in situ biostimulation treatment on groundwater bacterial community by DGGE. FEMS Microbiol Ecol, 2000. 32(2): p. 129-141.
5. Yu Z and Morrison M. Comparisons of different hypervariable regions of rrs genes for use in fingerprinting of microbial communities by PCR-denaturing gradient gel electrophoresis. Appl Environ Microbiol, 2004. 70(8): p. 4800-6.
6. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ *et al*. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci U S A, 2011. 108 Suppl 1: p. 4516-22.
7. Apprill A, McNally S, Parsons R and Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol 2015, 75:129-137.
8 Parada, AE, Needham DM and Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 2016, 18(5), 1403–1414.

Supplementary Table S3 | Comparison of bacterial communities from different groups within non-nested and nested polymerase chain reaction (PCR) data sets on operational taxonomic unit (OTU) level

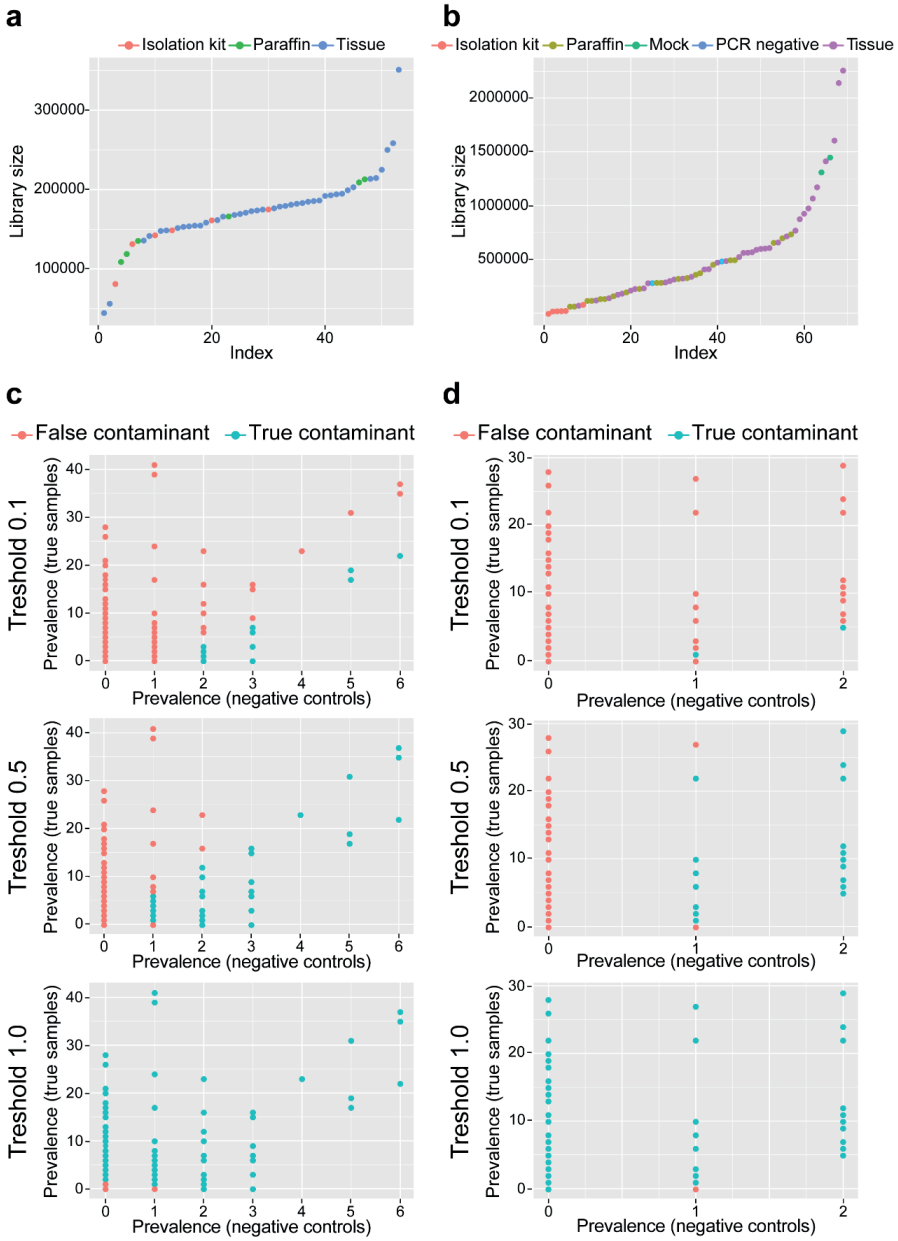
Non nested PCR dataset		**Weighted UniFrac**	**Bray-Curtis**
Overall	Pr(>F)	0.011	0.001
	R^2	0.086	0.109
	Homogeneity Pr(>F)	0.723	0.852
Tissues vs. paraffin	Pr(>F)	0.055	0.001
	R^2	0.045	0.062
	Homogeneity Pr(>F)	0.46	0.702
Tissues vs. DNA extraction negatives	Pr(>F)	0.019	0.001
	R^2	0.060	0.076
	Homogeneity Pr(>F)	0.733	0.737
Paraffin vs. DNA extraction negatives	Pr(>F)	0.565	0.319
	R^2	0.076	0.097
	Homogeneity Pr(>F)	0.644	0.303
Nested PCR dataset		**Weighted UniFrac**	**Bray-Curtis**
Overall	Pr(>F)	0.001	0.001
	R^2	0.374	0.345
	Homogeneity Pr(>F)	0.155	0.003
Tissues vs. paraffin	Pr(>F)	0.001	0.001
	R^2	0.299	0.289
	Homogeneity Pr(>F)	0.067	0.001
Tissues vs. PCR negatives	Pr(>F)	0.058	0.030
	R^2	0.055	0.049
	Homogeneity Pr(>F)	0.931	0.465
Tissue vs. DNA extraction negatives	Pr(>F)	0.001	0.001
	R^2	0.155	0.086
	Homogeneity Pr(>F)	0.482	0.394
Tissues vs. positive control	Pr(>F)	0.002	0.002
	R^2	0.103	0.101
	Homogeneity Pr(>F)	0.024	0.007
Paraffin vs. PCR negatives	Pr(>F)	0.234	0.208
	R^2	0.061	0.072
	Homogeneity Pr(>F)	0.701	0.354



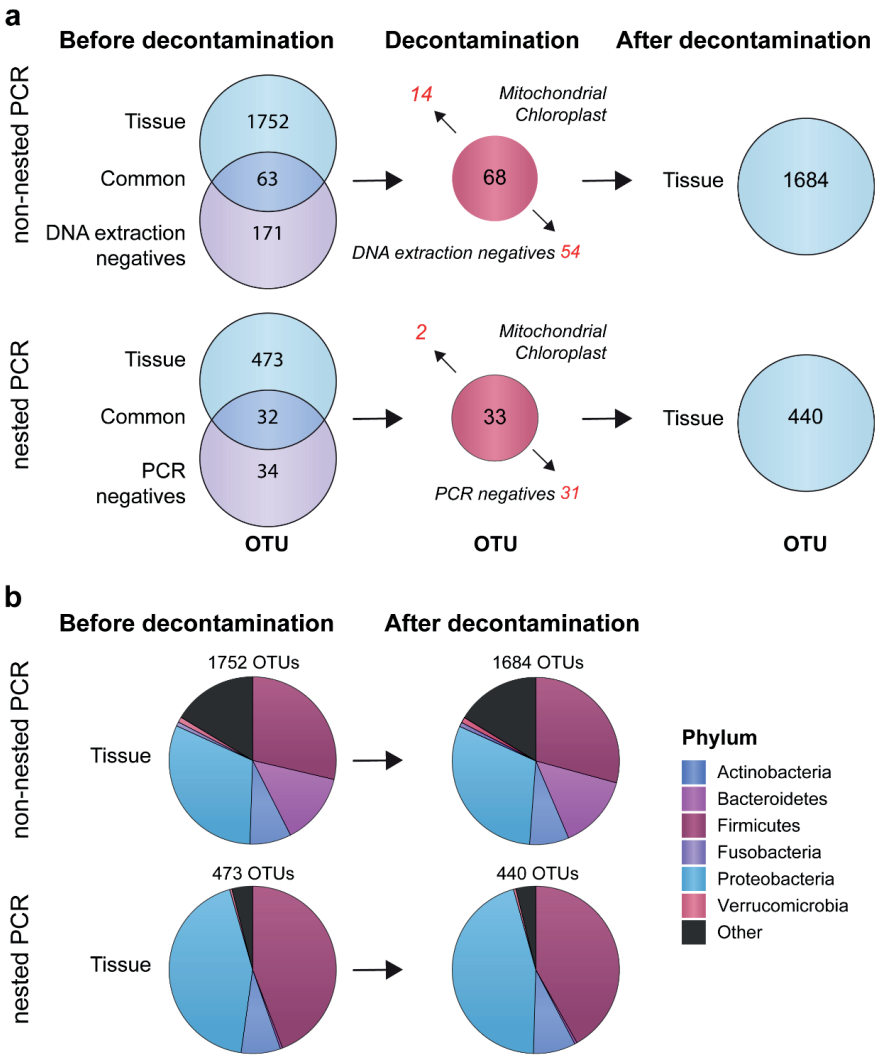
▲Supplementary Figure S1 | Bacterial analysis to examine extraction reagents and paraffin collections and to determine 16S rRNA gene copies in FFPE tissues. (a-c) Gel electrophoresis results following amplification with *Ralstonia* and/or *16S rRNA* gene primers to investigate bacterial presence in resuspension (1), binding (2), wash I-II (3-4) and elution (5) buffers of the RTP Bacteria DNA Mini Kit, extraction additives xylene (6), ethanol (7) and isopropanol (8), in addition to a set of paraffin collections (grains [P1-2], tissue processor machine [P3], embedding stations [P4-6]) processed by different kits. The individual gels in sections b and c represent the results obtained with two (b; upper and lower) and three (c; upper, middle and lower panel) different DNA isolation kits, respectively. The water (*) and blank (**) controls extracted with the DNA isolation kit were included, as well as positive (+) and negative (-) controls. **(d)** Quantitative polymerase chain reaction (qPCR) findings showing total *16S rRNA* gene copy numbers per µl DNA (log_{10}) in healthy tissue controls (C; n=12), normal adjacent tissue (ADJ; n=14), colorectal cancer (CRC; n=13) and paraffin controls (P; n=38). The mean value is shown for each group. Abbreviations for level of significance: n.s. not significant; *, $P<0.05$; **, $P<0.01$; ****, $P<0.0001$. Full length gel electrophoresis results are shown in Additional Figure A2.



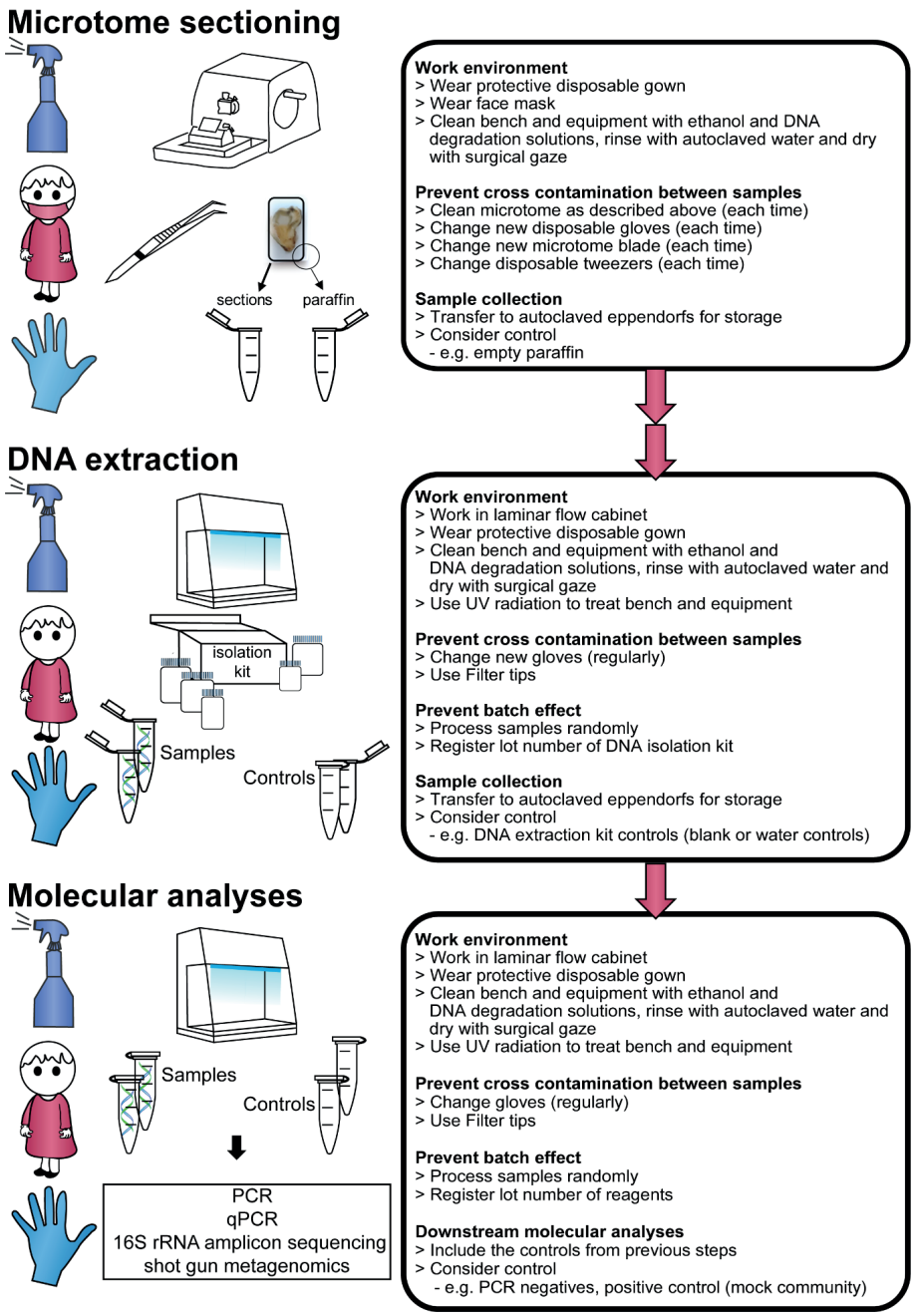
▲Supplementary Figure S2 | The relative abundance of bacterial families and genera in respectively DNA extraction negatives and mock controls. (a) The relative abundance plots at family level of DNA extraction negatives (n=6) that were processed with the different RTP Bacteria DNA Mini Kits A, B and C. The results of two different 16S rRNA amplicon sequencing approaches are shown. The non-nested PCR dataset demonstrated the presence of 53 bacterial families within the controls while a total of 6 families were detected in the nested PCR dataset. The asterisks (*) indicates bacterial families that were detected by both methods. **(b)** The relative abundance plot of theoretical (Theory) and experimental (Exp) mock controls of the nested dataset at genus level. The experimental mock communities underwent library preparation and 16S rRNA amplicon sequencing.



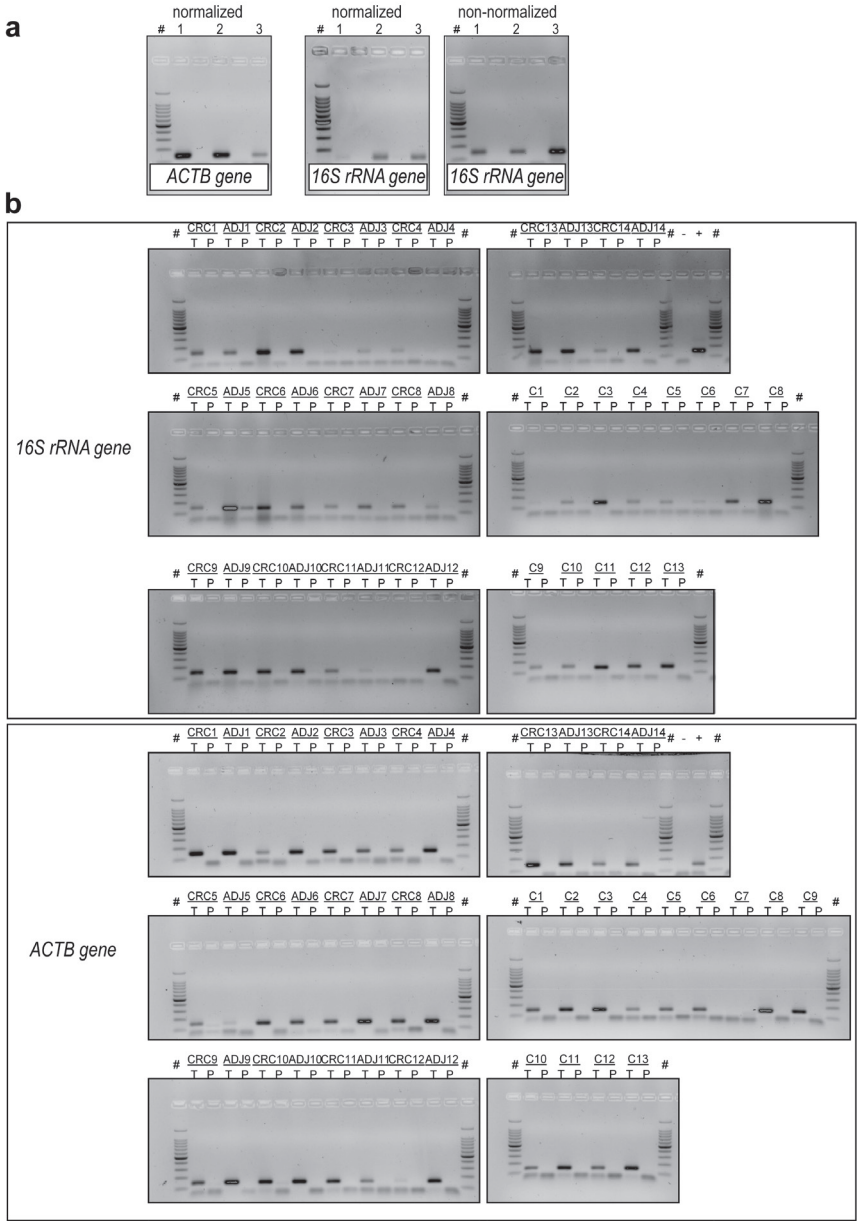
▲Supplementary Figure S3 | Library size and decontamination thresholds for the 16S rRNA amplicon sequencing data sets conducted with non-nested and nested polymerase chain reaction (PCR) approaches. (a,b) The library size of samples from the non-nested (a) and nested (b) PCR data sets **(c,d)** The plots represents the decontamination threshold 0.1 (default) and the more stringent thresholds 0.5 and 1.0 for the identification of contaminants by the prevalence method. The negative controls comprised of DNA extraction negatives in the non-nested data set (c) and PCR negatives in the nested data set (d).



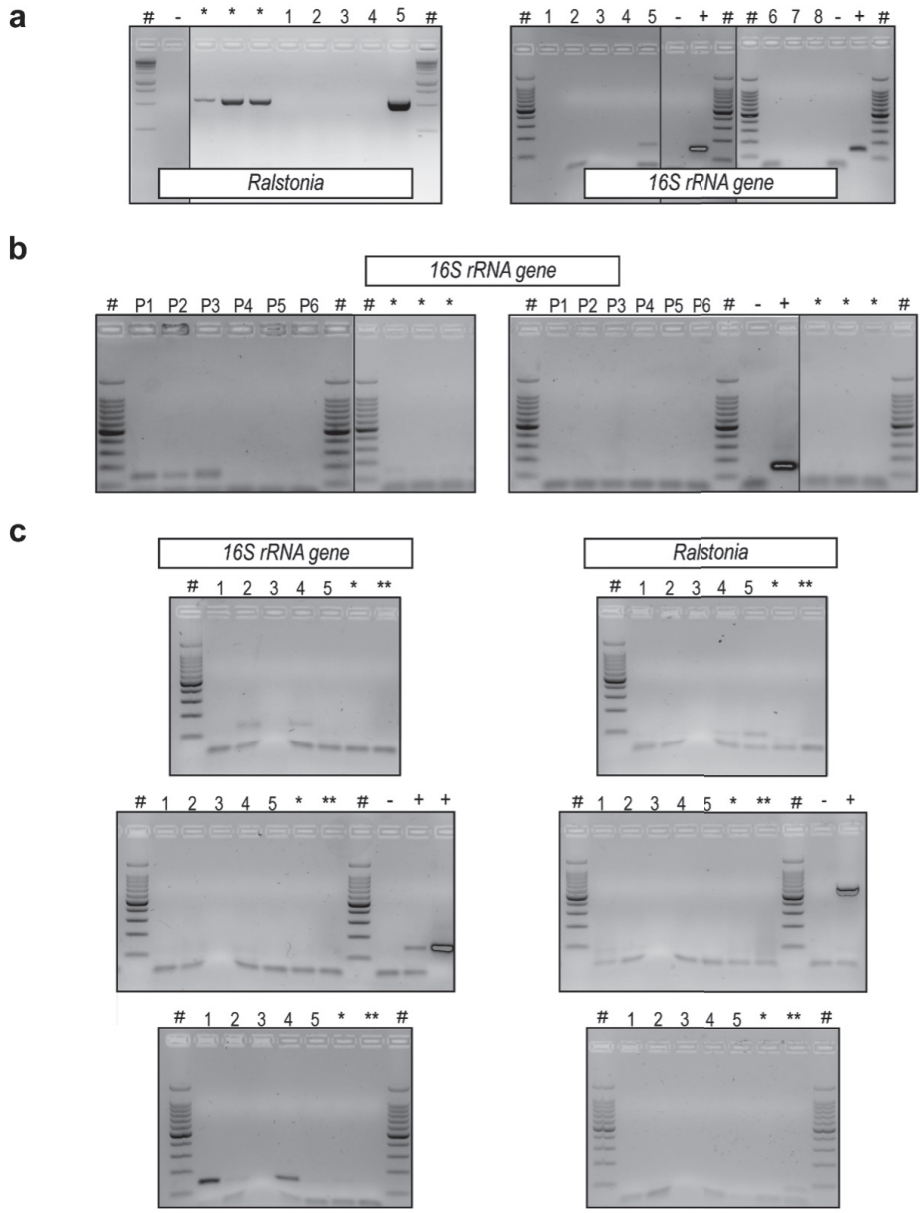
▲Supplementary Figure S4 | Decontamination process of the 16S rRNA amplicon data sets conducted with the non-nested and nested PCR approach. (a) The number of operation taxonomic units (OTUs) within tissues are shown before and after decontamination. The DNA extraction negatives in the non-nested data set and PCR negatives in the nested data set were used for decontamination based on the prevalence method. **(b)** The plot shows the phyla to which the different OTUs belong.



▲Supplementary Figure S5 | Processing formalin-fixed paraffin embedded (FFPE) tissue samples for microbial analyses. An overview of recommendations for handling FFPE tissue samples of low microbial biomass for downstream molecular analyses.



▲**Additional Figure A1 | Full length gel electrophoresis results. (a)** Gel results following *actin beta (ACTB)* and *16S rRNA* gene amplification, emphasizing the use of non-normalized (undiluted) template to improve bacterial DNA detection in FFPE tissue samples (n=3). **(b)** Results showing bacterial and human genomic DNA presence in 41 paired FFPE tissue (T) samples and their paired empty paraffin (P) controls. Tissues consist of colorectal cancer (CRC; n=14), normal adjacent (ADJ; n=14) and healthy control (C; n=13) tissues. The hash tack (#) represents the 100bp ladder. These uncropped findings are complementary to Figure 1c-e.


a
- * * * 1 2 3 4 5
Ralstonia
1 2 3 4 5 - + # # 6 7 8 - +
16S rRNA gene
b
16S rRNA gene
P1 P2 P3 P4 P5 P6 # # * * *
P1 P2 P3 P4 P5 P6 # - + * * *
c
16S rRNA gene
Ralstonia
1 2 3 4 5 * **
1 2 3 4 5 * ** # - + +
1 2 3 4 5 * ** # - +
1 2 3 4 5 * **

◂**Additional Figure A2 | Full length gel electrophoresis results. (a-c)** Gel results following amplification with *Ralstonia* and/or *16S rRNA* gene primers to investigate bacterial presence in resuspension (1), binding (2), wash I-II (3-4) and elution (5) buffers of the RTP Bacteria DNA Mini Kit, extraction additives xylene (6), ethanol (7) and isopropanol (8), in addition to a set of paraffin collections (grains [P1-2], tissue processor machine [P3], embedding stations [P4-6]) processed by different kits. The results in sections b and c represent the findings obtained with two (b; left and right panel) and three (c; upper, middle and lower panel) different DNA isolation kits, respectively. The water (*) and blank (**) controls extracted with the DNA isolation kit were included, as well as positive (+) and negative (-) controls. The hash tack (#) represents a 1kb ladder (a; left panel) or 100bp ladder (a [right panel], b, c). These uncropped findings are complementary to Supplementary Figure S1a-c.

5

CHAPTER 5

First steps towards combining fecal immunochemical testing with the gut microbiome in colorectal cancer screening

E.J. Grobbee[1], S.Y. Lam[1], G.M. Fuhler[1], B. Blakaj[1], S.R. Konstantinov[1], M.J. Bruno[1], M.P. Peppelenbosch[1], E.J. Kuipers[1], M.C.W. Spaander[1]

[1] *Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands.*

United European Gastroenterol J. 2020 Apr;8(3):293-302.

Abstract

Objectives: Many countries use fecal immunochemical testing (FIT) to screen for colorectal cancer. There is increasing evidence that fecal microbiota play a crucial role in colorectal cancer carcinogenesis. We assessed the possibility of measuring fecal microbial features in FIT as potential future biomarkers in colorectal cancer screening.

Methods: Bacterial stability over time and the possibility of bacterial contamination were evaluated using quantitative polymerase chain reaction analysis. Positive FIT samples (*n*=200) of an average-risk screening cohort were subsequently analyzed for universal 16S and bacteria *Escherichia coli* (*E. coli*), *Fusobacterium nucleatum* (*F. nucleatum*), Bacteroidetes and *Faecalibacterium prausnitzii* (*F. prausnitzii*) by quantitative polymerase chain reaction (qPCR). The results were compared with colonoscopy findings.

Results: Fecal microbiota in FIT were stably measured up to six days for *E. coli* (*P*=0.53), *F. nucleatum* (*P*=0.30), Bacteroidetes (*P*=0.05) and *F. prausnitzii* (*P*=0.62). Overall presence of bacterial contamination in FIT controls was low. Total bacterial load (i.e. 16S) was significantly higher in patients with colorectal cancer and high-grade dysplasia (*P*=0.006). For the individual bacteria tested, no association was found with colonic lesions.

Conclusion: These results show that the fecal microbial content can be measured in FIT samples and remains stable for six days. Total bacterial load was higher in colorectal cancer and high-grade dysplasia. These results pave the way for further research to determine the potential role of microbiota assessment in FIT screening.

Keywords:

Microbiome; colorectal cancer; screening; fecal occult blood test.

1 Introduction

Colorectal cancer (CRC) is a major cause of cancer-related morbidity and mortality [1]. The etiology of CRC is complex and not yet completely understood. There is increasing attention for the gut microbiota and its role in colorectal carcinogenesis [2-4]. It is estimated that at least 20%, and perhaps more of the cancer burden worldwide can be attributed to microbial agents [5].

An association between CRC and specific fecal bacteria was already reported a long time ago [6]. In a small Dutch study of 12 patients with *Streptococcus bovis* bacteremia, CRC was diagnosed in eight patients and gastric cancer in one [7]. CRC appears to have a complex etiology with potential etiological contribution of multiple bacterial species playing different roles [3, 4, 8]. Most gut bacteria cannot easily be cultivated, yet sequencing of bacterial deoxyribonucleic acid (DNA) following polymerase chain reaction (PCR) allows for the identification of the composition of the fecal microbiota. Evidence to date suggests that inflammatory processes triggered by enterotoxigenic bacteria can contribute to CRC development by facilitating DNA damage in intestinal epithelial cells [4, 9]. The ensuing accumulation of genetic lesions can contribute to oncogenesis along the adenoma-carcinoma sequence. Several studies have shown that the bacterial composition of malignant lesions differs from that of surrounding normal tissue [4, 10, 11]. While most previous research has focused on the unravelling of the complex microbial composition in CRC and the role of the gut microbiota in the pathogenesis of CRC, it is of interest to see whether altered bacterial presence may be valuable in improving screening strategies.

In the past decennia, an increasing number of countries have embarked on CRC screening. Many of those use fecal immunochemical tests (FIT) as their screening method [12]. FITs rely on the measurement of trace amounts of blood from neoplastic lesions. However, not all lesions bleed (in e.g., serrated adenomas) and conversely, occult blood can be detected in fecal samples of healthy individuals [13]. In spite of high participation rates and a relatively high sensitivity for CRC of 75-85% depending on the cut-off used, the sensitivity of FIT for detection of advanced adenomas is much lower and generally ranges below 50% [14, 15]. For this reason, there is an urgent need for additional markers to increase FIT sensitivity without losing its specificity, as the latter is of crucial importance in a screening setting. Investigation of fecal bacterial features could present one such possible additional marker [16]. Hence, it would be of great interest to detect bacterial features in the rest materials of FIT screenees, which would preclude additional material collection from screenees.

The aim of our study was to evaluate the possibility of measuring fecal microbiota in FIT

in relation to endoscopic findings, to evaluate their stability over time and to assess the effect of potential bacterial contaminants in downstream PCR analysis. For this proof of principal, we have selected four different bacterial markers for quantitative PCR (qPCR) analysis: suspected driver bacteria of the Enterobacteriaceae (*Escherichia coli* [*E. coli*]), Bacteroidetes species, the most often associated CRC bacterium *Fusobacterium nucleatum* (*F. nucleatum*) and the anti-inflammatory *Faecalibacterium prausnitzii* (*F. prausnitzii*) which was described to be less prevalent in CRC patients [17-24]. In addition, taxonomic profiling was carried out for six pooled samples from patients with different endoscopy outcomes to evaluate the feasibility of using FIT fluid for future 16S rRNA gene sequencing purposes.

2 Materials and Methods

2.1 Patients, FIT screening and data collection

Details about the design of this ongoing population-based FIT CRC screening program have been described previously [25]. In short, demographic data of all individuals between 50 and 74 years living in the southwest of The Netherlands were randomly obtained from municipal population registers and were invited for FIT screening biennially. At present, four rounds of FIT screening have taken place. For this study, only FIT samples of the end of the third and beginning of the fourth screening round were used, aiming for a total of 200 FITs to be included. Recruitment of this third and fourth screening round took place between February 2013 and August 2014. In the third screening round, all invitees received the OC-sensor (Eiken, Japan). In the fourth screening round, invitees were randomized between the OC-sensor and FOB-Gold (Sentinel, Italy). Participants were instructed to send the FIT within one day after collection and to keep the FIT in the refrigerator until sending it to the laboratory. A cut-off of ≥10 µg hemoglobin per gram (Hb/g) feces was used to refer screenees for colonoscopy within four weeks. All colonoscopies were performed by gastro-enterologists with an experience based on at least 1000 colonoscopies. All lesions were evaluated by trained gastrointestinal pathologists according to the Vienna criteria [26]. Advanced adenomas (AA) were defined as an adenoma with a diameter ≥10 mm, and/or with a ≥25% villous component, and/or high-grade dysplasia (HGD). Advanced neoplasia included AA and CRC, with the most advanced lesion used for analysis. Serrated polyps (SP) were defined as serrated adenomas (with or without dysplasia) and hyperplastic polyps. For this study only FIT-positive screenees were included.

The study was approved by the Medical Ethical Committee of the Erasmus MC University Medical Center Rotterdam (reference number: MEC-2014-212) on 6 May 2014. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki as reflected

in a prior approval by the institution's human research committee. All participants gave written informed consent to participate in CRC screening.

2.2 Bacterial quantitative analysis

After occult blood measurement, FIT samples were stored at -20˚C until analysis. DNA was isolated from FIT liquid by Wizard Genomic DNA Purification kit (Promega, Leiden, The Netherlands) with modifications. Information on primers, PCR and qPCR analyses can be found in the **Supplementary methods**.

2.3 Microbial stability and contamination in FIT

For analysis of the stability of the microbial content of FIT over time, seven FITs from stool samples with (n=2) and without blood (n=5) of healthy volunteers were collected and stored at -20˚C immediately or after 24, 48, 72, 96, 120 and 144 hours (h) in order to mimic FIT transit time. DNA was isolated from all samples upon thawing, and the presence of *E. coli*, *F. nucleatum*, Bacteroidetes, *F. prausnitzii* and universal bacterial 16S were detected by PCR and qPCR as described above. *F. nucleatum* was below detection level in all FIT samples of our healthy donors and could not be included in the analysis. To test for unintentional bacterial contamination, FIT controls with and without blood but not containing feces underwent the same procedure for comparison to FIT with blood and fecal material.

2.4 16S rRNA gene sequencing pilot

To determine the feasibility of using FIT material for 16S rRNA gene sequencing, a pilot was conducted using six different pooled DNA samples from patients with the following conditions: 1) no endoscopic findings, 2) tubular adenoma, 3) HGD, 4) CRC, 5) sessile adenoma and 6) hyperplastic polyp. All samples were shipped on dry ice to the Macrogen Institute in Seoul, Korea. The V3-4 region of the *16S rRNA* gene was amplified and sequenced on the Illumina MiSeq platform. The sequencing data was subsequently processed using the SILVA database for taxonomic profiling at genus level, allowing for the global assessment of the bacterial presence of our selected markers in FIT material.

2.5 Statistical analysis

Descriptive data were reported as proportions or means with the standard deviation. For non-normally distributed data, the median and interquartile range (IQR) were given. Chi-Square tests were used to analyse categorical data; continuous data were analyzed using Student's t-tests or one-way analysis of variance. Linear regression analysis was

used to assess bacterial load and transit time. Correction for multiple testing was done according to Bonferroni, resulting in a two-sided *P*-value of <0.01 that was considered to be statistically significant. Statistical analysis was performed using IBM SPSS version 21.0.

3 Results

3.1 FIT screenees

A total of 200 samples from FIT positive screenees were collected. Of these, 20 samples had to be discarded for various reasons (e.g. inclusion of multiple samples from the same screenee, sample misclassification or missing pathology outcome), resulting in 180 samples available for analysis of microbial content. These included 119 OC-sensor tests (66%) and 61 FOB-Gold tests (34%). Of those, 56% were male with a median age of 64 years (IQR 58-69 years). Median fecal Hb concentration was 21µg Hb/g feces (IQR 13-55µg Hb/g feces). All screenees included in this study underwent complete colonoscopy and in 31% patients (*n*=55) advanced neoplasia was detected, of whom 5 were diagnosed with CRC. All colonoscopy findings are described in **Table 1**.

3.2 Stability of microbiota in FIT over time and bacterial contaminants

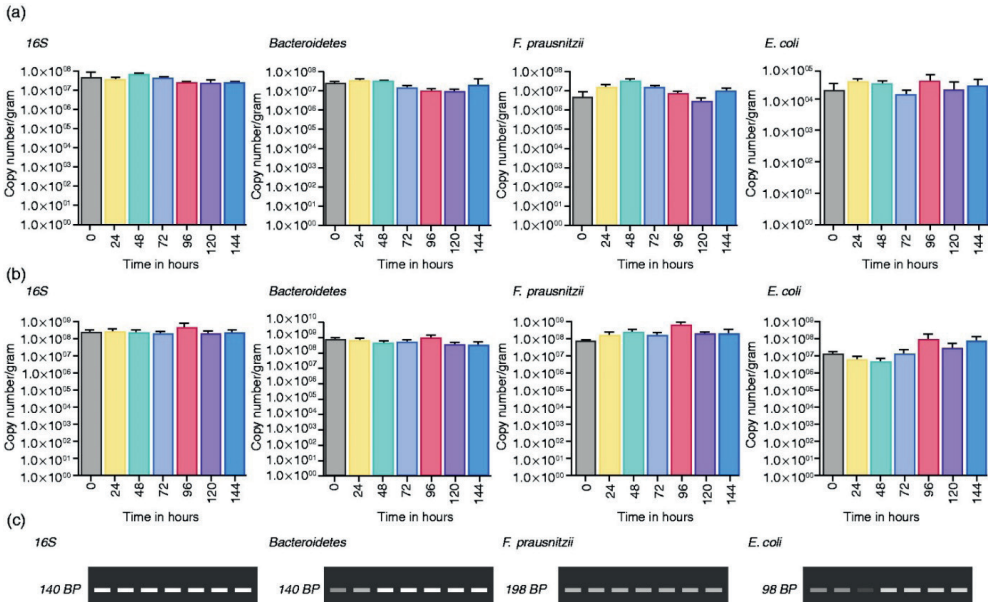
Transit time of the FIT from screenee to the laboratory could potentially affect the microbial composition detected. Although growth of anaerobic bacteria is not expected and FIT buffer contains bacteriostatic sodiumazide, degradation of bacterial DNA might occur. Therefore, we first analyzed the stability of the bacterial composition in FIT. Universal bacterial 16S, Bacteroidetes, *F. prausnitzii* and *E. coli* DNA was consistently detected by (q)PCR, with no significant loss in detection levels for up to 144h in FIT positive (**Figure 1a**) and FIT negative (**Figure 1b,c**) samples. Microbial contamination did not influence our results as indicated by relatively low levels of bacterial contamination detected in FIT controls (**Figure 2).** The 16S and individual bacterial marker levels in FIT fluid without fecal content were similar to water controls and several times lower than FIT fluid containing fecal material.

The average time between fecal sampling by the screenee and analysis at the laboratory (i.e. transit time) was 1 day (IQR 1-2 days), with 91% of FITs arriving at the laboratory within two days after sampling. For all screenees, the correlation between absolute copy number of the four bacteria and transit time was evaluated (**Figure 3**). No significant decrease in fecal microbiota was seen up to six days for *E. coli* (*P*=0.53), *F. nucleatum* (*P*=0.30), Bacteroidetes (*P*=0.05) and *F. prausnitzii* (*P*=0.62).

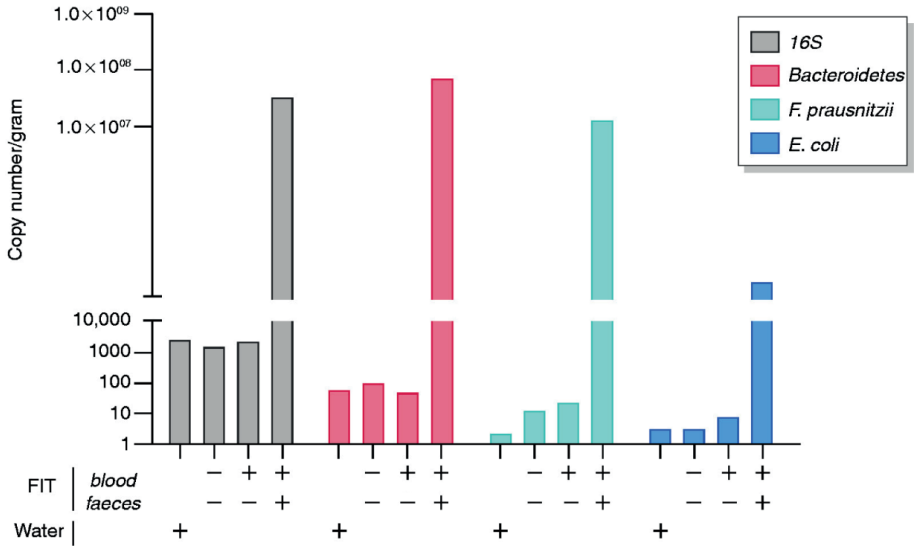
Table 1 | Most advanced lesion at colonoscopy of fecal immunochemical test (FIT) positive screenees

Finding at colonoscopy	***n***	**%**
Normal	41	23
Serrated polyps	25	14
Tubular adenoma <10 mm	59	32
Tubular adenoma ≥10 mm	33	18
(Tubulo)villous adenoma	14	8
High-grade dysplasia	3	2
Colorectal carcinoma	5	3
Total	180	100

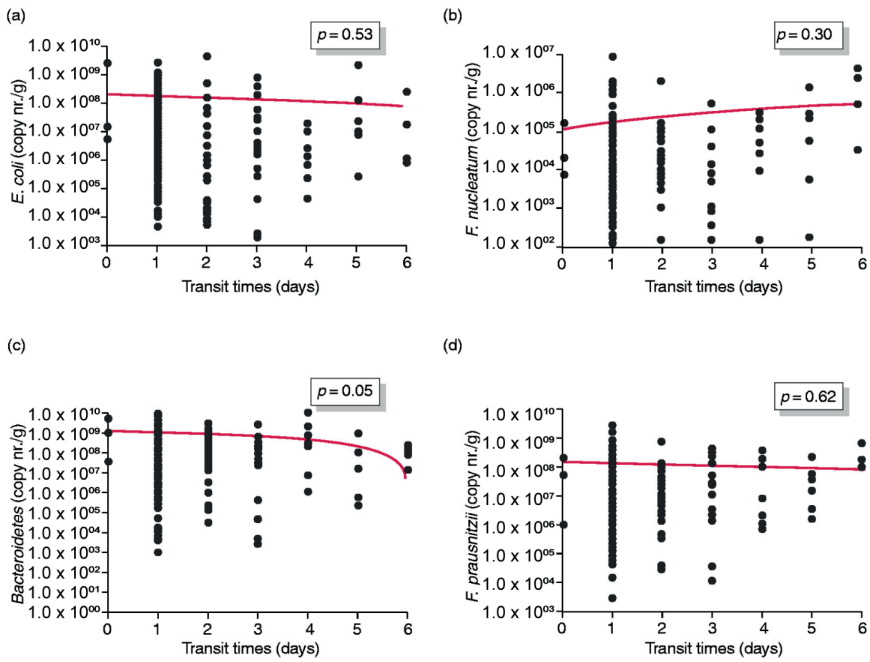
Abbreviations: nr, number.



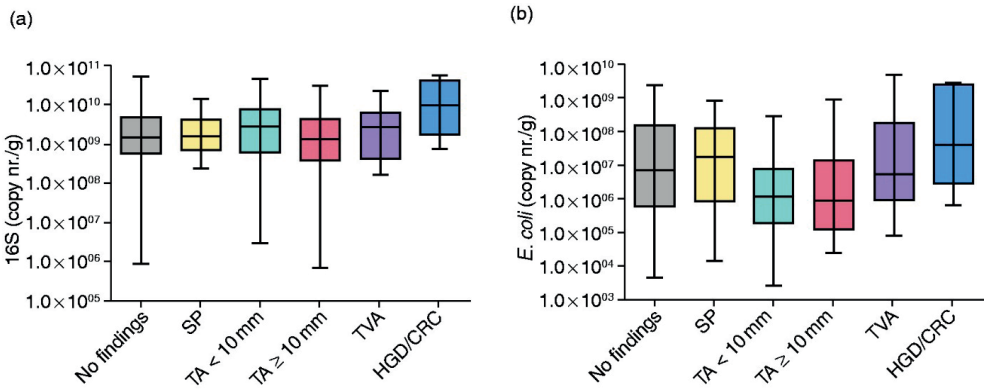
▲**Figure 1 | (a-c)** Stability of bacterial composition over time for fecal immunochemical test (FIT) positive (*n*=2) (a) and FIT negative specimens (*n*=5) (b,c). This has been depicted for universal 16S and specific markers Bacteroidetes, *Faecalibacterium prausnitzii* and *Escherichia coli*. Bacterial composition remained stable up to at least 144 hours.



▲Figure 2 | Quantitative polymerase chain reaction (qPCR) assessment of unintentional bacterial contamination. Fecal immunochemical test controls in either the presence or the absence of occult blood and/or faecal material were tested for 16S, Bacteroidetes, *Faecalibacterium prausnitzii* and *Escherichia coli*. Water controls were concurrently processed for qPCR analysis.



▲Figure 3 | (a-d) Transit time (interval between fecal sampling and arrival of the fecal immunochemical test (FIT) at the laboratory) and absolute copy number/gram (nr./g) FIT for *Escherichia coli* (a), *Fusobacterium nucleatum* (b), Bacteroidetes (c), *Faecalibacterium prausnitzii* (d).



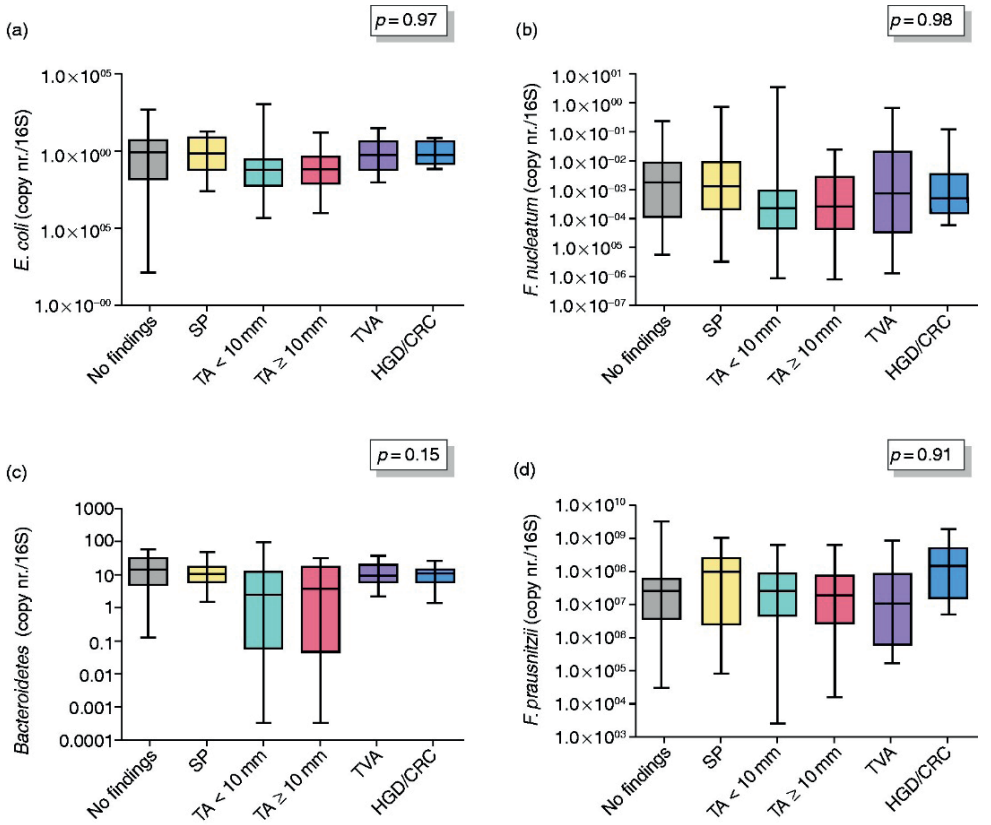
▲**Figure 4 | (a-b)** Copy number per gram (nr./g) fecal immunochemical test (FIT) liquid for 16S (a) and *Escherichia coli* (b) according to colonoscopy outcomes. Abbreviations: SP, serrated polyp; TA, tubular adenoma; TVA, (tubulo)villous adenoma; HGD, high grade dysplasia; CRC, colorectal cancer.

5

3.3 Microbiota in FIT and findings at colonoscopy

For all samples, copy number per gram (copy nr./g) FIT liquid was calculated for the total number of bacteria (i.e. 16S) and the four predefined bacteria. A significant difference was seen for 16S, with increasing abundance of total bacterial content in screenees with high-grade dysplasia and CRC (P=0.006) (**Figure 4a**). Significant differences between the groups were seen for *E. coli* (P=0.05) (**Figure 4b**). Post hoc testing revealed that in particular patients with tubular and villous adenoma showed lower levels of *E. coli*, albeit not significant (P=0.07). For *F. nucleatum*, *F. prausnitzii*, and Bacteroidetes, no association was observed between the presence of the bacteria and any particular lesion (**Supplementary Figure S1**). No significant association between amounts of bacteria and presence of advanced neoplasia was observed (**Supplementary Figure S2**).

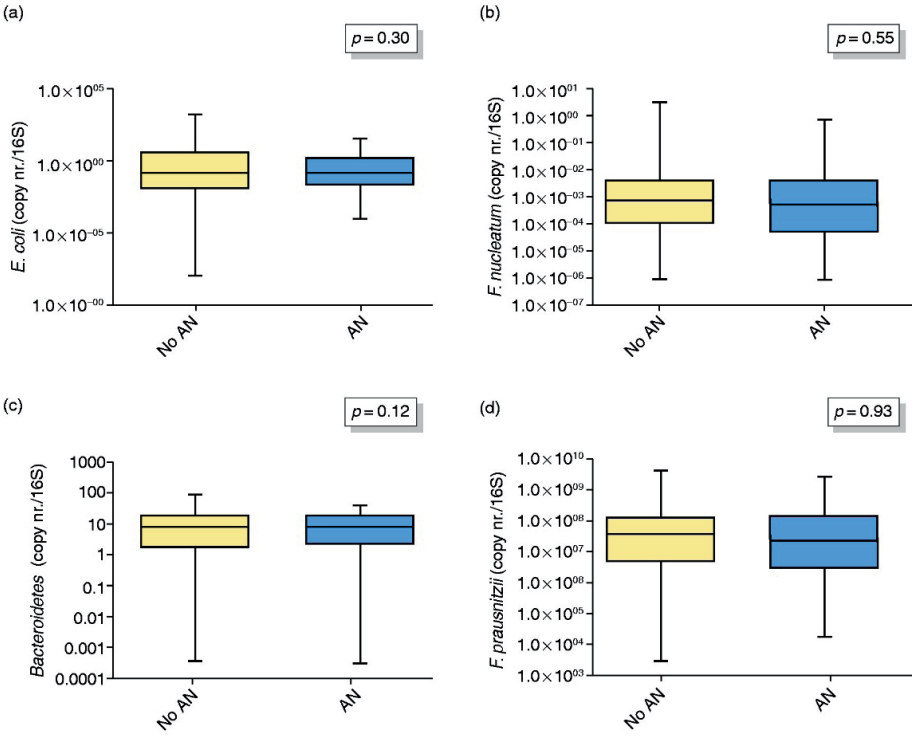
To correct for potential differences in amount of fecal matter in the FIT, the bacteria were also calculated relative to the total bacterial presence as determined by universal 16S (copy nr./g of 16S). No significant differences were found when evaluating FIT microbiota according to all colonoscopy findings for *E. coli* (P=0.97), *F. nucleatum* (P=0.98), Bacteroidetes (P=0.15) and *F. prausnitzii* (P=0.91) (**Figure 5**). In addition, no significant differences in microbiota were found between screenees with and without advanced neoplasia (*E. coli*, P=0.30; *F. nucleatum,* P=0.55; Bacteroidetes P=0.12; *F. prausnitzii*, P=0.93) (**Figure 6**). When evaluating FIT microbiota according to location of the most advanced lesion (i.e. distal vs. proximal), again no significant differences were seen for all four bacteria (**Supplementary Figure S3**).



▲**Figure 5 | (a-d)** Absolute copy number (nr.) per 16S and most advanced colonoscopy finding for *Escherichia coli* (a), *Fusobacterium nucleatum* (b), Bacteroidetes (c) and *Faecalibacterium prausnitzii* (d). Abbreviations: SP, serrated polyp, TA: tubular adenoma; TVA, (tubulo)villous adenoma; HGD: high grade dysplasia; CRC, colorectal cancer.

3.4 16S rRNA gene sequencing of pooled samples

Although it is well known that the overall bacterial abundance in FIT material is relatively low compared to stool specimens, 16S rRNA gene sequencing data were generated for all six pooled DNA samples from patients with respectively 1) no endoscopic findings, 2) tubular adenoma, 3) HGD, 4) CRC, 5) sessile adenoma and 6) hyperplastic polyp. Taxonomic classification indicated that Bacteroidetes in addition to genus *Faecalibacterium* (with regard to *F. prausnitzii*) and family Enterobacteriaceae (with regard to *E. coli*) were present in all samples (**Supplementary Figure S4**). Genus *Fusobacterium* (with regard to *F. nucleatum*) was detected in the pooled DNA samples from tubular adenomas, suggesting that this lower abundant bacterial marker is present in FIT material. More importantly, these findings confirm the ability to use FIT specimens for qPCR analysis as well as future 16S rRNA gene sequencing purposes.



▲**Figure 6 |** Absolute copy number (nr.) per 16S for screenees with no advanced neoplasia (No AN) and advanced neoplasia (AN) for *Escherichia coli* (a), *Fusobacterium nucleatum* (b), Bacteroidetes (c), *Faecalibacterium prausnitzii* (d).

4 Discussion

Our results show that fecal microbial DNA can be isolated from FIT samples and remains stable for up to six days. The inclusion of FIT controls in qPCR analysis allowed the assessment of bacterial contamination which appeared to be of minimal impact as the detected levels were similar to water controls. When the qPCR findings were put against the endoscopic findings, screenees with HGD and CRC had a higher load of total universal 16S. With respect to specific microbial features, no relation was found between numbers of specific bacteria and colonoscopy findings relative to total 16S, except that numbers of *E. coli* were reduced in patients with tubular and villous adenoma. With regard to location of the lesion, no differences were found between a lesion in the distal or proximal colon and number of fecal bacteria observed.

E. coli and Bacteroidetes species have been suggested to promote inflammation driving the colorectal epithelium to a carcinogenic state [18, 20, 21]. This state of inflammation and dysbiosis gives room for opportunistic bacteria, such as *F. nucleatum*, to further induce

carcinogenesis, whereas anti-inflammatory bacteria such as *F. prausnitzii* may be 'crowded out' [4, 21]. Consequently, various bacteria take part in the process of carcinogenesis, with many of these bacteria being variably present during carcinogenesis [4]. This could explain why lower concentrations of *E. coli* were found in tubular and villous adenomas, although screenees with normal colonoscopy and those with CRC had similar concentrations. Our findings did not support a role for *E. coli* and Bacteroidetes as additional biomarker in FIT samples to identify FIT positive screenees at risk of carrying advanced adenomas. Previous studies have suggested a role for *F. nucleatum* in CRC, in particular the detection of sessile serrated lesions as the mucus cap on these lesions might cause higher levels of *F. nucleatum* [19]. Additional detection of sessile serrated lesions would be especially valuable in FIT screening, as FIT is known to have a poor sensitivity for these lesions [27]. However, our results did not show any association between *F. nucleatum* and hyperplastic polyps or serrated lesions compared to other neoplasia or a normal colon (P=0.82; data not shown). This could be because *F. nucleatum* is not sensitive enough by itself as a biomarker in a screening setting due to overabundance in healthy subjects [28].

We found that fecal microbial DNA remained stable over six days, which is in line with findings from a previous study comparing different collection methods of feces including FIT [29]. Furthermore, one other study has looked specifically at isolating bacterial DNA from FIT samples [30]. Its findings are in line with ours showing that fecal material contained in FIT sampling is sufficient to perform microbiota characterization and is representative to bacterial findings in a full stool sample.

Most other studies regarding the role of the microbiome in colorectal carcinogenesis have looked specifically at the microbiome at and around the tumor site and it is conceivable that a fecal sample obtained by FIT is not representative of onsite mucosal dysbiosis [31, 32]. However, microbiota on mucosa retrieved during colonoscopy or surgery could be influenced by the bowel preparation that all patients undergo prior to the intervention. Furthermore, most of these studies had a case-control design and were thus prone to overestimate diagnostic performance [33]. To date, there are a small number of studies that looked at the fecal microbiome in FIT screenees, showing a difference in overall fecal microbiome between healthy patients and patients with colorectal adenomas [34, 35]. Two of these studies analyzed the microbiota in full stool samples and not in the FIT samples themselves, making comparison with our data complex. However, a full stool sample may ask for a considerable effort from the screenee, making the design undesirable in a screening setting as it might hamper participation rates. Baxter *et al.* used 16S rRNA sequencing of stool samples to identify a microbiota-based model to predict colonic

lesions and subsequently showed that this model also worked on DNA isolated from FITs indicating that FIT fluid may indeed provide additional biomarkers for CRC detection [30, 35]. While 16S rRNA gene sequencing data may give a more in depth analysis of the microbiome as seen in patients with different degrees of intestinal malignancy, its use for diagnostic purposes of individual patients may be cumbersome. Furthermore, in our pilot study, it was not possible to retrieve detailed taxonomic information on species level as with a limited sequencing depth, 16S rRNA gene sequencing did not pick up specific markers at species level such as *F. nucleatum*. For implementation in diagnostic laboratories, it might be preferable to find microbial biomarkers which may be identified in FIT by readily available PCR techniques rather than 16S or metagenomic sequencing efforts. However, with current literature not agreeing on the actual absence or presence of bacterial featured in colonic lesions, identifying the right biomarker to investigate is key.

The strengths of our study include the fact that all FIT samples were retrieved from a population-based CRC screening cohort, consisting of average risk screenees, resulting in a high external validity. Also, as gut microbiota were measured in FIT samples, no additional stool samples were required from the participants. It is the first study comparing microbial features between previously untreated patients across the adenoma-to-carcinoma range, including all different stages of malignancy. Furthermore, we included FIT samples that tested positive for occult blood for both lesions and non-lesions, precluding the possibility of a bias introduced by potential microbe-blood interactions [36]. In order to appreciate our results, some limitations also need to be addressed. At present, the exact pathway and role of the gut microbiota is unknown. Since no known common suspects have been consistently identified, we have selected four bacteria for this qPCR study, but the inclusion of other bacteria could be of more interest in the future. As only FIT-positive subjects underwent colonoscopy, it was not possible to evaluate prime indicators of diagnostic performance, including sensitivity, specificity and the area under the receiver-operating curve. However, we considered analysis of only positive FITs justified, as in the end, this is the population for whom biomarkers would be of benefit to avoid unnecessary endoscopic screening. Furthermore, we used the most advanced lesion detected during colonoscopy, while screenees sometimes have more than one lesion. The presence of multiple lesions could theoretically lead to our findings being an underestimation of the relation between fecal microbiome and colonic neoplasia, although for screening purposes subjects at highest risk (i.e. with advanced neoplasia) are of most interest. Another important limitation is the small sample size and the absence of negative controls in the 16S rRNA gene sequencing pilot. PCR analysis indicates that, while orders of multitude lower, contaminants will be present during isolation and amplification of DNA from FIT fluids. Although the markers

of choice were detectable in the pooled samples, which shows their feasibility to use for qPCR purposes, future studies should incorporate negative controls to confirm the detection of biological relevant signals and to control for bacterial contamination.

In conclusion, our results illustrate that the gut microbial markers can be stably measured in FIT samples in CRC screening, with a higher total bacterial load for CRC and high-grade dysplasia. The need to increase FIT sensitivity, especially for advanced adenomas, remains of evident importance and further studies should be conducted to determine the role of microbiota and preferably specific biomarkers in FIT.

5 Funding

Erasmus MC MRACE grant no 105565 (a population-based study on *F. nucleatum* and CRC). The funding source had no involvement in the study design, collection of data, analysis and interpretation of the data, in the writing of the report, nor in the decision to submit the paper.

References

1. Kuipers EJ, Grady WM, Lieberman D, Seufferlein T, Sung JJ, Boelens PG *et al*. Colorectal cancer. Nat Rev Dis Primers, 2015. 1: p. 15065.
2. Narayanan V, Peppelenbosch MP and Konstantinov SR. Human fecal microbiome-based biomarkers for colorectal cancer. Cancer Prev Res (Phila), 2014. 7(11): p. 1108-11.
3. Gagnière J, Raisch J, Veziant J, Barnich N, Bonnet R, Buc E *et al*. Gut microbiota imbalance and colorectal cancer. World J Gastroenterol, 2016. 22(2): p. 501-18.
4. Tjalsma H, Boleij A, Marchesi JR and Dutilh BE. A bacterial driver-passenger model for colorectal cancer: beyond the usual suspects. Nat Rev Microbiol, 2012. 10(8): p. 575-82.
5. zur Hausen H. The search for infectious causes of human cancers: where and why (Nobel lecture). Angew Chem Int Ed Engl, 2009. 48(32): p. 5798-808.
6. Klein RS, Recco RA, Catalano MT, Edberg SC, Casey JI, and Steigbigel NH. Association of Streptococcus bovis with carcinoma of the colon. N Engl J Med, 1977. 297(15): p. 800-2.
7. Kuipers EJ and de Jong A. [Gastrointestinal disorders and Streptococcus bovis bacteremia]. Gastro-intestinale aandoeningen en Streptococcus bovis-bacteriëmie. Ned Tijdschr Geneeskd, 1990. 134(28): p. 1337-9.
8. Sears CL and Garrett WS. Microbes, microbiota, and colon cancer. Cell Host Microbe, 2014. 15(3): p. 317-28.
9. Bhatt AP, Redinbo MR and Bultman SJ. The role of the microbiome in cancer development and therapy. CA Cancer J Clin, 2017. 67(4): p. 326-344.
10. Flanagan L, Schmid J, Ebert M, Soucek P, Kunicka T, Liska V *et al*. Fusobacterium nucleatum associates with stages of colorectal neoplasia development, colorectal cancer and disease outcome. Eur J Clin Microbiol Infect Dis, 2014. 33(8): p. 1381-90.
11. Marchesi JR, Dutilh BE, Hall N, Peters WH, Roelofs R, Boleij A *et al*. Towards the human colorectal cancer microbiome. PLoS One, 2011. 6(5): p. e20447.
12. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJ, Young GP *et al*. Colorectal cancer screening: a global overview of existing programmes. Gut, 2015. 64(10): p. 1637-49.
13. van Doorn SC, Stegeman I, Stroobants AK, Mundt MW, de Wijkerslooth TR, Fockens P *et al*. Fecal immunochemical testing results and characteristics of colonic lesions. Endoscopy, 2015. 47(11): p. 1011-7.
14. Lee JK, Liles EG, Bent S, Levin TR and Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med, 2014. 160(3): p. 171.
15. Imperiale TF, Gruber RN, Stump TE, Emmett TW and Monahan PO. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps: A Systematic Review and Meta-analysis. Ann Intern Med, 2019. 170(5): p. 319-329.
16. 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. 68(9): p. 1723-1725.
17. Lopez-Siles M, Martinez-Medina M, Surís-Valls R, Aldeguer X, Sabat-Mir M, Duncan SH *et al*. Changes in the Abundance of Faecalibacterium prausnitzii Phylogroups I and II in the Intestinal Mucosa of Inflammatory Bowel Disease and Patients with Colorectal Cancer. Inflamm Bowel Dis, 2016. 22(1): p. 28-41.
18. Liang Q, Chiu J, Chen Y, Huang Y, Higashimori A, Fang J *et al*. Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer. Clin Cancer Res, 2017. 23(8): p. 2061-2070.

19. Ito M, Kanno S, Nosho K, Sukawa Y, Mitsuhashi K, Kurihara H *et al*. Association of Fusobacterium nucleatum with clinical and molecular features in colorectal serrated pathway. Int J Cancer, 2015. 137(6): p. 1258-68.

20. Leung A, Tsoi H and Yu J. Fusobacterium and Escherichia: models of colorectal cancer driven by microbiota and the utility of microbiota in colorectal cancer screening. Expert Rev Gastroenterol Hepatol, 2015. 9(5): p. 651-7.

21. Sears CL and Pardoll DM. Perspective: alpha-bugs, their microbial partners, and the link to colon cancer. J Infect Dis, 2011. 203(3): p. 306-11.

22. 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(1): p. 70-78.

23. 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(4): p. 679-689.

24. Balamurugan R, Rajendiran E, George S, Samuel GV and Ramakrishna BS. Real-time polymerase chain reaction quantification of specific butyrate-producing bacteria, Desulfovibrio and Enterococcus faecalis in the feces of patients with colorectal cancer. J Gastroenterol Hepatol, 2008. 23(8 Pt 1): p. 1298-303.

25. Grobbee EJ, van der Vlugt M, van Vuuren AJ, Stroobants AK, Mundt MW, Spijker WJ *et al*. A randomised comparison of two faecal immunochemical tests in population-based colorectal cancer screening. Gut, 2017. 66(11): p. 1975-1982.

26. Schlemper RJ, Riddell RH, Kato Y, Borchard F, Cooper HS, Dawsey SM *et al*. The Vienna classification of gastrointestinal epithelial neoplasia. Gut, 2000. 47(2): p. 251-5.

27. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP *et al*. Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med, 2014. 370(14): p. 1287-97.

28. Kostic AD, Chun E, Robertson L, Glickman JN, Gallini CA, Michaud M *et al*. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. Cell Host Microbe, 2013. 14(2): p. 207-15.

29. 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(2): p. 115-123.

30. Baxter NT, Koumpouras CC, Rogers MA, Ruffin MTt and Schloss PD. DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. Microbiome, 2016. 4(1): p. 59.

31. Zoetendal EG, von Wright A, Vilpponen-Salmela T, Ben-Amor K, Akkermans AD, and de Vos WM. Mucosa-associated bacteria in the human gastrointestinal tract are uniformly distributed along the colon and differ from the community recovered from feces. Appl Environ Microbiol, 2002. 68(7): p. 3401-7.

32. Chen W, Liu F, Ling Z, Tong X and Xiang C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. PLoS One, 2012. 7(6): p. e39743.

33. Deeks JJ BP, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Available from: http://srdta.cochrane.org/.

34. Goedert JJ, Gong Y, Hua X, Zhong H, He Y, Peng P *et al*. Fecal Microbiota Characteristics of Patients with Colorectal Adenoma Detected by Screening: A Population-based Study. EBioMedicine, 2015. 2(6): p. 597-603.

35. Baxter NT, Ruffin MTt, Rogers MA and Schloss PD. Microbiota-based model improves the sensitivity of fecal immunochemical test for detecting colonic lesions. Genome Med, 2016. 8(1): p. 37.

36. Schroedl W, Kleessen B, Jaekel L, Shehata AA and Krueger M. Influence of the gut microbiota on blood acute-phase proteins. Scand J Immunol, 2014. 79(5): p. 299-304.

Supplementary information

Supplementary methods

Bead-beating was performed 3 times for 30 seconds to lyse bacteria. Protein was precipitated from the supernatant by Protein Precipitation Buffer, followed by isopropanol precipitation of deoxyribonucleic acid (DNA). DNA was washed with 70% ethanol and resuspended in Tris-EDTA (TE) buffer. The concentration was measured by Nanodrop (Thermo Fisher Scientific, Waltham, MA) and adjusted to 10ng/µl. Bacterial DNA (*E. coli, F. nucleatum,* Bacteroidetes (*Bacteroides-Prevotella-Porphyromonas*), *F. prausnitzii* and universal bacterial 16S) was detected by polymerase chain reaction (PCR) or quantitative PCR (qPCR). Specificity of primers (**Supplementary Table S1**) was determined by primer blast search against all bacteria (http://www.ncbi.nlm.nih.gov/tools/primer-blast/).

In addition, DNA was amplified by PCR using GoTaq polymerase (Promega, Leiden, The Netherlands) and 35 cycles (95˚ for 15", 56˚ for 30", 72˚ for 30") on a 2720 Thermal Cycler (Applied Biosystems, Thermo Fisher Scientific) and PCR products were verified by 2% agarose gel electrophoresis. Analysis of bacterial abundance was performed by qPCR. Amplification was performed in a 20µl volume containing 10µl SYBR® Select Master Mix for CFX (Thermo Fisher Scientific), 2µl forward and reverse primer (end concentration 1µM), 7µl H_2O and 1µl template. Cycle conditions were 95˚ for 10 min (initial denaturation), followed by 40 cycles of (95˚ for 15", 56˚ for 30", 72˚ for 30") and melt curve analysis on StepOnePlus real time PCR System (Applied Biosystems). For each bacterium, DNA for standard curves was prepared by PCR using DNA isolated from fecal immunochemical tests (FITs) as template. PCR products were purified using Invisorb® Fragment CleanUp kit (Stratec molecular, Berlin, Germany), DNA concentration was measured and confirmed by sanger sequencing. DNA dilutions ranging from 0.0001-10 ng/µl were prepared. The amount of DNA in test samples was inferred from their Ct value through calculation of standard curves run on each separate plate, and corrected for total amount of DNA present in the FIT liquids. Copy number per gram (copy nr./g) PCR product was inferred from the weight of one PCR product as calculated by the accumulated weight of the base pairs in the product, and results were presented as absolute bacterial content (copy nr./g FIT fluid) or as ratio of the total amount of 16S copies/g FIT fluid. Samples in which bacterial DNA was undetectable were given a value smaller than the lowest detectable value for that PCR product, in order to prevent loss of these samples from group comparisons.

Supplementary Table S1 | Primer sequences

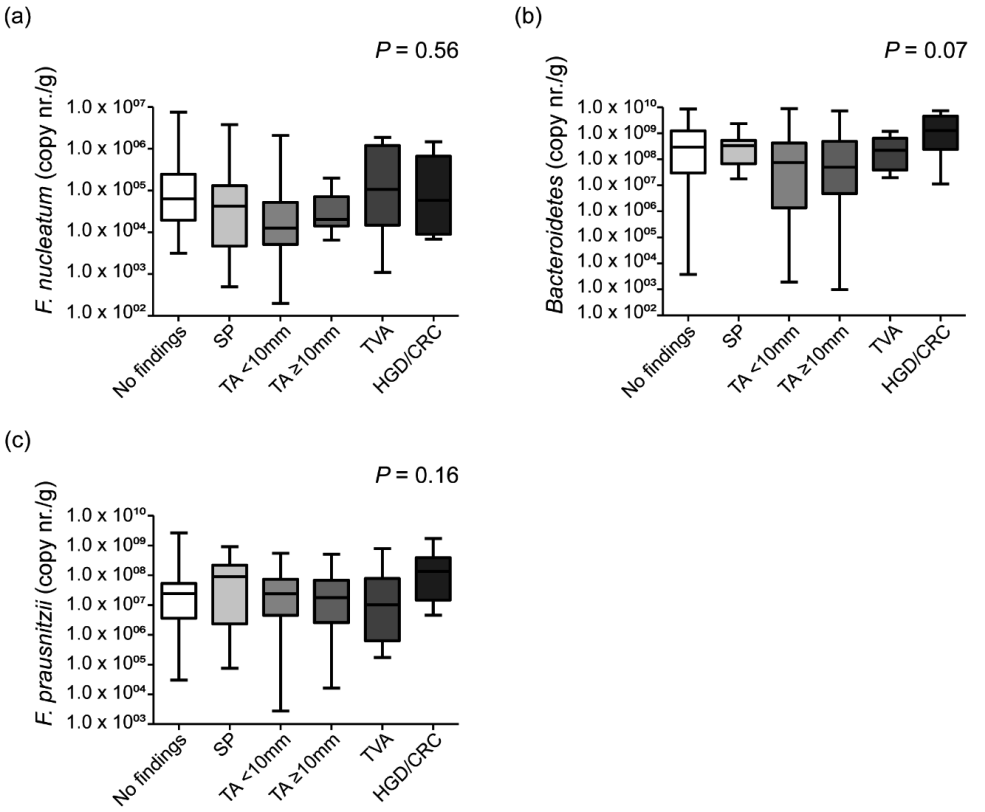
Target	Sequence (5'-3')		Size (bp)	Ref
Universal 16S	Forward	CGGTGAATACGTTCCCGG	145	1-4
	Reverse	TACGGCTACCTTGTTACGACTT		
F. nucleatum	Forward	CTTAGGAATGAGACAGAGATG	140	5-6
	Reverse	TGATGGTAACATACGAAAGG		
E. coli	Forward	CATGCCGCGTGTATGAAGAA	96	7-9
	Reverse	CGGGTAACGTCAATGAGCAAA		
Bacteroidetes	Forward	GGTGTCGGCTTAAGTGCCAT	140	3, 10-11
	Reverse	CGGACGTAAGGGCCGTGC		
F. prausnitzii	Forward	GATGGCCTCGCGTCCGATTAG	198	12-13
	Reverse	CCGAAGACCTTCTTCCTCC		

Bacterial abbreviations: F. nucleatum, Fusobacterium nucleatum; E. coli, Escherichia coli; F. prausnitzii, Faecalibacterium prausnitzii. Other abbreviations: bp, base pair.

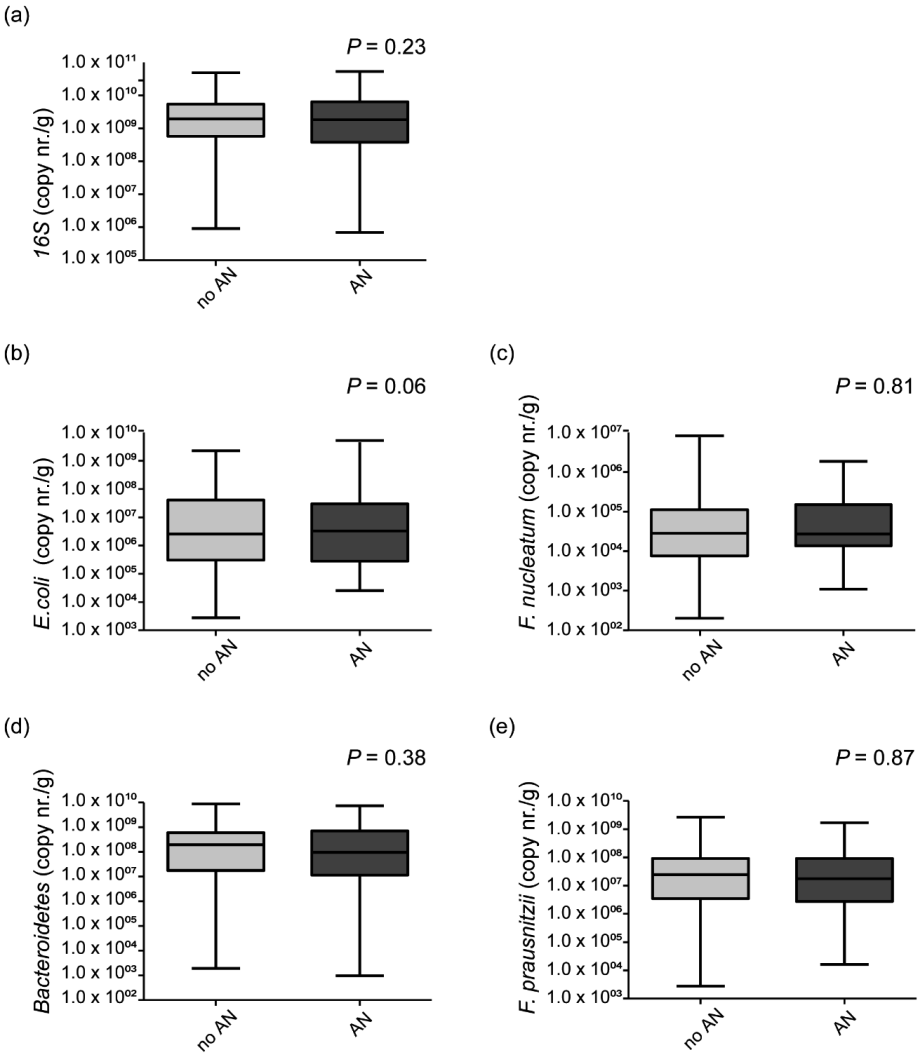
Supplementary references

1. Suzuki MT, Taylor LT and DeLong EF. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl Environ Microbiol, 2000. 66(11): p. 4605-14.
2. Rodes L, Saha S, Tomaro-Duchesneau C and Prakash S. Microencapsulated Bifidobacterium longum subsp. infantis ATCC 15697 favorably modulates gut microbiota and reduces circulating endotoxins in F344 rats. Biomed Res Int, 2014. 2014: p. 602832.
3. Kabeerdoss J, Ferdous S, Balamurugan R, Mechenro J, Vidya R, Santhanam S *et al*. Development of the gut microbiota in southern Indian infants from birth to 6 months: a molecular analysis. J Nutr Sci, 2013. 2: p. e18.
4. Furet JP, Firmesse O, Gourmelon M, Bridonneau C, Tap J, Mondot S *et al*. Comparative assessment of human and farm animal faecal microbiota using real-time quantitative PCR. FEMS Microbiol Ecol, 2009. 68(3): p. 351-62.
5. Periasamy S, Chalmers NI, Du-Thumm L and Kolenbrander PE. Fusobacterium nucleatum ATCC 10953 requires Actinomyces naeslundii ATCC 43146 for growth on saliva in a three-species community that includes Streptococcus oralis 34. Appl Environ Microbiol, 2009. 75(10): p. 3250-7.
6. Borgo PV, Rodrigues VA, Feitosa AC, Xavier KC and Avila-Campos MJ. Association between periodontal condition and subgingival microbiota in women during pregnancy: a longitudinal study. J Appl Oral Sci, 2014. 22(6): p. 528-33.
7. Arthur JC, Gharaibeh RZ, Mühlbauer M, Perez-Chanona E, Uronis JM, McCafferty J *et al*. Microbial genomic analysis reveals the essential role of inflammation in bacteria-induced colorectal cancer. Nat Commun, 2014. 5: p. 4724.
8. Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan TJ *et al*. Intestinal inflammation targets cancer-inducing activity of the microbiota. Science, 2012. 338(6103): p. 120-3.

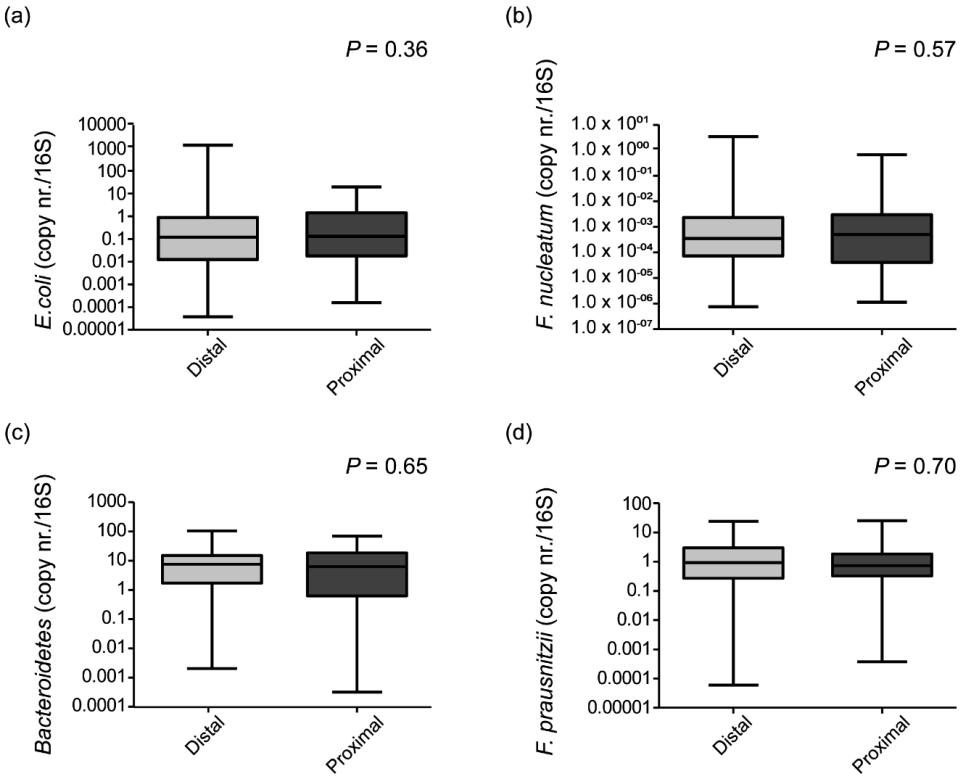
9. Huijsdens XW, Linskens RK, Mak M, Meuwissen SG, Vandenbroucke-Grauls CM, and Savelkoul PH. Quantification of bacteria adherent to gastrointestinal mucosa by real-time PCR. J Clin Microbiol, 2002. 40(12): p. 4423-7.
10. Rinttilä T, Kassinen A, Malinen E, Krogius L and Palva A. Development of an extensive set of 16S rDNA-targeted primers for quantification of pathogenic and indigenous bacteria in faecal samples by real-time PCR. J Appl Microbiol, 2004. 97(6): p. 1166-77.
11. Malinen E, Rinttilä T, Kajander K, Mättö J, Kassinen A, Krogius L *et al*. Analysis of the fecal microbiota of irritable bowel syndrome patients and healthy controls with real-time PCR. Am J Gastroenterol, 2005. 100(2): p. 373-82.
12. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, and Macfarlane S. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. Appl Environ Microbiol, 2007. 73(22): p. 7435-42.
13. Zhang M, Zhou L, Wang Y, Dorfman RG, Tang D, Xu L *et al*. Faecalibacterium prausnitzii produces butyrate to decrease c-Myc-related metabolism and Th17 differentiation by inhibiting histone deacetylase 3. Int Immunol, 2019. 31(8): p. 499-514.



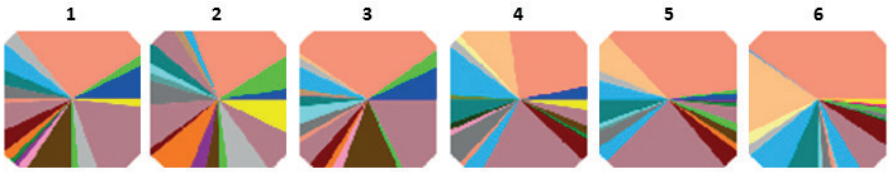
▲**Supplementary Figure S1 | (a-c)** Bacterial copy number per gram (nr./g) fecal immunochemical test (FIT) liquid for *Fusobacterium nucleatum* (a), *Bacteroidetes* (b) and *Faecalibacterium prausnitzii* (c) for the most advanced colonoscopy finding. Abbreviations: SP, serrated polyp; TA, tubular adenoma; TVA, (tubulo)villous adenoma; HGD, high grade dysplasia; CRC, colorectal cancer.



▲**Supplementary Figure S2 | (a-e)** Bacterial copy number per gram (nr./g) fecal immunochemical test (FIT) liquid for 16S (a), *Escherichia coli* (b), *Fusobacterium nucleatum* (c), Bacteroidetes (d) and *Faecalibacterium prausnitzii* (e) stratified for subjects with and without advanced neoplasia (AN).



▲**Supplementary Figure S3 | (a-d)** Absolute copy number (nr.) per 16S for *Escherichia coli* (a), *Fusobacterium nucleatum* (b), Bacteroidetes (c) and *Faecalibacterium prausnitzii* (d), stratified for the location of the most advanced finding (proximal versus distal colon).



Samples	1	2	3	4	5	6
Bacteroidetes_Bacteroidia_Bacteroidales_Bacteroidaceae_ Bacteroides	24.03	19.78	28.49	22.38	31.8	38.32
Firmicutes_Clostridia_Clostridiales_Ruminococcaceae_ Faecalibacterium	3.24	1.05	2.54	3.49	3.38	4.41
Proteobacteria_Gammaproteobacteria_Enterobacteriales_ Enterobacteriaceae; other	0.46	0.14	0.49	0.08	0.11	0.00
Fusobacteria_Fusobacteriia_Fusobacteriales_ Fusobacteriaceae_Fusobacterium	0.00	0.03	0.00	0.00	0.00	0.00

▲Supplementary Figure S4 | 16S rRNA gene sequencing pilot with pooled samples. Taxonomic classification at genus level is displayed for six pooled samples containing 1) no endoscopic findings, 2) tubular adenoma, 3) high grade dysplasia, 4) colorectal cancer, 5) sessile adenoma and 6) hyperplastic polyp. The relative abundance of specific bacterial genera (%) is given for each pooled sample.

6

CHAPTER 6

Prediction and treatment of radiation enteropathy: can intestinal bugs lead the way?

S.Y. Lam[1], M.P. Peppelenbosch[1], G.M. Fuhler[1]

[1] Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands.

Clin Cancer Res.2019 Nov 1;25(21):6280-6282.

CCR translations - Commentary on Reis Ferreira et al., 2019

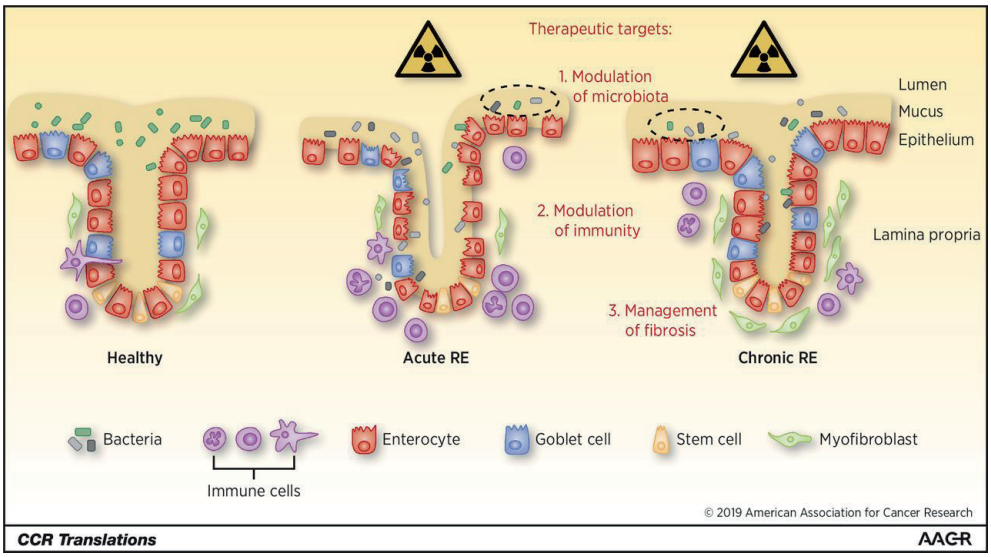
Abstract

Radiation-induced gastrointestinal toxicity is a significant comorbidity affecting many cancer patients. Intestinal microbial changes are observed in patients suffering from radiation enteropathy, although a causal relationship with disease activity has yet to be proven. Implementation of bacterial profiling in clinical care could improve recognition and management of this debilitating disease.

In this issue of Clinical Cancer Research, Ferreira and colleagues investigated the contribution of bacterial microbes in the complex pathogenesis of radiation enteropathy (RE) and suggested that microbial profiles might be relevant in the prediction, prevention or treatment of RE in the future [1]. Many cancer patients suffer from gastrointestinal (GI) symptoms after radiotherapy treatment, but a better recognition of symptoms as well as a better arsenal of treatment modalities is needed to care for these patients. New findings in translational microbiome research may serve as a tool in the assessment, identification and management of patients at risk of RE.

Radiation therapy plays a major role in the treatment and survival of oncology patients, but is accompanied by severe adverse side-effects. The intestines and the colon in particular are affected when the pelvis or the abdomen is included in the radiation field, resulting in early (acute) and/or delayed (chronic) RE. While early radiation toxicity generally occurs during treatment exposure and subsides thereafter, delayed RE is now stated to be one of the most common causes of gastrointestinal complaints, with a prevalence reported to exceed even that of inflammatory bowel diseases (IBD) [2]. Clinically, IBD and RE share similarities, as both are characterized by bloody stool and diarrhea and both are accompanied by mucosal immune cell infiltrate and inflammation, epithelial barrier breach (in case of acute RE) and fibrosis (in case of chronic RE), as well as alterations in the intestinal microbiota (**Figure 1**).

Modulation of IBD disease activity through manipulation of the microbiome is now receiving vast attention, and it is therefore only logical that RE should follow suit. The use of probiotics, specific diets and fecal microbial transplants (FMT) have all been advocated for IBD treatment, and may also hold promise for management of RE. The intestinal bacterial community indeed shows clear signs of disbalance post radiation [3-5], although it is not fully understood whether radiation-induced GI symptoms are caused by the disruption of the microbiota or *vice versa*. An altered bacterial profile prior to radiotherapy was demonstrated in patients who subsequently developed diarrhea, suggesting that pre-existing changes in the gut microbiota exist [4]. *Croprococcus* was evidently enriched before radiation in patients at risk of RE and may serve as a potential biomarker [5]. The innovative microbiota and radiotherapy-induced gastrointestinal side effects (MARS) study of Ferreira *et al.* expands on findings by investigating both acute and late effects of RE by using fecal samples from these cohorts in addition to mucosal biopsies from patients with ≥1 year of follow-up [1]. Importantly, this largest clinical study to date confirmed a trend for lower fecal bacterial diversity prior to development of acute RE and found a non-significant pattern in the late RE group.



▲**Figure 1 | The pathology of radiation enteropathy (RE) and its potential treatment options.** The healthy epithelium sustains damage upon irradiation, which causes a decrease in barrier function and translocation of luminal bacteria. The composition of the bacterial flora is altered in patients with acute RE, which may contribute to an inflammatory response which can resolve the bacterial influx, but may also cause collateral damage through the production of reactive oxygen species and inflammatory cytokines. Ferreira *et al.* now show that while microbial balance may still exist in chronic RE to some extent, there is also an altered immunological state, as evidenced by altered cytokine levels in mucosal biopsies from patients with ≥1 year of follow up after radiotherapy. Treatment of RE may in future be aimed at modulating all aspects of RE, including the microbiota, its interaction with the immune system, and immune-mediated fibrotic responses.

Mouse models have allowed mechanistic investigation of the microbiota in radiotherapy-induced GI toxicity, and potential treatments thereof. A significant shift in gut microbial composition restricted to the damaged mucosa was shown in irradiated mice [6]. Interestingly, germ free mice that were inoculated with the microbiota of these irradiated mice showed more severe pathological features following irradiation, demonstrating that bacterial disbalance may drive radiation-induced toxicity. FMT improved GI functionality and intestinal integrity after irradiation in a mouse model, suggesting that resetting the microbiota may be beneficial for the treatment of RE [7]. In humans, modulation of the microbial community for the treatment of RE symptoms has so far predominantly been tested with the use of probiotics, which appears to be effective and safe [8]. Furthermore, application of low fermentable oligosaccharides, disaccharides, monosaccharides and polyols (FODMAP) diet was shown to improve symptoms in a small pilot study in RE patients with symptoms of irritable bowel syndrome, but whether the GI microbial community was involved was not investigated [9]. Of note, however, Ferreira *et al.* did not identify any specific

changes in mucosal microbiota in late RE. Furthermore, when investigating specific fecal taxa, a link between short-chain fatty acids producing bacteria (e.g. *Roseburia*, *Clostridium IV* and *Phascolarctobacterium*) and disease severity was observed, which is surprising given that these bacteria are generally associated with a healthy microbiome. Thus, to what extent direct modulation of the microbiota in humans may contribute to prevention or treatment of RE remains to be seen.

It is conceivable that restoring the balance between host immunity and the microbiota is equally important for resolution of RE symptoms. Similar to IBD [10], Ferreira *et al.* showed a clear relationship between microbial signatures and cytokine patterns, with a depletion of cytokines regulating intestinal homeostasis (IL-7, IL-12/IL-23p40, IL-15, IL-16) and an inverse correlation of IL-15 with *Roseburia* and *Propionibacterium*. These structural and functional changes emphasize the complexity of RE, and demonstrate the existence of an immunity-microbiome axis in RE. In mice, IL-1β secretion was found to be a major mediator in the sequelae of microbial induced radiation damage, suggesting that treatment with IL-1 inhibitors might be considered in clinical practice [6]. Furthermore, the mortality of gastrointestinal acute radiation syndrome (GIARS) mice treated with 7-9 GY was shown to be associated with bacterial translocation, but could be controlled by modulation of macrophage polarization using CCL1 antisense oligodeoxynucleotide (ODN) therapy [11]. Hydrogen-water therapy similarly improved GI tract function and epithelial integrity, but its mechanism of action was via the reduction of the innate immune cell bacterial sensor MyD88 in the small intestines [12]. Altogether, these murine studies show that modulation of immunity may resolve microbial disbalance and alleviate RE symptoms. Importantly, inhibition of immune reactions has also been shown beneficial for prevention of immune-mediated fibrosis in IBD patients, and may be speculated to also alleviate fibrosis during chronic RE. For patients with IBD, many immune modulators are now used in clinical practice. For instance, increased levels of IL12/IL23p40 and TNFα are seen in IBD patients, and treatment with anti-p40 and anti-TNFα antibodies is now commonly prescribed. However, while a non-significant increase in TNFα levels in RE may indicate that some patients might benefit from anti-TNFα treatment, RE patients show reduced rather than increased mucosal levels of IL12/IL23p40 [1], making anti-p40 treatment obsolete. Thus, treatment strategies for RE based on immune modulation need to be carefully tailored.

The classic 'target cell theory' with the epithelium as the only determinant of early pathology has currently been abandoned and replaced by the concept that epithelial injury, the microvasculature, the immune system, the enteric nervous system, the intestinal microbiota and host factors are involved in pathogenesis of RE [2]. It has also

been underscored that acute and chronic RE are distinct features, and that delayed radiation induced GI toxicity symptoms are not uncommon and often progressive in nature. Although the clinical implementation of bacterial microbes in toxicological effects of radiation therapy is in its infancy, the findings thus far have been promising. Future work needs to combine clinical and laboratory efforts to establish optimal recognition and management strategies for patients with RE. A baseline screening tool to predict patient responses to cancer therapy and the risk for treatment related toxicity would be beneficial for a more personalized treatment approach.

References

1. Reis Ferreira M, Andreyev J, Mohammed K, Truelove L, Gowan SM, Li J *et al*. Microbiota and radiotherapy-induced gastrointestinal side-effects (MARS) study: a large pilot study of the microbiome in acute and late radiation enteropathy. Clin Cancer Res, 2019.
2. Hauer-Jensen M, Denham JW and Andreyev HJ. Radiation enteropathy--pathogenesis, treatment and prevention. Nat Rev Gastroenterol Hepatol, 2014. 11(8): p. 470-9.
3. Nam YD, Kim HJ, Seo JG, Kang SW and Bae JW. Impact of pelvic radiotherapy on gut microbiota of gynecological cancer patients revealed by massive pyrosequencing. PLoS One, 2013. 8(12): p. e82659.
4. Wang A, Ling Z, Yang Z, Kiela PR, Wang T, Wang C *et al*. Gut microbial dysbiosis may predict diarrhea and fatigue in patients undergoing pelvic cancer radiotherapy: a pilot study. PLoS One, 2015. 10(5): p. e0126312.
5. Wang Z, Wang Q, Wang X, Zhu L, Chen J, Zhang B *et al*. Gut microbial dysbiosis is associated with development and progression of radiation enteritis during pelvic radiotherapy. J Cell Mol Med, 2019. 23(5): p. 3747-3756.
6. Gerassy-Vainberg S, Blatt A, Danin-Poleg Y, Gershovich K, Sabo E, Nevelsky A *et al*. Radiation induces proinflammatory dysbiosis: transmission of inflammatory susceptibility by host cytokine induction. Gut, 2018. 67(1): p. 97-107.
7. Cui M, Xiao H, Li Y, Zhou L, Zhao S, Luo D *et al*. Faecal microbiota transplantation protects against radiation-induced toxicity. EMBO Mol Med, 2017. 9(4): p. 448-461.
8. Wang YH, Yao N, Wei KK, Jiang L, Hanif S, Wang ZX *et al*. The efficacy and safety of probiotics for prevention of chemoradiotherapy-induced diarrhea in people with abdominal and pelvic cancer: a systematic review and meta-analysis. Eur J Clin Nutr, 2016. 70(11): p. 1246-1253.
9. Larsen T, Hausken T, Otteraaen Ystad S, Hovdenak N, Mueller B, and Lied GA. Does the low FODMAP diet improve symptoms of radiation-induced enteropathy? A pilot study. Scand J Gastroenterol, 2018. 53(5): p. 541-548.
10. van der Giessen J, Binyamin D, Belogolovski A, Frishman S, Tenenbaum-Gavish K, Hadar E *et al*. Modulation of cytokine patterns and microbiome during pregnancy in IBD. Gut, 2019.
11. Suzuki F, Loucas BD, Ito I, Asai A, Suzuki S, and Kobayashi M. Survival of Mice with Gastrointestinal Acute Radiation Syndrome through Control of Bacterial Translocation. J Immunol, 2018. 201(1): p. 77-86.
12. Xiao HW, Li Y, Luo D, Dong JL, Zhou LX, Zhao SY *et al*. Hydrogen-water ameliorates radiation-induced gastrointestinal toxicity via MyD88's effects on the gut microbiota. Exp Mol Med, 2018. 50(1): p. e433.

7

CHAPTER 7

Value of *VAV3* methylation in stool DNA might be restricted to non-thiopurine-treated inflammatory bowel disease patients

S.Y. Lam[1], J. Yu[2], M.P. Peppelenbosch[1]

[1] Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands; [2] Department of Medicine and Therapeutics, Institute of Digestive Disease, State Key Laboratory of Digestive Disease, Li Ka Shing Institute of Health Sciences and CUHK-Shenzhen Research Institute, The Chinese University of Hong Kong, Hong Kong.

Clin Gastroenterol Hepatol. 2020 Feb;18(2):520.

Commentary on Kisiel et al., 2019

Fecal testing in general and fecal immunochemical testing (FIT) for hemoglobin in particular have become important screening modalities for the prevention and early detection of sporadic colorectal cancer (CRC). Constant improvements in feces-based screening methodology have now resulted in satisfactory sensitivity and specificity, even in population-wide approaches [1]. However, the use of FIT is less suitable for specific patient populations at risk for CRC, especially in patients with inflammatory bowel disease (IBD) which often display colonic luminal bleeding due to ulceration. In this case, fecal hemoglobin is not a reliable marker for the detection of high-grade dysplasia (HGD) or neoplastic growth and alternative strategies are needed.

The cancer process is characterized by alternative epigenetic modification of DNA and testing for alternative methylation in stool DNA has been proposed as a screening strategy for CRC development [2]. This has not been translated in specific protocols suitable for clinical implementation for the monitoring of potential cancer development in patients with IBD. Therefore, we have read with great interest the article "Analysis of DNA Methylation at Specific Loci in Stool Samples Detects Colorectal Cancer and High-Grade Dysplasia in Patients With Inflammatory Bowel Disease" by Kisiel *et al.* which has appeared in the April 2019 issue of your journal. This study has greatly documented the usefulness of determining methylated levels of the *BMP3* and *VAV3* genes in a multitarget stool DNA test [3]. Although this study represents an important breakthrough as it allows for non-invasive screening of HGD and CRC in the high-risk group of IBD patients, the manuscript does not address the potential mechanistic basis of the observations made and the possible relation of this mechanistic basis for the clinical presentation of screening results.

The bone morphogenetic protein (BMP) family represents an important family of tumor suppressive morphogens in the colon (e.g. development of juvenile polyposis is associated with mutations in the BMP pathway [4] and the association of differential BMP family orthologue methylation with the presence of HGD is not surprising. *VAV3* is a guanine nucleotide exchange factor and involved in GTP loading of p21RAC, an event linked with local disease activity in IBD, which can be clinically targeted by thiopurine therapy [5]. The study of Kisiel *et al*. does not provide information on the thiopurine-treatment history of the patients involved. In the presence of thiopurine treatment, p21RAC will be pharmacologically inactivated and under these conditions there should be no competitive advantage for lesions with altered *VAV3* methylation. Therefore, we feel that value of *VAV3* methylation for monitoring progression towards HGD and cancer, as identified by Kisiel *et al.,* is likely to be restricted to the non-thiopurine-treated fraction of IBD patients.

References

1. Wieten E, Schreuders EH, Nieuwenburg SA, Hansen BE, Lansdorp-Vogelaar I, Kuipers EJ *et al*. Effects of Increasing Screening Age and Fecal Hemoglobin Cutoff Concentrations in a Colorectal Cancer Screening Program. Clin Gastroenterol Hepatol, 2016. 14(12): p. 1771-1777.
2. Ng JM and Yu J. Promoter hypermethylation of tumour suppressor genes as potential biomarkers in colorectal cancer. Int J Mol Sci, 2015. 16(2): p. 2472-96.
3. Kisiel JB, Klepp P, Allawi HT, Taylor WR, Giakoumopoulos M, Sander T *et al*. Analysis of DNA Methylation at Specific Loci in Stool Samples Detects Colorectal Cancer and High-Grade Dysplasia in Patients With Inflammatory Bowel Disease. Clin Gastroenterol Hepatol, 2019. 17(5): p. 914-921 e5.
4. Hardwick JC, Kodach LL, Offerhaus GJ and van den Brink GR. Bone morphogenetic protein signalling in colorectal cancer. Nat Rev Cancer, 2008. 8(10): p. 806-12.
5. Parikh K, Zhou L, Somasundaram R, Fuhler GM, Deuring JJ, Blokzijl T *et al*. Suppression of p21Rac signaling and increased innate immunity mediate remission in Crohn's disease. Sci Transl Med, 2014. 6(233): p. 233ra53.

8

CHAPTER 8

A microbiome study to explore the gut-skin axis in hidradenitis suppurativa

S.Y. Lam[1*], D. Radjabzadeh[2*], H. Eppinga M[1,3], Y.R.A. Nossent[1], H.H. van der Zee[3], R. Kraaij[2], S.R. Konstantinov[1], G.M. Fuhler[1], E.P. Prens[3], H.B. Thio[3], M.P. Peppelenbosch[1]
* These authors contributed equally

[1] Department of Gastroenterology and Hepatology, [2] Department of Internal Medicine,[3] Department of Dermatology, Erasmus MC University Medical Center Rotterdam, the Netherlands.

Based on J Dermatol Sci. 2021 Mar;101(3):218-220.

Abstract

Background: The simultaneous occurrence of gut and skin diseases has led to the postulation of a gut-skin axis in which the gut microbiota are thought to be involved. Although dysbiosis of the cutaneous microbiome is observed in the chronic inflammatory skin disease hidradenitis suppurativa (HS), it remains unknown whether gastrointestinal microbial disturbances are engaged in the complex pathogenesis.

Objectives: This exploratory 16S ribosomal RNA (rRNA) amplicon sequencing study investigates fecal microbiome signatures in HS.

Methods: 17 HS and 20 healthy subjects were recruited to donate feces for bacterial DNA isolation and 16S rRNA amplicon sequencing. A subset provided an axillary skin sample for concurrent analysis.

Results: The fecal microbiome in HS and healthy subjects did not show differences in biodiversity (bacterial richness [P=0.483], Shannon [P=0.821] and inverse Simpson [P=0.916] indices), nor in community structure (Bray-Curtis [P=0.106] and Jaccard [P=0.103] metrics). Nevertheless, several bacterial taxa were differentially abundant in feces of patients and healthy controls, including the enrichment of *Robinsoniella* and depletion of Christensenellaceae in HS. The cutaneous microbiome in HS did not differ from controls.

Conclusion: Microbial differences in the overall composition of the fecal microbiome were not apparent in HS, but individual taxonomic bacterial differences were found. Future studies are warranted to elucidate the role of the microbiota as potential part of the gut-skin axis.

Keywords:

Gastrointestinal microbiome; hidradenitis suppurativa; gut; skin; gut-skin axis.

1 Introduction

The epithelial barrier of the human body is the first line of defence against microbial invasion. Whereas the inner surface is largely lined by epithelium of the gastrointestinal (GI) tract, the skin forms the protective outer layer. Both the gut and skin are home to a diverse range of bacterial microbes as well as viruses, fungi and yeasts, the dynamics of which are normally kept in balance to maintain homeostasis. Phyla Firmicutes and Bacteroidetes are the major bacterial representatives of the gut [1], but also constitute a part of the skin microbiome, which is predominated by Actinobacteria [1-5]. While the skin microbiome is further shaped by physiological microenvironments at sebaceous, moist and dry skin sites [1-6], it has become evident that gut and skin bacterial communities constitute distinct ecosystems [1]. Nevertheless, the concurrent existence of gut and skin diseases has led to the postulation of a gut-skin axis, in which gut microbiota are believed to be important contributors to skin pathologies [7-10].

Hidradenitis suppurativa (HS) is a chronic recurrent inflammatory disease of the hair follicle which mainly affects the intertriginous skin areas by the formation of nodules, sinus tracts and abscesses. The underlying mechanism of this debilitating disease is complex and not fully understood, but the inflammatory nature of cutaneous manifestations has led to the investigation of microbial factors using culture based approaches [11-14], in addition to high throughput sequencing [12-17]. The skin microbiome of healthy and HS subjects differs [14-16], but also lesional and non-lesional skin sites in HS appear to cluster separately [13]. Large bacterial aggregates have been demonstrated in HS lesions [18], and more in depth investigation indicated the predominance of anaerobic bacteria (*Porphyromonas spp.* and *Prevotella spp.*) in these lesions [13, 16, 17]. While dysbiosis of the cutaneous microbiome is apparent in HS, it remains unknown whether the gut microbiota are altered in this chronic disease. Metabolic by-products of bacteria might enter the systemic circulation and influence skin physiology, but also the immune and neuroendocrine system are thought to be part of this complex communication between the gut and skin [7-9].

Considering the potential systemic effects of the gut, perturbation of the microbiota might exert negative effects at distant skin sites. A role for the microbes in a gut-skin axis is widely postulated and microbial alterations have been demonstrated in the stool of psoriasis patients [19-22]. However, the fecal bacteria have not been fully explored in HS. Previously, we did not found differences between patients and healthy controls for fecal bacterial markers *Faecalibacterium prausnitzii* and *Escherichia coli* [23]. Here, we performed a more in depth analysis of the fecal microbiome using 16S ribosomal RNA (rRNA) amplicon sequencing.

2 Material and methods

2.1 Study population

A total of 17 HS patients and 20 healthy controls were enrolled in this exploratory study (**Table 1**). Patients were recruited after verification of the diagnosis by an experienced dermatologist at the department of Dermatology of the Erasmus MC University Medical Center Rotterdam, the Netherlands. No oral antibiotics and no topical treatment with antibiotics/steroids were allowed in respectively eight weeks and seven days before sampling. Other exclusion criteria were inflammatory bowel disease (IBD), active infection, malignancy and pregnancy. Clinical data collection included age, gender, ethnicity, body mass index (BMI), smoking and disease activity according to the Hurley clinical staging system. The control group consisted of healthy subjects without any cutaneous and GI manifestations. Informed consent was obtained from all study subjects and the study was approved by the medical ethical committee of the Erasmus MC University Medical Center Rotterdam (METC number MEC-2014-371). More details of individual study subjects are described in **Supplementary Table S1**.

2.2 Sample collection

Fecal samples were collected from all study subjects (n=37) and stored at -80˚C within 48 hours of collection (**Table 1)**. A subset of healthy controls (n=7) and HS patients (n=6) also provided an axillary skin sample, resulting in 7 healthy skin swabs, 6 HS lesional swabs and 3 paired non-lesional swabs (**Supplementary Table S1**). Since both armpits were affected in 3 out of 6 HS subjects, a non-lesional axillary sample could not be obtained from these patients. Sterile cotton swabs were applied for 30 seconds of rubbing of the skin without opening/manipulating of HS lesions and directly preserved in a solution containing phosphate buffered saline (PBS) and cell lysis buffer (Life Technologies, Carlsbad, CA) at a 1:1 ratio for storage at -20˚C.

2.3 Bacterial DNA extraction and 16S rRNA amplicon sequencing

Bacterial DNA was isolated from fecal specimens (20mg) as previously described [23]. In short, 1ml of cell lysis buffer (Life Technologies) was added to the sample, the suspension vortexed and incubated for 15 minutes at room temperature. After three times 30 seconds of bead-beating, samples were centrifuged, whereupon protein precipitation buffer was added at a 3:1 ratio. Following centrifugation, DNA was precipitated from the supernatant sing 100% isopropanol at 1:1 ratio, the pellet washed with 100µl 70% ethanol and the DNA udissolved in 50µl Tris-EDTA (TE) buffer. All skin samples were concordantly processed

Table 1 | Baseline characteristics of hidradenitis suppurativa (HS) patients and healthy controls

	HS		Healthy		*P*-value
General demographics					
Total	17		20		-
Age, median (range), y	47.0	(25-60)	29.5	(22-63)	0.053[1]
Female, n (%)	11	(64.7)	13	(65.0)	1.000[2]
Smoking, n (%)	11	(64.7)	1	(5.0)	0.000[2]
BMI, median (range), y	27.1	(22.7-36.4)	22.9	(20.4-37.2)	0.011[1]
Caucasian, n (%)	13	(76.5)	15	(75.0)	1.000[2]
HS characteristics					
Age at diagnosis, median (range), y	29	(15-55)	-		-
Disease duration, mean (SD), y	14.1	(8.9)	-		-
Therapy					
Topical, *n* (%)[a]	9	(52.9)	-		-
Anti-TNFα, n (%)	1	(5.9)	-		-
HS severity[b]					
Hurley score, n (%)					
I	2	(11.8)	-		-
II	14	(82.4)	-		-
III	1	(5.9)	-		-
Sample collection					
Feces, n (%)	17	(100)	20	(100)	-
Skin swab, n (%)					
Non-lesion, n (%)	3	(17.6)	-		-
HS lesion, n (%)	6	(35.3)	-		-
Healthy skin, n (%)	-		7	(35.0)	-

[a] Topical medication was not applied seven days before sample collection

[b] Disease severity scores using the Hurley clinical staging system (I, II, III)

[1] Mann-Whitney U test.

[2] Fisher exact test.

Abbreviations: BMI, body mass; n, number; SD, standard deviation; TNF, tumor necrosis factor; y, years.

and stored at -20˚C until further downstream analysis. 16S rRNA amplicon sequencing was performed at the Macrogen Institute in Seoul, Korea. The hyper variable region V3-V4 of the 16S rRNA gene was amplified using the Bakt_341 and Bakt_805R primers (forward: 5'-CCTACGGGNGGCWGCAG-3', reverse: 5'- GACTACHVGGGTATCTAATCC-3') and amplicons sequenced on the Illumina MiSeq platform (Illumina, San Diego, CA).

2.4 Analysis of clinical data

Baseline characteristics were reported using frequency (percentage) for categorical variables and mean (standard deviation) or median (range) for continuous variables. All analyses were performed with SPSS statistics Version 24.0 (IBM, Armonk, NY, USA). Fisher exact or Mann-Whitney U tests were applied for the comparison of respectively categorical and continuous variables. The threshold of significance was set at a *P*-value of <0.05.

2.5 Analysis of 16S rRNA amplicon sequences

The 16S rRNA amplicon sequences have been deposited in the NCBI SRA database under the BioProject ID: PRJNA687260. The raw data was processed to filter out low quality reads, to merge read pairs and to remove chimeras. Non-chimeric reads were clustered into operational taxonomic units (OTUs) using a 97% similarity cut-off. Taxonomy assignment was conducted using the 16S rRNA gene reference of the SILVA database (version 128) [24] and RDP classifier (version 2.12) [25]. To control for differences in sequencing depth between samples, rarefaction at a depth of 11.500 reads was applied to the dataset. Bar plots were used to display the relative abundance of individual samples and sample groups at phylum and genus levels. Alpha diversity was assessed by calculation of the bacterial richness, Shannon Index and inverse Simpson Index at genus level. Group results were visualized in boxplots and statistical differences between groups tested using the Wilcoxon test with a *P*-value of <0.05 as cut-off. Multidimensional scaling (MDS) plots were constructed using Bray-Curtis and Jaccard metrics for group comparison based on bacterial community composition at genus level. Statistical differences were tested with permutational multivariate analysis of variance (PERMANOVA), the 'adonis' function of the vegan package [26] in R. To compare differential abundance of taxa between groups, linear discriminant analysis (LDA) scores were calculated using the linear discriminant analysis effect size (LEfSe) tool (with α parameter set to 0.05 and logarithmic LDA score set to 4.0) [27]. Considering potential confounding effects of body weight (BMI≤25 versus BMI>25) and smoking status (smokers versus non-smokers) on the fecal microbiota, α- and β diversity were investigated in a sensitivity analysis for HS subjects. LEfSe analysis permitted the comparison of fecal samples of HS and healthy subjects with respect to subclasses BMI and smoking status.

3 Results

3.1 Patient characteristics

The fecal microbiome was characterized in 17 HS patients and 20 healthy controls, while

the cutaneous microbiome was studied in a subset (**Supplementary Table S1**). Overall, HS patients were slightly older (median age 47.0 (range 25–60) vs 29.5 (range 22–63) years, $P = 0.053$), were more frequently smokers (64.7 vs 5.0 %, $P = 0.000$) and had a higher median BMI (27.1 vs 22.9, $P = 0.011$) in comparison to healthy controls (**Table 1**).

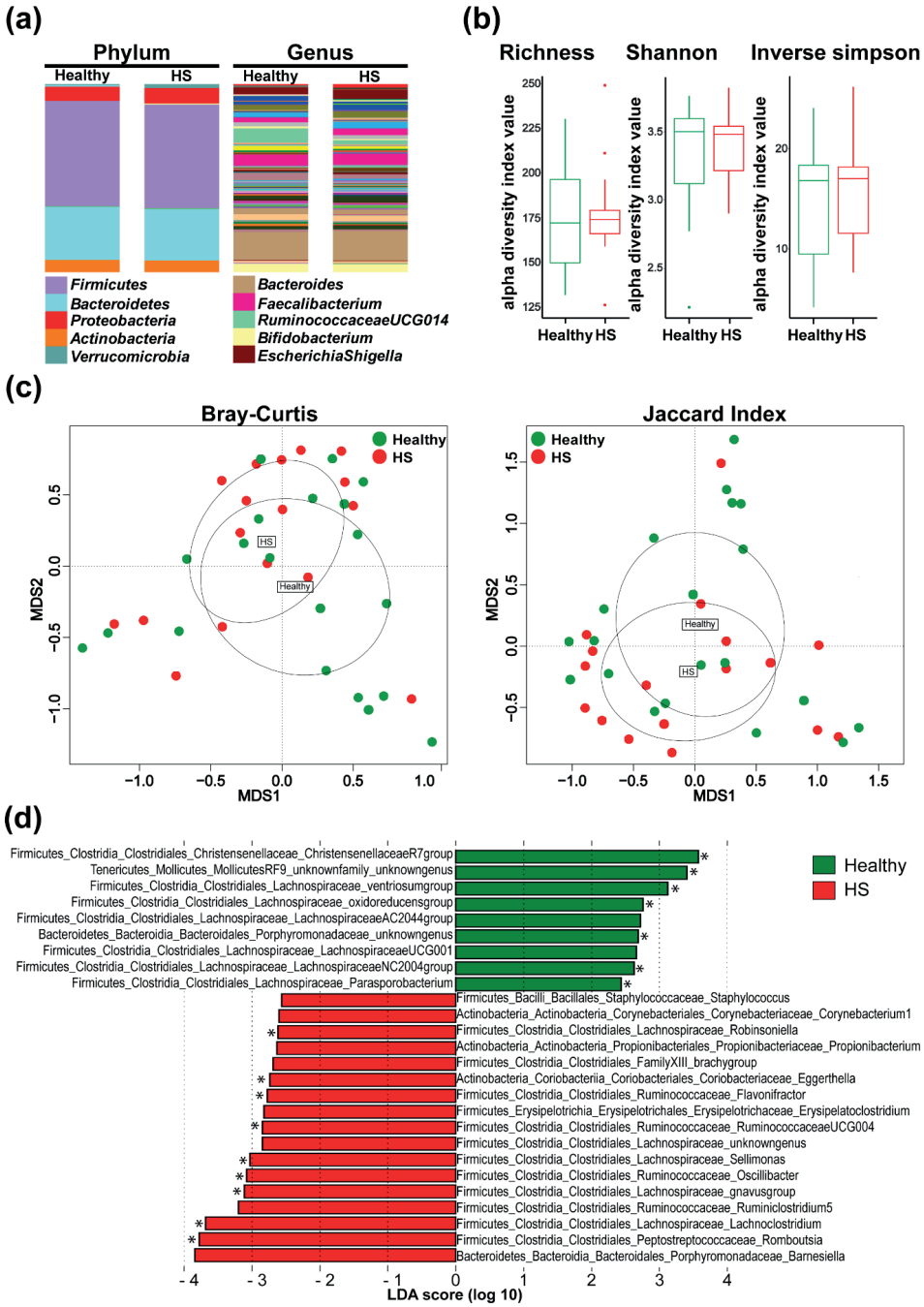
3.2 Fecal bacterial taxonomic differences between HS patients and controls

Microbial characterization confirmed that the gut and skin are home to distinct bacterial communities (**Supplementary Figure S1**). The fecal microbiome of both HS and controls was predominated by Firmicutes (55.3% and 56.0%, respectively) at phylum level, with *Bacteroides* (15.6% and 14.9%) presenting the most abundant genus (**Figure 1a**). There were no differences between HS patients and controls in bacterial richness (P=0.483), Shannon (P=0.821) and inverse Simpson (P=0.916) indices (**Figure 1b**), nor in bacterial community structure based on Bray-Curtis (R2=0.042, P=0.106) or Jaccard (R2=0.039, P=0.103) metrics (**Figure 1c**). In addition, subjects when stratified by smoking status or BMI (**Supplementary Figure S2**).

LEfSe analysis indicated considerable taxonomic differences between feces of HS patients and healthy controls (**Figure 1d**). From 17 taxa that were more abundant in HS (**Supplementary Figure S3**), genus *Robinsoniella* was observed in HS patients (n=10; 59%) while it was absent in all healthy controls tested. Remarkably, *Robinsoniella* was not found in the patient with Hurley class III in contrast to *Barnesiella* which was particularly abundant (19.4%) in this individual. While the majority of taxa were present at different abundance in both HS and healthy control groups, *Sellimonas* was more commonly found in patients (n=13; 76%) than controls (n=5; 25%). In contrast, a member of the Christensenellaceae family was depleted in HS patients in comparison to healthy controls (**Supplementary Figure S4**). Another 8 taxa that were also depleted in HS included several representatives of family Lachnospiraceae. Moreover, the majority of differentially abundant bacterial taxa were also found when accounted for subclasses smoking status and BMI but not *Barnesiella* (**Figure 1d** and **Supplementary Figure S3-4**).

3.3 Skin microbiome alterations observed in a subset of HS patients

Characterization of the cutaneous microbiome in healthy (n=7) and HS (n=9) skin samples showed relative abundance profiles dominated by Firmicutes, Bacteroidetes, Proteobacteria and Actinobacteria (**Figure 2a**). Although inter-individual differences were apparent at genus level, the lesional skin samples from two HS patients (subjects 14 and


(a)
Phylum
Genus
Healthy
HS
Firmicutes
Bacteroidetes
Proteobacteria
Actinobacteria
Verrucomicrobia
Bacteroides
Faecalibacterium
RuminococcaceaeUCG014
Bifidobacterium
EscherichiaShigella
(b)
Richness
Shannon
Inverse simpson
alpha diversity index value
Healthy HS
(c)
Bray-Curtis
Jaccard Index
MDS1
MDS2
(d)
Firmicutes_Clostridia_Clostridiales_ChristensenellaceaeR7group
Tenericutes_Mollicutes_MollicutesRF9_unknownfamily_unknowngenus
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_ventriosumgroup
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_oxidoreducensgroup
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_LachnospiraceaeAC2044group
Bacteroidetes_Bacteroidia_Bacteroidales_Porphyromonadaceae_unknowngenus
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_LachnospiraceaeUCG001
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_LachnospiraceaeNC2004group
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_Parasporobacterium
Firmicutes_Bacilli_Bacillales_Staphylococcaceae_Staphylococcus
Actinobacteria_Actinobacteria_Corynebacteriales_Corynebacteriaceae_Corynebacterium1
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_Robinsoniella
Actinobacteria_Actinobacteria_Propionibacteriales_Propionibacteriaceae_Propionibacterium
Firmicutes_Clostridia_Clostridiales_FamilyXIII_brachygroup
Actinobacteria_Coriobacteriia_Coriobacteriales_Coriobacteriaceae_Eggerthella
Firmicutes_Clostridia_Clostridiales_Ruminococcaceae_Flavonifractor
Firmicutes_Erysipelotrichia_Erysipelotrichales_Erysipelotrichaceae_Erysipelatoclostridium
Firmicutes_Clostridia_Clostridiales_RuminococcaceaeUCG004
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_unknowngenus
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_Sellimonas
Firmicutes_Clostridia_Clostridiales_Ruminococcaceae_Oscillibacter
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_gnavusgroup
Firmicutes_Clostridia_Clostridiales_Ruminococcaceae_Ruminiclostridium5
Firmicutes_Clostridia_Clostridiales_Lachnospiraceae_Lachnoclostridium
Firmicutes_Clostridia_Clostridiales_Peptostreptococcaceae_Romboutsia
Bacteroidetes_Bacteroidia_Bacteroidales_Porphyromonadaceae_Barnesiella
LDA score (log 10)

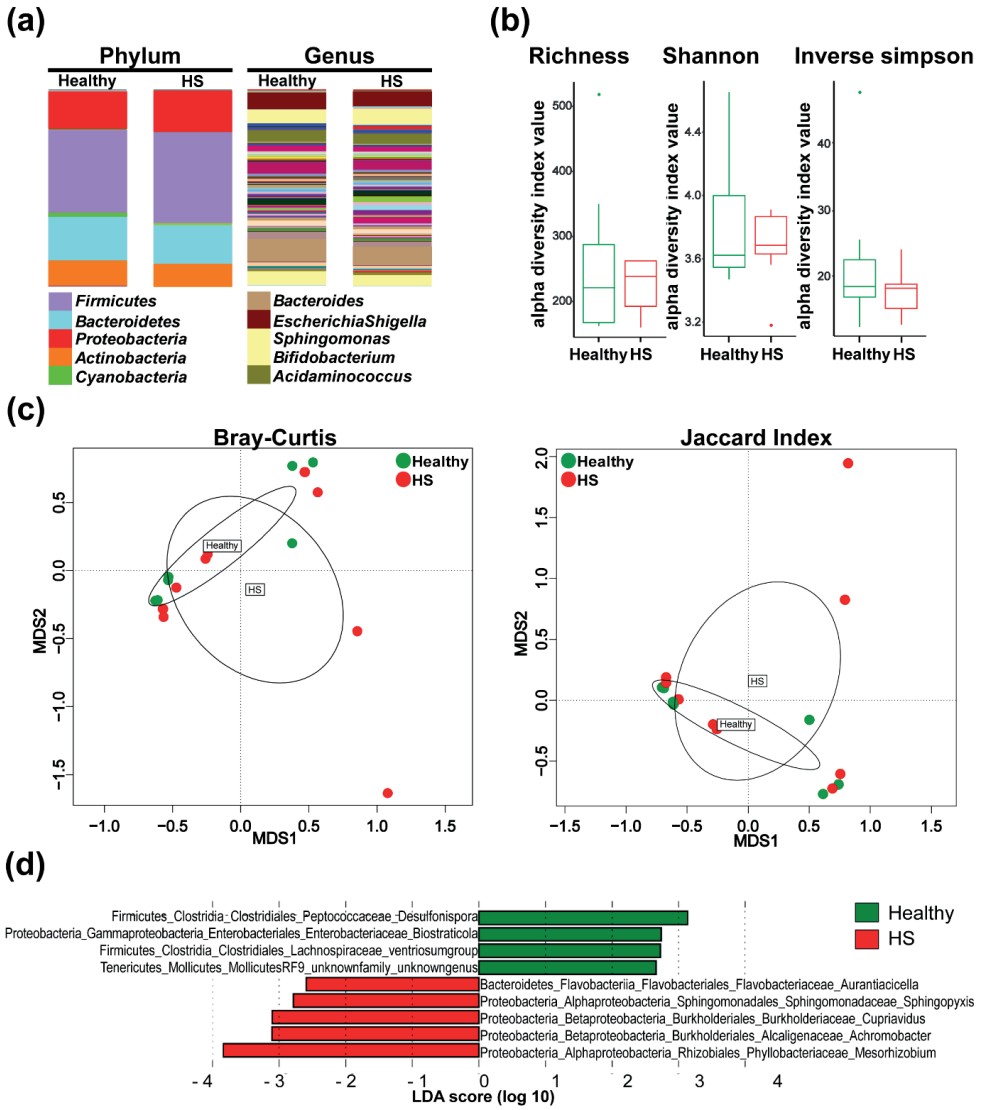
◂Figure 1 | Fecal microbial composition and individual bacterial differences between hidradenitis suppurativa (HS) patients and controls. 16S rRNA amplicon sequencing results comparing fecal samples from HS patients (n=17) and healthy controls (n=20). **(a)** Bar plots showing the relative abundance of HS and healthy control groups at phylum and genus level. **(b)** Boxplots showing the fecal bacterial biodiversity measured by richness, Shannon index and the inverse Simpson index at genus level. Each plot represents the 25-75% interquartile range and the median. **(c)** Multidimensional scaling (MDS) plots illustrate the fecal bacterial composition of HS patients and healthy controls using Bray-Curtis and Jaccard metrics at genus level. **(d)** Results of linear discriminant analysis (LDA) effect size (LEfSe) analysis for fecal samples. The histogram plot shows the differentially abundant taxa in fecal samples of HS patients and healthy controls. The asterisks (*) mark the taxa that are differently abundant among the two groups irrespective of smoking habits and body mass index (BMI).

15) exhibited a clearly different bacterial profile (**Supplementary Figure S5**). No differences in bacterial richness (*P*=0.791), Shannon (*P*=0.758) and inverse Simpson (*P*=0.536) indices were observed between healthy controls and HS patients (**Figure 2b**). While the two HS-lesional samples mentioned above separated from other samples, overall community composition did not differ between two groups based on Bray-Curtis (R2=0.055, *P*=0.507) and Jaccard (R2=0.047, *P*=0.680) metrics (**Figure 2c**). Further stratification into healthy skin, HS lesion and HS non-lesion groups did not show additional differences in α-diversity or β-diversity (not shown). Due to the small sample size, both lesion and non-lesion of the HS group combined were contrasted against the healthy skin group by LEfSe analysis (**Figure 2d**). Several features were found to be differentially abundant between these two groups. *Mesorhizobium* presented in all HS skin samples and in particular the lesional (8.4%) compared to the non-lesional (0.3%) skin site of subject 14, but the relevance and robustness of this genus remains uncertain.

4 Discussion

This exploratory 16S rRNA amplicon sequencing study investigated microbial signatures in the chronic inflammatory skin disease HS by characterization of the fecal microbiome, in addition to the cutaneous microbiome within a smaller subset. Previous findings of a distinct microbiome at the gut and the skin were confirmed [1], and in depth investigation of the fecal and cutaneous microbiome did not show overall differences in bacterial biodiversity or community composition between patients and healthy controls. However, we have found interesting alterations in the relative abundance of several individual bacterial taxa in HS. Although previous studies reported on the bacterial taxa changes of the skin in HS [13, 15], we are among the first that report various feature differences in fecal samples of HS patients.

While in health, the gut microbial community is in homeostasis with the host, bacterial perturbations have been associated with many diseases. The most striking finding in our



▲Figure 2 | Comparison of microbial composition of skin samples from hidradenitis suppurativa (HS) patients and controls. 16S rRNA amplicon sequencing results of skin samples from HS patients (n=6) and healthy controls (n=7). Samples are divided in two groups: healthy skin (n=7) and HS skin (n=9), the latter consisting of lesions (n=6) and non-lesions (n=3). **(a)** Bar plots showing the relative abundance of skin samples at phylum and genus level. **(b)** Boxplots showing the cutaneous bacterial biodiversity measured by richness, Shannon index and the inverse Simpson index at genus level. Each plot represents the 25-75% interquartile range and the median. **(c)** Multidimensional scaling (MDS) plots illustrate the bacterial composition of healthy and HS skin groups using Bray-Curtis and Jaccard metrics at genus level. **(d)** Results of linear discriminant analysis (LDA) effect size (LEfSe) analysis. The histogram plot shows the differentially abundant bacterial taxa in healthy and HS skin groups.

study was the overrepresentation of genus *Robinsoniella* in feces of HS patients. *Robinsoniella peoriensis* is the only species in this genus and has been detected in many different biological samples, including blood [28, 29], abdominal fluid aspirate [29] and (surgical) wounds [29, 30]. With the report that this anaerobic bacterium was found in the fecal content of premature neonates [31], a role of as a potential invader of the gut and pathogen in HS might be speculated. Another feature over-represented in HS was *Sellimonas*. The overabundance was demonstrated in fecal samples from patients with rheumatoid arthritis (RA) [32] and ankylosing spondylitis (AS) [33], suggesting a role of intestinal microbiota beyond the gut. The co-existence of RA, AS, psoriasis and IBD has been reported [34-38], indicating a common etiological overlap between such chronic inflammatory conditions. Indeed, alterations in intestinal bacterial features are commonly shared among patients with IBD, RA and spondyloarthritis, and include a decrease of *Faecalibacterium* [39]. Despite the fact that HS has also been reported to co-occur with IBD [40-46], an IBD-like microbial signature with a decrease of *Faecalibacterium prausnitzii* was not found in our previous study using quantitative polymerase chain reaction [23], and also not here. Since not all HS patients suffer from IBD (and indeed none of the patients in our cohort did), other specific bacterial taxa might be relevant in the complex pathogenesis of HS.

The most abundant taxon associated with HS was *Barnesiella*, but not when stratified for subclasses smoking or BMI status. This genus was previously reported to be protective against vancomycin-resistant *Enterococcus* (VRE) colonisation [47], and correlated to the healthy individuals in meta-analysis of 1252 controls and 1796 patients with intestinal disease (IBD, colorectal cancer and *Clostridium difficile* infection) [48]. Therefore, the association of *Barnesiella* with HS seems counterintuitive, but might be explained by the overrepresentation of this genus (19.4%) in the stool of the Hurley class III patient. One potentially beneficial feature that was underrepresented in HS is the *Christensenellaceae R-7 group*. Christensenellaceae is reported to be more abundant in healthy controls without gut diseases [48], and linked to human health in general [49]. While an association with lower BMI has been suggested [49, 50], this taxon was depleted in feces of HS patients compared to healthy controls irrespective of BMI in the current study.

This study has several limitations that need to be addressed. First, the collection of fecal samples as a proxy for the gut microbiota may have underrepresented the microbial communities at the epithelial barrier. Nevertheless, the fecal content still represents the luminal bacteria within the gut and is less invasive to collect. The same holds true for collection of skin swabs compared to biopsies, as the latter might provide more insight

into the bacteria of deeper cutaneous layers. Secondly, the low number of skin samples did not allow replication of previous findings of a distinct cutaneous microbiome in HS [13, 15] and precluded an investigation of the full spectrum of clinical HS patients. The inclusion of other predilection sites would have been informative, but was beyond the scope of this study. Third, 16S rRNA amplicon sequencing did not allow taxonomic classification to species level, but that was accepted in this exploratory study which allowed more insight into the fecal bacteria in HS. To date, only one small study reported on the fecal microbiome [51]. No differences in community structure were found between HS patients (n=3) and controls (n=3), but four genera (*Bilophila*, *Holdemania*, *Lachnobacterium*, *Veillonella*) were differently abundant. To confirm their and our findings, future studies are necessary.

In conclusion, individual bacterial taxa in fecal samples from HS patients support the possibility of a role for intestinal microbial alterations in this chronic inflammatory skin disease. Future studies are necessary to further investigate the gut microbiota as potential part of the gut-skin axis in HS. The inclusion of patients with different disease severity and the exploration of diet and lifestyle effects on the microbiota is warranted to better understand the aetiology of HS.

References

1. Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, and Knight R. Bacterial community variation in human body habitats across space and time. Science, 2009. 326(5960): p. 1694-7.
2. 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. 324(5931): p. 1190-2.
3. Grice EA and Segre JA. The skin microbiome. Nat Rev Microbiol, 2011. 9(4): p. 244-53.
4. Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH *et al*. Biogeography and individuality shape function in the human skin metagenome. Nature, 2014. 514(7520): p. 59-64.
5. Byrd AL, Belkaid Y and Segre JA. The human skin microbiome. Nat Rev Microbiol, 2018. 16(3): p. 143-155.
6. Findley K, Oh J, Yang J, Conlan S, Deming C, Meyer JA *et al*. Topographic diversity of fungal and bacterial communities in human skin. Nature, 2013. 498(7454): p. 367-70.
7. O'Neill CA, Monteleone G, McLaughlin JT and Paus R. The gut-skin axis in health and disease: A paradigm with therapeutic implications. Bioessays, 2016. 38(11): p. 1167-1176.
8. Vaughn AR NM, Clark AK, Sivamani RK. Skin-gut axis: The relationship between intestinal bacteria and skin health. World J Dermatol 2017; 6(4): 52-58.
9. Salem I, Ramser A, Isham N and Ghannoum MA. The Gut Microbiome as a Major Regulator of the Gut-Skin Axis. Front Microbiol, 2018. 9: p. 1459.
10. Ellis SR, Nguyen M, Vaughn AR, Notay M, Burney WA, Sandhu S *et al*. The Skin and Gut Microbiome and Its Role in Common Dermatologic Conditions. Microorganisms, 2019. 7(11).
11. Nikolakis G, Liakou AI, Bonovas S, Seltmann H, Bonitsis N, Join-Lambert O *et al*. Bacterial Colonization in Hidradenitis Suppurativa/Acne Inversa: A Cross-sectional Study of 50 Patients and Review of the Literature. Acta Derm Venereol, 2017. 97(4): p. 493-498.
12. Guet-Revillet H, Coignard-Biehler H, Jais JP, Quesne G, Frapy E, Poiree S *et al*. Bacterial pathogens associated with hidradenitis suppurativa, France. Emerg Infect Dis, 2014. 20(12): p. 1990-8.
13. Guet-Revillet H, Jais JP, Ungeheuer MN, Coignard-Biehler H, Duchatelet S, Delage M *et al*. The Microbiological Landscape of Anaerobic Infections in Hidradenitis Suppurativa: A Prospective Metagenomic Study. Clin Infect Dis, 2017. 65(2): p. 282-291.
14. Riverain-Gillet E, Guet-Revillet H, Jais JP, Ungeheuer MN, Duchatelet S, Delage M *et al*. The surface microbiome of clinically unaffected skinfolds in hidradenitis suppurativa: a cross-sectional culture based and 16s rrna gene amplicon sequencing study in 60 patients. J Invest Dermatol, 2020.
15. Ring HC, Thorsen J, Saunte DM, Lilje B, Bay L, Riis PT *et al*. The Follicular Skin Microbiome in Patients With Hidradenitis Suppurativa and Healthy Controls. JAMA Dermatol, 2017. 153(9): p. 897-905.
16. Naik HB, Jo JH, Paul M and Kong HH. Skin microbiota perturbations are distinct and disease severity-dependent in hidradenitis suppurativa. J Invest Dermatol, 2019.
17. Ring HC, Sigsgaard V, Thorsen J, Fuursted K, Fabricius S, Saunte DM *et al*. The microbiome of tunnels in hidradenitis suppurativa patients. J Eur Acad Dermatol Venereol, 2019. 33(9): p. 1775-1780.
18. Ring HC, Bay L, Nilsson M, Kallenbach K, Miller IM, Saunte DM *et al*. Bacterial biofilm in chronic lesions of hidradenitis suppurativa. Br J Dermatol, 2017. 176(4): p. 993-1000.
19. Scher JU, Ubeda C, Artacho A, Attur M, Isaac S, Reddy SM *et al*. Decreased bacterial diversity characterizes the altered gut microbiota in patients with psoriatic arthritis, resembling dysbiosis in inflammatory bowel disease. Arthritis Rheumatol, 2015. 67(1): p. 128-39.

20. Codoner FM, Ramirez-Bosca A, Climent E, Carrion-Gutierrez M, Guerrero M, Perez-Orquin JM *et al*. Gut microbial composition in patients with psoriasis. Sci Rep, 2018. 8(1): p. 3812.
21. Shapiro J, Cohen NA, Shalev V, Uzan A, Koren O, and Maharshak N. Psoriatic patients have a distinct structural and functional fecal microbiota compared with controls. J Dermatol, 2019. 46(7): p. 595-603.
22. Hidalgo-Cantabrana C, Gomez J, Delgado S, Requena-Lopez S, Queiro-Silva R, Margolles A *et al*. Gut microbiota dysbiosis in a cohort of patients with psoriasis. Br J Dermatol, 2019.
23. Eppinga H, Sperna Weiland CJ, Thio HB, van der Woude CJ, Nijsten TE, Peppelenbosch MP *et al*. Similar Depletion of Protective Faecalibacterium prausnitzii in Psoriasis and Inflammatory Bowel Disease, but not in Hidradenitis Suppurativa. J Crohns Colitis, 2016. 10(9): p. 1067-75.
24. 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. 41(Database issue): p. D590-6.
25. Wang Q, Garrity GM, Tiedje JM and Cole JR. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol, 2007. 73(16): p. 5261-7.
26. Oksanen J, Blanchet F, Friendly M, Kindt R, Legendre P, McGlinn D *et al*. Vegan: community ecology package. 2019.
27. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS *et al*. Metagenomic biomarker discovery and explanation. Genome Biol, 2011. 12(6): p. R60.
28. Jeon Y, Kim TS, Kim HB, Park KU, Song J, and Kim EC. First Korean case of Robinsoniella peoriensis bacteremia in a patient with aspiration pneumonia. Ann Lab Med, 2012. 32(5): p. 370-4.
29. Gomez E, Gustafson DR, Colgrove R, Ly T, Santana R, Rosenblatt JE *et al*. Isolation of Robinsoniella peoriensis from four human specimens. J Clin Microbiol, 2011. 49(1): p. 458-60.
30. Cotta MA, Whitehead TR, Falsen E, Moore E and Lawson PA. Robinsoniella peoriensis gen. nov., sp. nov., isolated from a swine-manure storage pit and a human clinical source. Int J Syst Evol Microbiol, 2009. 59(Pt 1): p. 150-5.
31. Ferraris L, Aires J and Butel MJ. Isolation of Robinsoniella peoriensis from the feces of premature neonates. Anaerobe, 2012. 18(1): p. 172-3.
32. Sun Y, Chen Q, Lin P, Xu R, He D, Ji W *et al*. Characteristics of Gut Microbiota in Patients With Rheumatoid Arthritis in Shanghai, China. Front Cell Infect Microbiol, 2019. 9: p. 369.
33. Chen Z, Qi J, Wei Q, Zheng X, Wu X, Li X *et al*. Variations in gut microbial profiles in ankylosing spondylitis: disease phenotype-related dysbiosis. Ann Transl Med, 2019. 7(20): p. 571.
34. Halling ML, Kjeldsen J, Knudsen T, Nielsen J and Hansen LK. Patients with inflammatory bowel disease have increased risk of autoimmune and inflammatory diseases. World J Gastroenterol, 2017. 23(33): p. 6137-6146.
35. Yang BR, Choi NK, Kim MS, Chun J, Joo SH, Kim H *et al*. Prevalence of extraintestinal manifestations in Korean inflammatory bowel disease patients. PLoS One, 2018. 13(7): p. e0200363.
36. Stolwijk C, van Tubergen A, Castillo-Ortiz JD and Boonen A. Prevalence of extra-articular manifestations in patients with ankylosing spondylitis: a systematic review and meta-analysis. Ann Rheum Dis, 2015. 74(1): p. 65-73.
37. Essers I, Ramiro S, Stolwijk C, Blaauw M, Landewe R, van der Heijde D *et al*. Characteristics associated with the presence and development of extra-articular manifestations in ankylosing spondylitis: 12-year results from OASIS. Rheumatology (Oxford), 2015. 54(4): p. 633-40.
38. Eppinga H, Poortinga S, Thio HB, Nijsten TEC, Nuij V, van der Woude CJ *et al*. Prevalence and Phenotype of Concurrent Psoriasis and Inflammatory Bowel Disease. Inflamm Bowel Dis, 2017. 23(10): p. 1783-1789.

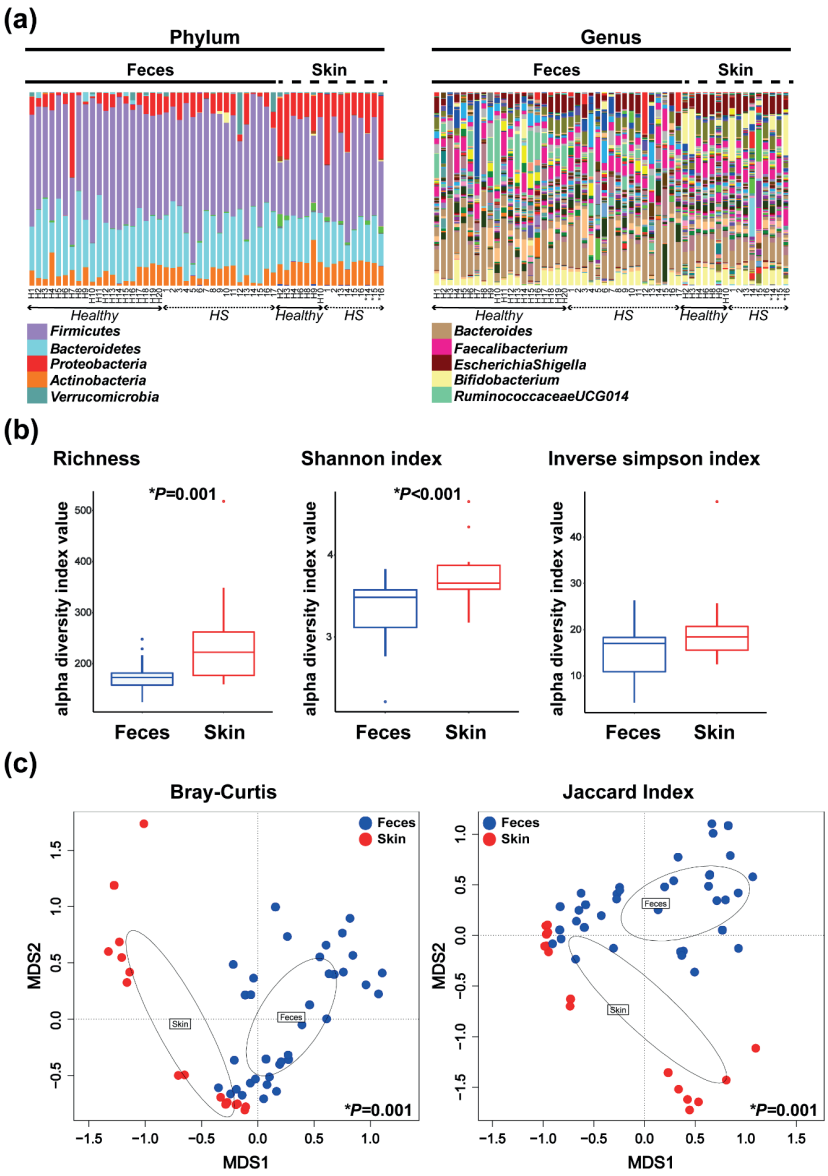
39. Salem F, Kindt N, Marchesi JR, Netter P, Lopez A, Kokten T *et al*. Gut microbiome in chronic rheumatic and inflammatory bowel diseases: Similarities and differences. United European Gastroenterol J, 2019. 7(8): p. 1008-1032.
40. van der Zee HH, van der Woude CJ, Florencia EF and Prens EP. Hidradenitis suppurativa and inflammatory bowel disease: are they associated? Results of a pilot study. Br J Dermatol, 2010. 162(1): p. 195-7.
41. van der Zee HH, de Winter K, van der Woude CJ and Prens EP. The prevalence of hidradenitis suppurativa in 1093 patients with inflammatory bowel disease. Br J Dermatol, 2014. 171(3): p. 673-5.
42. Deckers IE, Benhadou F, Koldijk MJ, Del Marmol V, Horvath B, Boer J *et al*. Inflammatory bowel disease is associated with hidradenitis suppurativa: Results from a multicenter cross-sectional study. J Am Acad Dermatol, 2017. 76(1): p. 49-53.
43. Shalom G, Freud T, Ben Yakov G, Khoury R, Dreiher J, Vardy DA *et al*. Hidradenitis Suppurativa and Inflammatory Bowel Disease: A Cross-Sectional Study of 3,207 Patients. J Invest Dermatol, 2016. 136(8): p. 1716-1718.
44. Cices A, Ibler E, Majewski S, Huynh T, Sable KA, Brieva J *et al*. Hidradenitis suppurativa association at the time of, or subsequent to, diagnosis of inflammatory bowel disease in a large U.S. patient population. J Eur Acad Dermatol Venereol, 2017. 31(7): p. e311-e312.
45. Egeberg A, Jemec GBE, Kimball AB, Bachelez H, Gislason GH, Thyssen JP *et al*. Prevalence and Risk of Inflammatory Bowel Disease in Patients with Hidradenitis Suppurativa. J Invest Dermatol, 2017. 137(5): p. 1060-1064.
46. Chen WT and Chi CC. Association of Hidradenitis Suppurativa With Inflammatory Bowel Disease: A Systematic Review and Meta-analysis. JAMA Dermatol, 2019.
47. Ubeda C, Bucci V, Caballero S, Djukovic A, Toussaint NC, Equinda M *et al*. Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. Infect Immun, 2013. 81(3): p. 965-73.
48. Mancabelli L, Milani C, Lugli GA, Turroni F, Cocconi D, van Sinderen D *et al*. Identification of universal gut microbial biomarkers of common human intestinal diseases by meta-analysis. FEMS Microbiol Ecol, 2017. 93(12).
49. Waters JL and Ley RE. The human gut bacteria Christensenellaceae are widespread, heritable, and associated with health. BMC Biol, 2019. 17(1): p. 83.
50. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R *et al*. Human genetics shape the gut microbiome. Cell, 2014. 159(4): p. 789-99.
51. Kam S, Collard M, Lam J and Alani RM. Gut Microbiome Perturbations in Patients with Hidradenitis Suppurativa: A Case Series. J Invest Dermatol, 2021. 141(1): p. 225-228 e2.

Supplementary information

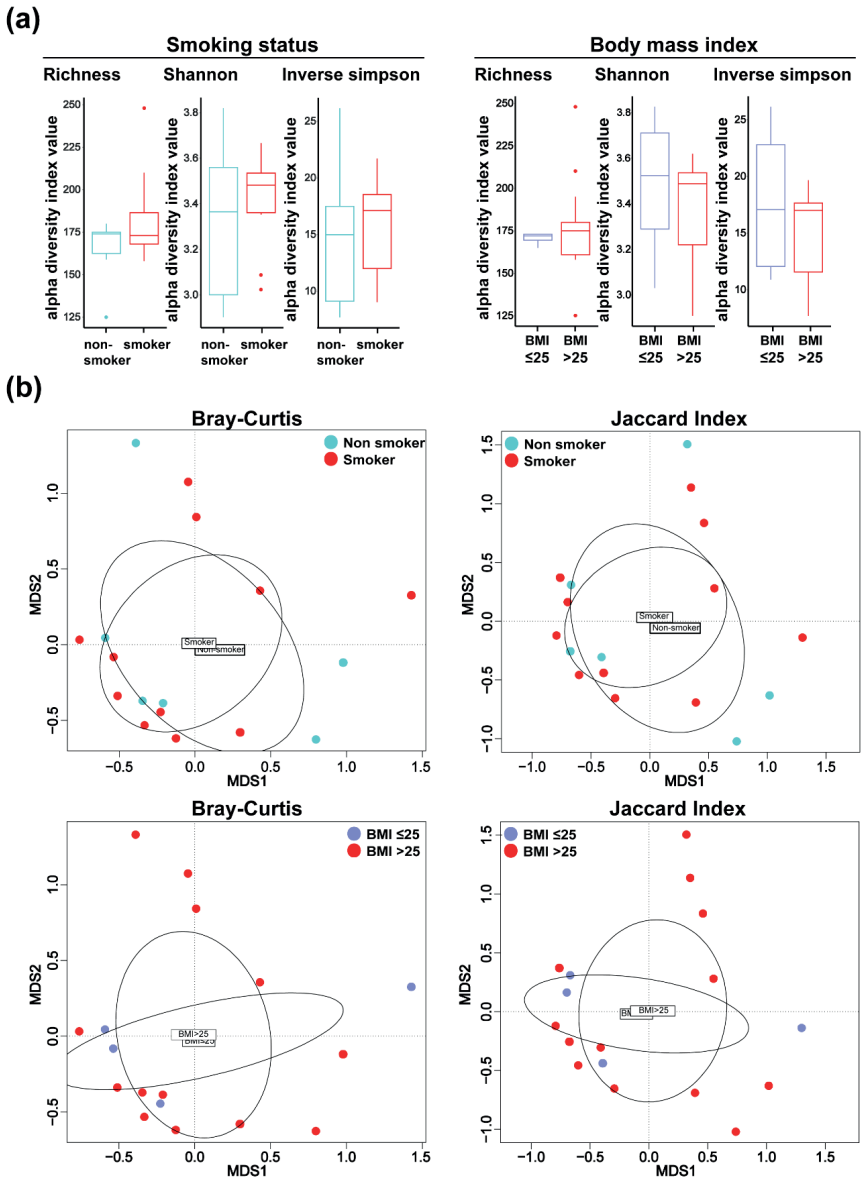
Supplementary Table S1 | Characteristics of hidradenitis suppurativa (HS) patients and healthy controls

n[a]	Gender	Age (y)	Ethnicity	Smoker	BMI	HS disease (y)[b]			Skin swab[c]
						Age	*Duration*	*Hurley*	
1	Female	25	Caucasian	Yes	36.4	16	9	I	Yes; lesion
2	Female	41	Caucasian	Yes	25.9	19	22	II	Yes; lesion
3	Female	55	Caucasian	Yes	31.1	29	26	II	No
4	Male	45	Arab-Berber	No	27.4	17	28	II	No
5	Female	51	Caucasian	Yes	26.0	38	13	I	No
6	Male	59	Caucasian	Yes	36.0	48	11	II	No
7	Male	45	Indian descent	No	25.8	23	22	II	No
8	Female	56	Caucasian	No	35.3	53	3	II	No
9	Female	44	Indian descent	No	22.9	39	5	II	No
10	Female	30	Caucasian	Yes	32.0	20	10	II	No
11	Female	47	Caucasian	Yes	23.2	15	32	II	No
12	Female	58	Caucasian	Yes	27.1	48	10	II	No
13	Male	60	Caucasian	Yes	31.7	55	5	II	Yes; lesion
14	Female	53	Afro-Caribbean	No	26.0	38	15	II	Yes; both
15	Male	32	Caucasian	Yes	22.7	18	14	III	Yes; both
16	Male	55	Caucasian	Yes	23.9	45	10	II	Yes; both
17	Female	25	Caucasian	No	28.0	21	4	II	No
H1	Female	25	Caucasian	No	22.6	-	-	-	No
H2	Female	29	Caucasian	No	22.4	-	-	-	Yes
H3	Female	29	Asian	No	22.4	-	-	-	Yes
H4	Male	24	Asian	No	30.1	-	-	-	Yes
H5	Female	29	Caucasian	No	20.4	-	-	-	No
H6	Male	30	Mixed	No	23.1	-	-	-	Yes
H7	Female	22	Asian	No	22.3	-	-	-	No
H8	Male	23	Caucasian	No	20.5	-	-	-	Yes
H9	Female	27	Mixed	No	37.2	-	-	-	Yes
H10	Female	27	Caucasian	No	22.8	-	-	-	Yes
H11	Female	26	Caucasian	No	21.5	-	-	-	No
H12	Female	53	Caucasian	No	25.8	-	-	-	No
H13	Female	52	Caucasian	No	22.4	-	-	-	No
H14	Male	52	Caucasian	No	28.1	-	-	-	No
H15	Male	54	Caucasian	Yes	28.4	-	-	-	No
H16	Female	53	Caucasian	No	29.1	-	-	-	No
H17	Male	63	Caucasian	No	25.1	-	-	-	No
H18	Female	51	Caucasian	No	23.6	-	-	-	No
H19	Female	45	Caucasian	No	28.7	-	-	-	No
H20	Male	44	Caucasian	No	20.8	-	-	-	No

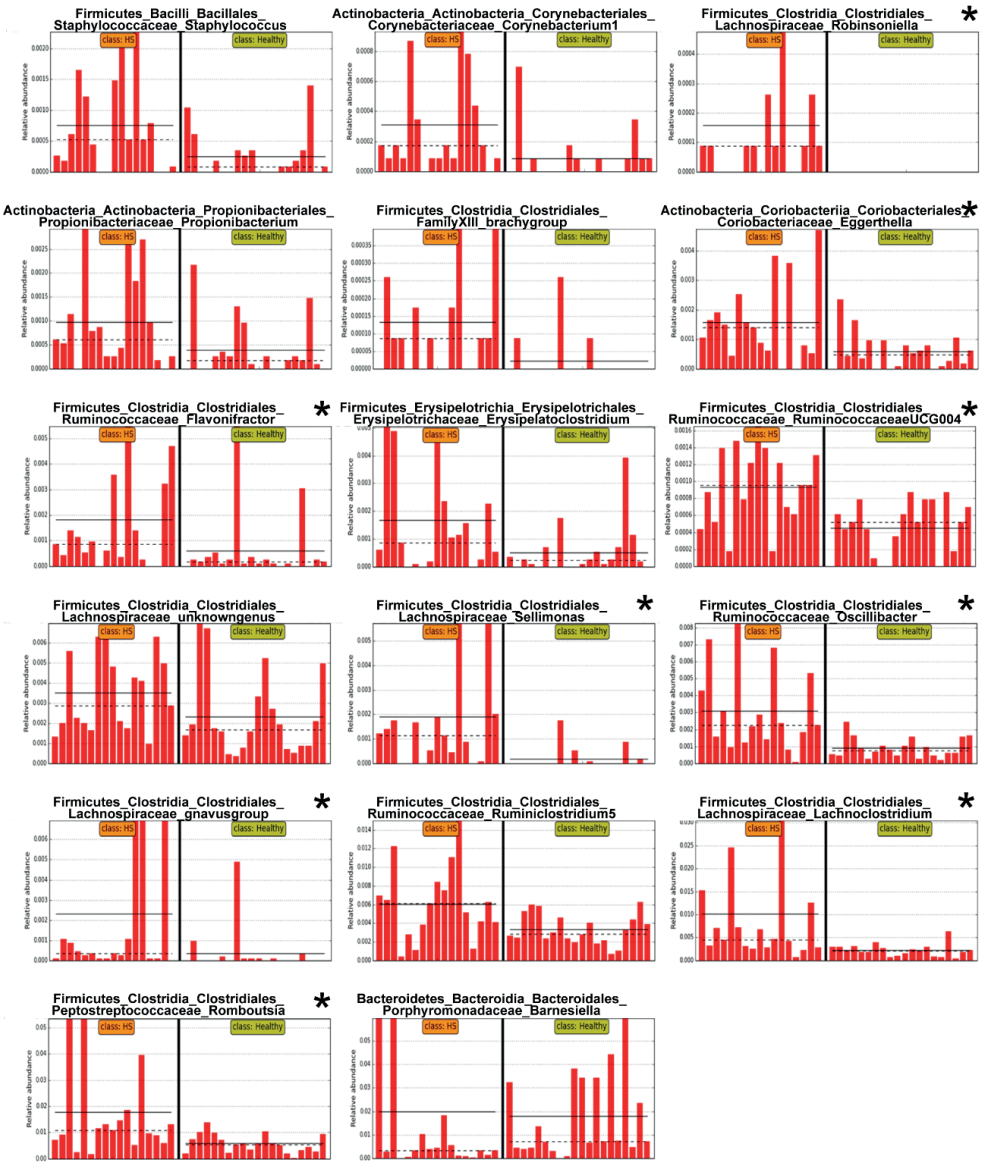
[a] HS subjects are annotated as 1-17 and healthy subjects as H1-20.
[b] Age at diagnosis HS, disease duration and disease severity scores based on Hurley clinical staging system (I, II, III)
[c] Axillary skin swab collection: no or yes (HS lesion only or both HS lesion and HS non-lesion).
Abbreviations: BMI, body mass index; n, number; y, years.



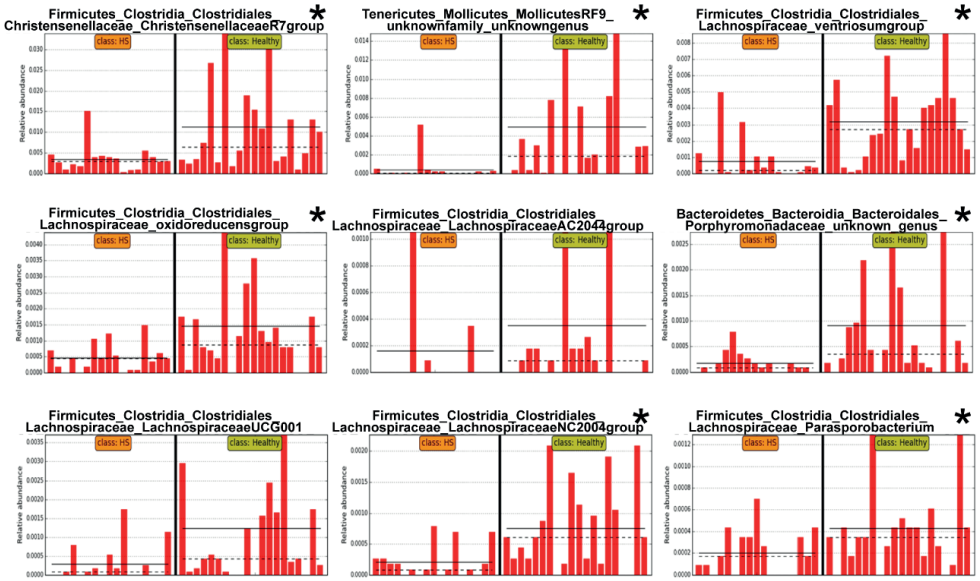
▲Supplementary Figure S1 | Characterization of the skin and fecal microbiome from hidradenitis suppurativa (HS) patients and healthy controls. 16S rRNA amplicon sequencing results comparing fecal (n=37) and skin (n=16) samples that were collected from HS patients and healthy controls. **(a)** Bar plots showing the relative abundance of individual fecal and skin samples at phylum and genus level. Healthy subjects are annotated as H1-20 and HS subjects as 1-17. The asterisks (*) mark the non-lesional samples of HS patients. **(b)** Boxplots showing the bacterial biodiversity of fecal and skin samples measured by richness, Shannon index and the inverse Simpson index at genus level. Each plot represents the 25-75% interquartile range and the median. **(c)** Multidimensional scaling (MDS) plots illustrate the bacterial composition of fecal and skin samples using Bray-Curtis and Jaccard metrics at genus level.



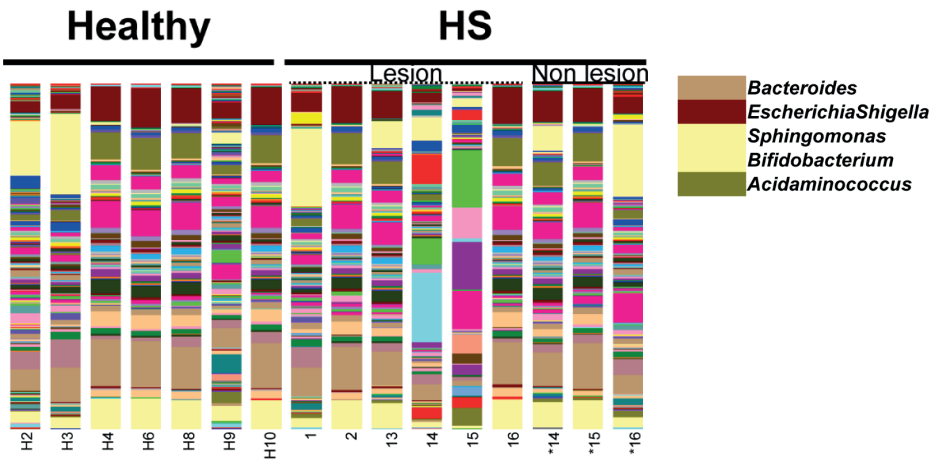
▲Supplementary Figure S2 | Comparison of microbial composition of fecal samples from hidradenitis suppurativa (HS) patients stratified for smoking status and body mass index (BMI). 16S rRNA amplicon sequencing results comparing fecal samples from HS patients (n=17) based on smoking status (smoking [n=11] versus non-smoking [n=6]) and BMI (≤25 [n=4] versus >25 [n=13]). **(a)** Boxplots showing the fecal bacterial biodiversity measured by richness, Shannon index and the inverse Simpson index at genus level. Each plot represents the 25-75% interquartile range and the median. **(b)** Multidimensional scaling (MDS) plots illustrate the fecal bacterial composition of smokers versus non-smokers (upper panels) and BMI≤25 versus BMI>25 (lower panels) using Bray-Curtis and Jaccard metrics at genus level.



▲Supplementary Figure S3 | Linear discriminant analysis (LDA) effect size (LEfSe) analysis results of taxa enriched in feces of hidradenitis suppurativa (HS) patients. The histogram depicts the relative abundance of 17 taxa that were differentially more abundant in fecal samples of HS patients (n=17) in comparison to healthy controls (n=20) at genus level. The horizontal solid line represents the group mean and the dashed line the group median. The asterisks (*) mark the taxa that are differently abundant among the two groups irrespective of smoking habits and body mass index (BMI).



▲Supplementary Figure S4 | Linear discriminant analysis (LDA) effect size (LEfSe) analysis results of taxa depleted in feces of hidradenitis suppurativa (HS) patients. The histogram depicts the relative abundance of 9 taxa that were depleted in fecal samples of HS patients (n=17) in comparison to healthy controls (n=20) at genus level. The horizontal solid line represents the group mean and the dashed line the group median. The asterisks (*) mark the taxa that are differently abundant among the two groups irrespective of smoking habits and body mass index (BMI).



▲Supplementary Figure S5 | The 16S rRNA amplicon sequencing results of skin samples from hidradenitis suppurativa (HS) patients (n=6) and healthy controls (n=7). Bar plots showing the relative abundance of individual skin samples at genus level. Healthy subjects are annotated as H2, 3, 4, 6, 8, 9 or 10, and HS subjects as 1, 2, 13, 14, 15 or 16. The asterisks (*) mark the non-lesional samples of HS patients.

9

CHAPTER 9

Paneth cell dysfunction in the ileocecal resection specimen as predictor of re-resection in Crohn's disease

E.M.J. Beelen[1], S.Y. Lam[1], G.M. Fuhler[1], W.R. Schouten[2], M.P. Peppelenbosch[1], C.J. van der Woude[1], A.C. de Vries[1]

[1] Department of Gastroenterology and Hepatology, Erasmus MC University Medical Center Rotterdam, the Netherlands; [2] Department of Surgery, Erasmus MC University Medical Center Rotterdam, the Netherlands.

In preparation

Abstract

Background: Paneth cells are essential for gut homeostasis and involved in Crohn's disease (CD) pathology. Markers of Paneth cell dysfunction may predict an unfavourable postoperative CD course. This study aims to explore the predictive value of markers of Paneth cell dysfunction for the risk of re-resection in CD patients.

Methods: In this single center case-control study, adult CD patients with intestinal re-resection (n=25) and without re-resection (n=25) during at least 10 years of follow-up were included. The proximal resection margin was processed for genotyping of Paneth cell-associated IBD risk loci, assessment of ER stress, quantification of Paneth cell-specific lysozyme, the total bacterial load and presence of *Faecalibacterium prausnitzii* (*F. prausnitzii*) and adherent-invasive *Escherichia coli* (*AIEC*).

Results: The majority (n=39) of patients carried either one or two *ATG16L1* risk alleles. The median number of GRP78 positive crypts was 8 (24%) in patients with re-resection and 14 (20%) in patients without re-resection (*P*=0.799). No difference was found in mean intensity, distribution or quantity of lysozyme staining. Total bacterial load and levels of *AIEC* were similar in both groups. Levels of *F. prausnitzii* appeared to be lower in patients with re-resection during follow-up, although not statistically significant.

Conclusion: In this explorative study, ER stress levels, lysozyme expression and genetic polymorphisms affecting Paneth cell function as well as bacterial markers *AIEC* or *F. prausnitzii* in resection specimens were not associated with re-resection. Thus, this study was unable to identify markers of Paneth cell function in resection specimens as predictors for postoperative surgical recurrence.

1 Introduction

Crohn's disease (CD) is a multifactorial disease in which genetic susceptibility contributes to altered immunological responses associated with a dysbiotic intestinal microbiome, resulting in chronic intestinal inflammation. Despite considerable advances in inflammatory bowel disease (IBD) management, with new diagnostic, treatment and monitoring strategies, up to 50% of patients will require intestinal resection during the disease course [1,2]. The prediction of the postoperative disease course in CD patients is challenging. A wide range of clinical and surgical predictors have been studied, often with contradicting results. Smoking, penetrating disease behaviour and a history of prior intestinal resection have been previously identified as individual predictors of postoperative recurrence [3]. However, even in the absence of such clinical risk factors, postoperative endoscopic recurrence rates are estimated to be as high as 50% within 18 months [4, 5]. Overall, approximately 30% of patients will require a re-resection of affected regions [6].

As prediction of the postoperative CD course with clinical markers thus far appears to be unsatisfactory, recent studies have also focused on histologic features in the ileocolonic resection specimen for the prediction of disease recurrence [7]. Clonal T cell expansions in the ileal mucosa are suggested to be associated with smoking and disease recurrence [8]. Furthermore, the presence of granulomas, (sub)mucosal lymphatic vessel density and plexitis have previously been associated with recurrence or reoperation. While in particular myenteric plexitis shows promise for the prediction of postoperative disease course [9,10], other studies were unable to validate these results and additional markers are still needed [11,12].

One such histologic element with potentially predictive potential is the Paneth cell, a specialized secretory cell in the small bowel, at the bases of the crypts of Lieberkühn. Paneth cells control microbial invasion in the intestine and help to protect the barrier function by secreting antimicrobial proteins, including α-defensins and lysozyme. Several lines of evidence have suggested a role for these cells in CD pathology. Mouse models have demonstrated that loss of Paneth cell function is associated with microbiome-dependent susceptibility to intestinal inflammation [13,14]. In human CD patients, reduced levels of α-defensins and morphological Paneth cell defects are seen, which are associated with microbial dysbiosis [15,16]. Genetic susceptibility appears to play a role in the contribution of Paneth cells to disease pathology as morphological Paneth cell aberrations were found to be more pronounced in patients carrying a single nucleotide polymorphism (SNP) in the autophagy gene *ATG16L1* [17]. *ATG16L1* plays a physiological role in protein turnover, and loss of this gene in mice results in the accumulation of endoplasmic reticulum (ER) stress

sensors [18]. Consistently, sustained ER stress levels as well as mucosal bacterial changes are observed in Paneth cells from patients carrying the *ATG16L1* risk allele [19].

Considering their role in gut homeostasis and CD pathology, markers of Paneth cell dysfunction might be considered as predictors for an unfavourable postoperative CD disease course. Indeed, a retrospective analysis of 102 ileocolonic CD resection specimens indicated a shorter time to postoperative disease recurrence in patients with abnormal Paneth cell granule morphology [20]. However, confirmation of these data is lacking. This study aims to explore the predictive value of markers of Paneth cell dysfunction for the risk of re-resection in CD patients, as a robust marker of CD prognosis.

2 Methods

2.1 Study design

In this explorative single center study, patients who underwent ileocecal resection for the indication of CD were identified from the hospital endoscopy registry system, Endobase (Olympus Medical Systems Europe, Hamburg, Germany). Patients ≥16 years old who were in surgical remission after ileocecal resection, and had at least 10 years of follow-up after resection were included. Demographics, disease characteristics and surgical characteristics were extracted from electronic patient files. The study population was divided into 2 groups, consisting of patients with and without re-resection during follow-up, indicating an unfavourable and favourable postoperative disease course, respectively.

For this study, formalin fixed paraffin embedded (FFPE) tissue resection specimens were obtained from the department of Pathology at the Erasmus MC University Medical Center Rotterdam. A total of 50 FFPE samples were available from CD patients (25 with re-resection and 25 without re-resection during the follow up). Serial 5µm and 10µm FFPE sections from the non-inflamed proximal resection margin of the ileocolonic resection specimen were prepared using a microtome.

This study was conducted in accordance with the protocol and the principles of the Declaration of Helsinki. Informed consent was not obtained considering the retrospective and anonymised character of the dataset. The protocol was approved by the Medical Ethical Review Committee of the Erasmus MC University Medical Center on the 29th of March 2017.

2.2 DNA isolation

DNA was isolated from five 10µm FFPE tissue sections using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. DNA concentration was measured using Nanodrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). All samples were stored at -20˚C until further downstream analyses.

2.3 ER-stress and lysozyme quantification

FFPE tissue sections (5µm) were stained with antibodies against GRP78, a marker of ER stress [18]. In short, slides were deparaffinised, endogenous peroxidases were blocked with 3% H_2O_2, and blocked with normal goat serum. Staining was performed with anti-BiP antibody (GRP78 1:400, Cell Signaling Technology, Leiden, the Netherlands) or anti-lysozyme (1:800 DAKO, Glostrup, Denmark) overnight at 4˚C followed by DAKO Envision + system-HRP (DAKO, Glostrup, Denmark) at room temperature for 1h, after which antibody binding was detected by 3-amino-9-ethylcarbazol (DAB, Sigma-Aldrich, St Louis, MO). Staining intensity was scored, blinded, using images obtained by 40x light microscopy (Zeiss, Axioskop, Oberkochen, Germany). GRP78 staining was scored based on the numbers of GRP78-positive Paneth cells in the crypts. Furthermore, intensity of GRP78 staining was scored on a scale from 0-2, 2 being the highest intensity. Representative examples of ER-stress positive and negative samples are provided in **Supplementary Figure S1**.

Previously, Paneth cell morphology as assessed by lysozyme staining of sections has been associated with Paneth cell defects and IBD [20]. Therefore, we chose to determine the immune reactivity intensity of lysozyme as a specific marker for Paneth cells. At least 100 well defined crypts per slide were assessed. A minimum of 100 Paneth cells per section were scored to quantify the overall intensity for lysozyme staining using the ImageJ software with the immunohistochemistry toolbox plugin [21]. Intensity of the staining is described as a number from 0-255 where 255 represents the lowest intensity (**Supplementary Figure S2**).

2.4 Genotyping of single nucleotide polymorphisms (SNPs)

Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) analysis was conducted for SNP *ATG16L1* (rs2241880) using forward primer 5'-GCTCTGTCACCATATCAAGCG-3' and reverse primer: 5'-AGGAGACGCTCTGCT CTTC-3' [19]. The reaction mixture contained colorless GoTaq® buffer (Promega, Maddison, WI), 1.25mM $MgCl_2$ (Promega), 0.167mM (each) deoxynucleotides (Roche Diagnostics,

Mannheim, Germany), 2.5U GoTaq®polymerase (Promega), 333nM of each primer (Sigma-Aldrich, St Louis, MO), 4µl DNA and water for a final volume of 30µl. Amplification was performed using the Applied Biosystems 2720 Thermal Cycler (Applied Biosystems, Waltham, MA) under the following conditions: initial denaturation of 7 minutes at 95˚C, 40 cycles consisting of 30 seconds denaturation at 95˚C, 30 seconds of annealing at 55˚C and 1 min of elongation at 72˚C, and final extension for 10 minutes at 72˚C. The amplification products were subjected to gel electrophoresis using 2% agarose gel in 1x TBE buffer containing Serva DNA stain G (Promega), and after confirmation used in a second reaction mixture together with 2µl 10x Buffer Bful (Thermofisher), 1µl restriction endonuclease Bful (Thermofisher) and 18µl water for a total volume of 31µl. Following overnight incubation of 10 hours at 37˚C, the enzyme was inactivated after 20 minutes at 80˚C. The digestion profiles were then viewed on gel to determine *ATG16L1* genotype status (GG, GA or AA) (**Supplementary Figure S3**).

In a subset of patients, additional genotyping was performed at KBioscience Ltd (Hertfordshire, UK) for CD-associated SNPs in XBP1 (rs35873774), NOD2 (rs2066844, rs2066845, rs2066847), IRGM (rs13361189), STAT3 (rs744166), NCF4 (rs4821544), CCR6 (rs2301436) and IL23 (rs11465804) [22,23].

2.5 Quantitative polymerase chain reaction (qPCR) for microbial analysis

To investigate the role of bacterial microbes in the disease course after ileocecal resection, the commensal *Faecalibacterium prausnitzii (F. prausnitzii)* and pathogenic adherent-invasive *Escherichia coli* (*AIEC*) were selected for qPCR analyses using the StepOnePlus Real-Time PCR system (Applied Biosystems). To control for the size of the resection specimen from which DNA was isolated, human genomic DNA present was quantified by measuring *ACTB* gene levels. Each qPCR reaction contained 10µl of SYBR® Select Master Mix (Applied Biosystems), 1µl of each forward and reverse primer (end concentration 0.5µM), 6µl of water and 2µl of non-normalized template for a total volume of 20µl. Each reaction mixture was subsequently subjected to the following PCR conditions: 4 minutes at 95˚C and 40 cycles of 15 seconds denaturation at 95˚C, 30 seconds annealing and 30 seconds elongation at 72˚C. The annealing temperatures for *F. prausnitzii* (forward: 5'- GATGGCCTCGCGTCCGATTAG-3', reverse: 5'-CCGAAGACCTTCTTCCTCC-3) [24-27], *AIEC* (forward: 5'-CCATTCATG CAGCAGCTCTTT-3', reverse: 5'- ATCGGACAACATTAGCGGTGT-3') [28,29] and *ACTB* gene (forward 5'-CTGGAACGGTGAAGGTGACA-3', reverse: 5'-AAGGGACTTCCTGTAACAATGCA-3') [30] primer sets (Sigma-Aldrich, St Louis, MO) were 58˚C, 60˚C and 60.5˚C respectively.

2.6 Data analysis

Statistical analysis of clinical data was performed using IBM SPSS for windows, version 24.0 (IBM Corp., Armonk, NY). Non-normally distributed variables are displayed as medians with range or interquartile range (IQR). Outcome data were compared between patient groups with and without re-resection using Mann-Whitney U test, Chi-square test or Fisher's exact test.

For the qPCR data analysis of bacterial markers, the $2^{-\Delta\Delta CT}$ method was used for calculation of the fold change. First, ΔCT_{sample} (= $CT_{bacterial\ target}$ – $CT_{ACTB\ target}$) was obtained for each sample by normalization to the amount of total human DNA. The average CT_{sample} of the group without re-resection (control group) was then used to calculate ΔΔCT (= ΔCT_{sample} – average $\Delta CT_{control\ group}$), after which the fold change derived from $2^{-\Delta\Delta CT}$. Differences between groups were measured using the Mann-Whitney U test. A p-value <0.05 was considered statistically significant.

3 Results

3.1 Study population

The study population comprised 50 patients (male/female 19/31; median age 26.6 years [IQR 21.1 – 35.8]), who underwent ileocecal resection between 1987 and 2007. Median follow-up time after resection was 17.7 years (IQR 14.4 – 21.41). During follow-up, 25 patients underwent a re-resection after a median of 8.0 years (IQR 3.8 – 14.6).

Ileocecal resection was performed for the indication of isolated ileal disease in 32 (64%) patients, whereas 18 (36%) patients also had colonic (limited to cecum or ascending colon) involvement. Half of patients (n=25) had stricturing disease, and 15 (30%) patients had penetrating disease. Twelve patients (24%) were active smokers at baseline. Ileocecal resection was performed early after CD diagnosis in a majority of patients, median disease duration was 1.9 years (IQR 0.4 – 5.7) and previous thiopurine and anti-TNFα use were 7 (14%) and 5 (10%), respectively. Patients in the group without re-resection during follow-up were significantly younger at the time of ileocecal resection (22.2 years [IQR 19.0 – 35.7]) as compared to patients with re-resection (29.0 years [IQR 24.5 – 37.0]), *P*=0.03. No differences in disease behaviour, disease localisation or smoking status were observed between the re-resection and no re-resection group (**Table 1**).

3.2 Lysozyme quantification

Staining for lysozyme was performed in 31 sections, of which 9 sections did not stain sufficiently. Therefore, 22 sections were included in the analysis, of which 10 (46%) with

Table 1 | Baseline clinical characteristics of study subjects

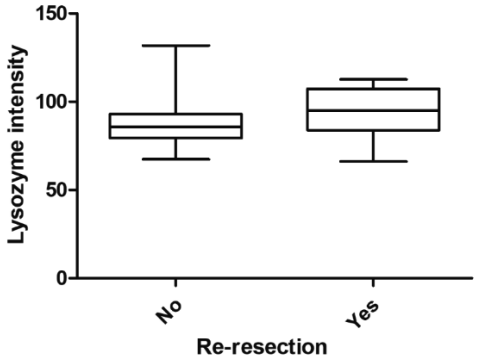
	No re-resection (*n*=25)	Re-resection (*n*=25)	*P*-value
Male gender, *n* (%)	11 (44)	8 (32)	0.38
Age (y), median (IQR)	22.2 (19.0 – 35.7)	29.0 (24.5 – 37.0)	0.03
Disease duration (y), median (IQR)	1.1 (0.1 – 6.6)	2.2 (0.9 – 5.8)	0.35
Montreal A, *n* (%)			
A1 < 17 y	6 (24)	2 (8)	0.08
A2 17-40 y	11 (44)	20 (80)	
A3 > 40 y	4 (16)	2 (8)	
Missing	4 (16)	1 (4)	
Montreal L, *n* (%)			
L1 Ileal	17 (68)	15 (60)	0.56
L3 Ileocolonic	8 (32)	10 (40)	
Montreal B, *n* (%)			
B1 Luminal	4 (16)	6 (24)	
B2 Stricturing	12 (48)	13 (52)	0.60
B3 Penetrating	9 (36)	6 (24)	
Perianal disease, *n* (%)	4 (16)	5 (20)	1.00
Active smoker, *n* (%)	5 (20)	7 (28)	0.16
Missing	6 (24)	11 (44)	
Medication prior to resection			
Thiopurine, *n (%)*	5 (20)	2 (8)	0.42
Anti-TNFα, *n* (%)	5 (20)	0 (0)	0.05

Abbreviations: anti-TNF, anti tumor necrosis factor; IQR, interquartile range; n, number; y, years.

and 12 (54%) without re-resection during follow-up. Per section, a median of 174.5 Paneth cells (IQR 157.8 – 202.0) were scored. A median of 7.8 (IQR 4.5 – 11.5) lysozyme positive Paneth cells per crypt were detected, 7.0 (IQR 6.1 – 8.5) in patients without re-resection, 9.5 (IQR 7.2 – 10.2) in patients with re-resection, *P*=0.08. Median of the measured lysozyme staining intensity was 85.7 (IQR 79.4 – 93.0) for the no re-resection group as compared to 95.0 (IQR 83.8 – 107.3) for the re-resection group, *P*=0.254 (**Figure 1**). No differences in the number or distribution of lysozyme positive granules in the Paneth cells was noted. Thus, these results indicate that based on lysozyme staining no numerical or morphological aberrations are present in patients requiring a re-resection.

3.3 Genotyping of risk genes in Crohn's disease (CD)

Multiple risk alleles have been identified for CD, some of which may also affect postoperative complications [31, 32]. In light of its importance for Paneth cell function, we determined the

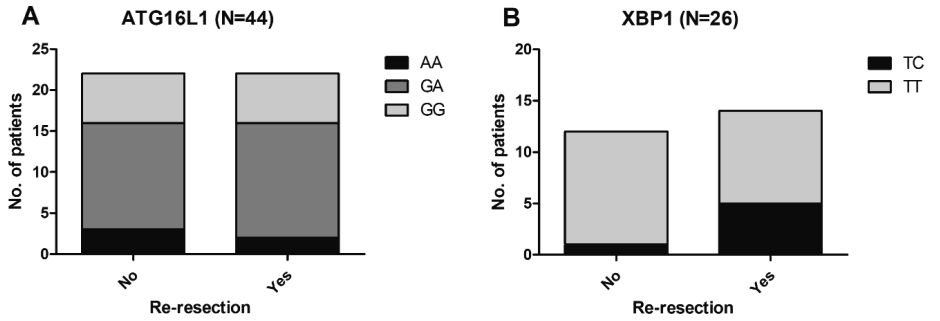


▲**Figure 1 |** Median and range of lysozyme staining intensity in patients with and without re-resection during follow-up

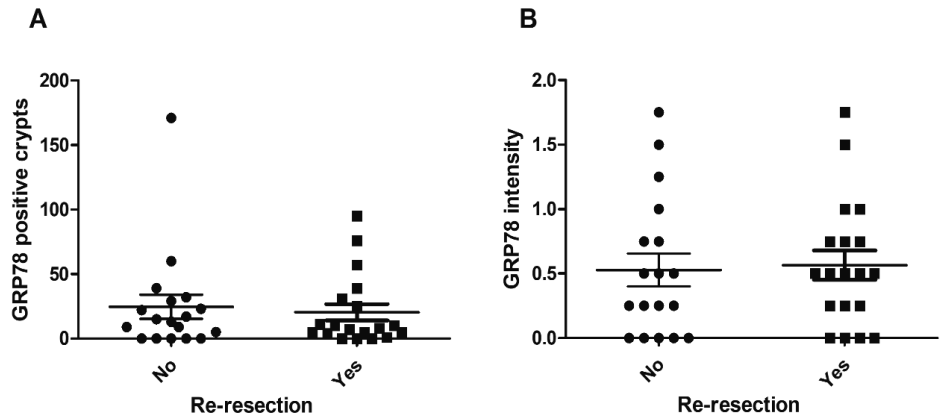
ATG16L1 rs2241880 SNP status in a total of 44 CD patients. The majority (n=39) of patients carried either one or two *ATG16L1* risk alleles (GA or GG), but no differences were found between the patient groups with and without re-resection (**Figure 2A**). For a total of 26 patients, the SNP status was also performed for other CD risk genes associated with innate immune function [22]. Interestingly, more patients in the re-resection group (n=5; 19%) carried the protective C allele of XBP1, an ER stress regulator known to affect Paneth cell function [13], compared to patients without re-resection (n=1; 4%), but this did not reach statistical significance (*P*=0.09) (**Figure 2B**). No differences were found between the re-resection and no re-resection group for NOD2, STAT3, NCF4, CCR5 and IL23 risk alleles.

3.4 Paneth cell ER-stress

As ER stress appears to be intrinsically linked to Paneth cell function in CD patients, we next investigated the levels of GRP78, a hallmark of ER stress, in resection specimens. FFPE sections of 37 patients, 18 with re-resection during follow-up and 19 without, were successfully analyzed for ER stress by GRP78 staining. A median of 66.0 (range 5 - 245) crypts were scored per section, and GRP78 staining intensity was scored. Overall, the median percentage of GRP78 positive crypts per section was 20.0 (IQR 3.6 – 41.0). The median number of GRP78 positive crypts did not differ between the re-resection (8 [IQR 4-31], median 24% of crypts per patient) and no re-resection group (14 [IQR 0-30], median 20% of crypts per patient), *P*=0.799 (**Figure 3A**). In addition, the intensity of the GRP78 staining did not differ between both groups. Median intensity was 0.5 (IQR 0.25 – 0.75) in the re-resection group versus 0.4 (IQR 0.00 – 0.81) in the group without re-resection, *P*=0.904 (**Figure 3B**).



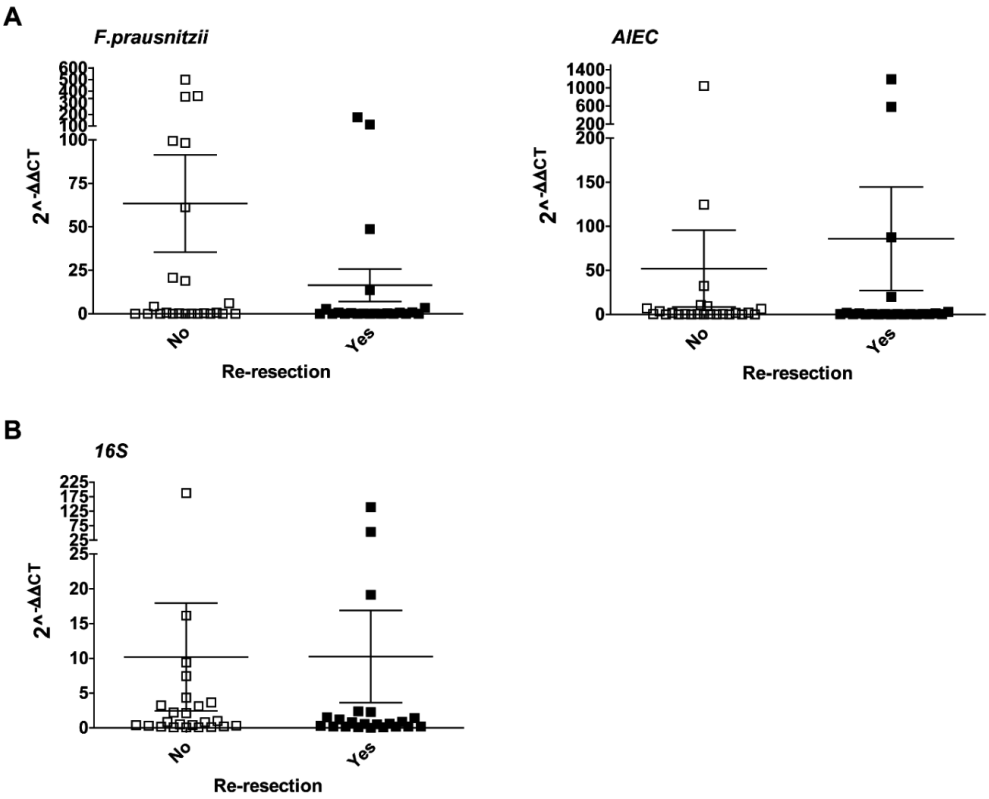
▲**Figure 2 |** Number of patients carrying one or two (GA or GG) of the risk alleles for *ATG16L1* **(A)** and the number of patients carrying the protective C allele for XBP1 **(B)** are presented for the groups with and without re-resection during follow-up.



▲**Figure 3 |** The number of GRP78 positive crypts **(A)** and the GRP78 intensity from 0-2 **(B)** in patients with and without re-resection.

3.5 Microbial analysis

The fold change levels of *F. prausnitzii* appeared to be lower in patients with re-resection in comparison to those without surgery during follow-up, although not statistically significant (*P*=0.72) (**Figure 4A**). In addition, no differences were detected between the groups for AIEC (*P*=0.77) or total bacterial load as determined by 16S (**Figure 4B**). Although only minor bacterial changes were found here, these findings might be a part of larger shifts at community level which warrants further exploration.



▲Figure 4 | Quantitative analysis of bacterial markers in Crohn's disease patients with and without re-resection during follow-up. The relative amount of each sample is normalized to human genomic DNA and is expressed in terms of fold change ($2^{\wedge-\Delta\Delta CT}$) for bacterial markers *Faecalibacterium prausnitzii* and the adherence-invasive *Escherichia coli (AIEC)* **(A)**, in addition to the total bacterial load which is expressed by 16S **(B)**.

4 Discussion

In this study, we investigated the potential of Paneth cell defects to act as a marker of surgical recurrence after intestinal resection for CD. We found that neither Paneth cell numbers, lysozyme expression levels nor distribution were associated with requirement of re-resection for CD. Genetic polymorphisms affecting Paneth cell function, ER stress levels and levels of *F. prausnitzii* or *AIEC* were also not predictive of re-resection. Thus, this study was unable to identify Paneth cell aberrations as either cause or marker for recurrent surgery.

ER-stress-induced Paneth cell dysfunction exacerbates disease in mice [14], potentially via modulation of the intestinal microbiota. Upon stimulation with Interferon gamma, Paneth cells release various anti-microbial compounds [33], which alter the intestinal microbial

composition. Microbial dysbalance is commonly observed in IBD patients, and in CD patients previous research showed an association with Paneth cell defects [16]. Here, we selected two bacterial features previously associated with Paneth cell alterations [16, 19, 34]. *F. prausnitzii* is a commensal butyrate producing, anti-inflammatory bacterium, which is reported to be decreased in CD patients. Interestingly, a higher risk of postoperative recurrence was observed in CD patients with low mucosal *F. prausnitzii* levels [35]. In our cohort, a reduced presence of this bacterium was also seen, although this did not reach statistical significance. Secondly, we investigated *AIEC*, a mucosal pathogen with an increased prevalence in ileal mucosa in CD patients [36]. The presence of this bacterium has previously been shown to be associated with postoperative endoscopic recurrence of disease [37, 38]. However, we were unable to replicate these findings. This may be due to the lower numbers in our study, or because our study used re-resection rather than endoscopic lesions as measure for recurrence.

In contrast to previous data, in this study genetic polymorphisms affecting Paneth cell function were not associated with re-resection. Previous data have suggested that IBD genetic risk factors are associated with altered Paneth cell function as well as microbial composition. In particular SNPs in the bacterial sensor *NOD2*, the ER stress regulator *XBP1*, and the autophagy genes *IRGM* and *ATG16L1*, are associated with microbial composition in ileal CD patients [39-41]. Genetic contribution to recurrence of disease after IBD surgery has also been investigated. Recently, a SNP in NOD2 was linked to a shorter time to recurrent ileocolectomy in CD [42]. However, another recent study showed that SNPs in *SMAD3* but not *NOD2*, *IRGM* or *ATG16L1* were associated with surgical recurrence. A larger study investigating 200 genetic IBD risk variants only found *CARD8* variants to be a risk factor for recurrent surgery [44]. It is conceivable that genetic risk factors should not be investigated individually, but as a whole, as the number of risk alleles is associated with both increased microbial alterations and odds ratios for individual risk alleles is low [45, 46]. Interestingly, Gathungu *et al.* showed GM-CSF cytokine and antibody levels at time of first surgery, but not *ATG16L1* status, were predictive of time to second surgery. However, when investigating the interaction between these markers, they observed that patients with the protective *ATG16L1* allele as well as low GM-CSF levels had significantly longer intervals to the second surgery [47]. GM-CSF signaling in general appears important in IBD, and several genetic polymorphisms are associated with impaired GM-CSF-mediated innate immune responses [23, 48]. These data would suggest that multiple genetic risk variants associated with IBD combine to affect severity of disease. Indeed, molecular classification of IBD disease groups on genetic rather than clinical parameters has been

proposed [49]. IBD is a multifactorial disease, in which immunity, microbiome and genetics affect each other reciprocally and cause and effect are not easily distinguished. Genetic risk alleles affect multiple cell types, and while we investigated the role of Paneth cells here, monocytes from patients with the *ATG16L1* risk allele also show impaired killing of AIEC [50]. Thus, investigation of multiple cell types, microbes and their interactions may yield better models to predict recurrence of disease after CD surgery.

One of the strengths of this study is the broad, extensive approach in analysis of markers for Paneth cell function. However, some limitations of this study need to be considered. The small population size might have contributed to the inability of reproducing Paneth cell markers as predictors for CD recurrence. In addition, in this study, in order to reliably assess Paneth cell lysozymes and ER stress, we collected ileal tissue free of inflammation. Therefore, other histologic factors that might be of influence e.g. (transmural) inflammation or granulomas were not assessed [12]. Submucosal or myenteric plexitis were not taken into account because current available studies show contradicting results [10, 12, 50] and considering the small, explorative character of this study, we chose to focus on thoroughly assessing one aspect of the resection specimen. Furthermore, the choice of re-resection as endpoint rather than endoscopic recurrence could be of influence. However, we chose the robust endpoint re-resection and long-term follow-up of more than 10 years, to form two distinct study groups in terms of postoperative prognosis. Furthermore, in a retrospective analysis like the present study, endoscopies might have been performed on different time points and for different indications, thereby excluding endoscopy as a reliable outcome to compare between study groups.

In this study, Paneth cell morphology in the resection specimen and factors known to modulate or be modulated by these cells were not associated with the risk of intestinal re-resection after ileocecal resection in CD. Therefore, markers of Paneth cell function in the resection specimen are currently not validated to predict postoperative recurrence after ileocecal resection in CD.

References

1. Bernell O, Lapidus A and Hellers G. Risk factors for surgery and postoperative recurrence in Crohn's disease. Ann Surg, 2000. 231(1): p. 38-45.
2. Frolkis AD, Dykeman J, Negrón ME, Debruyn J, Jette N, Fiest KM *et al*. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. Gastroenterology, 2013. 145(5): p. 996-1006.
3. De Cruz P, Kamm MA, Prideaux L, Allen PB and Desmond PV. Postoperative recurrent luminal Crohn's disease: a systematic review. Inflamm Bowel Dis, 2012. 18(4): p. 758-77.
4. Gionchetti P, Dignass A, Danese S, Magro Dias FJ, Rogler G, Lakatos PL *et al*. 3rd European Evidence-based Consensus on the Diagnosis and Management of Crohn's Disease 2016: Part 2: Surgical Management and Special Situations. J Crohns Colitis, 2017. 11(2): p. 135-149.
5. De Cruz P, Kamm MA, Hamilton AL, Ritchie KJ, Krejany EO, Gorelik A *et al*. Crohn's disease management after intestinal resection: a randomised trial. Lancet, 2015. 385(9976): p. 1406-17.
6. Beelen EMJ, van der Woude CJ, Pierik MJ, Hoentjen F, de Boer NK, Oldenburg B *et al*. Decreasing Trends in Intestinal Resection and Re-Resection in Crohn's Disease: A Nationwide Cohort Study. Ann Surg, 2019.
7. Bressenot A and Peyrin-Biroulet L. Histologic features predicting postoperative Crohn's disease recurrence. Inflamm Bowel Dis, 2015. 21(2): p. 468-75.
8. Allez M, Auzolle C, Ngollo M, Bottois H, Chardiny V, Corraliza AM *et al*. T cell clonal expansions in ileal Crohn's disease are associated with smoking behaviour and postoperative recurrence. Gut, 2019. 68(11): p. 1961-1970.
9. Lemmens B, de Buck van Overstraeten A, Arijs I, Sagaert X, Van Assche G, Vermeire S *et al*. Submucosal Plexitis as a Predictive Factor for Postoperative Endoscopic Recurrence in Patients with Crohn's Disease Undergoing a Resection with Ileocolonic Anastomosis: Results from a Prospective Single-centre Study. J Crohns Colitis, 2017. 11(2): p. 212-220.
10. Decousus S, Boucher AL, Joubert J, Pereira B, Dubois A, Goutorbe F *et al*. Myenteric plexitis is a risk factor for endoscopic and clinical postoperative recurrence after ileocolonic resection in Crohn's disease. Dig Liver Dis, 2016. 48(7): p. 753-8.
11. Ng SC, Lied GA, Kamm MA, Sandhu F, Guenther T, and Arebi N. Predictive value and clinical significance of myenteric plexitis in Crohn's disease. Inflamm Bowel Dis, 2009. 15(10): p. 1499-507.
12. Hammoudi N, Cazals-Hatem D, Auzolle C, Gardair C, Ngollo M, Bottois H *et al*. Association Between Microscopic Lesions at Ileal Resection Margin and Recurrence After Surgery in Patients With Crohn's Disease. Clin Gastroenterol Hepatol, 2020. 18(1): p. 141-149 e2.
13. Kaser A, Lee AH, Franke A, Glickman JN, Zeissig S, Tilg H *et al*. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell, 2008. 134(5): p. 743-56.
14. Adolph TE, Tomczak MF, Niederreiter L, Ko HJ, Böck J, Martinez-Naves E *et al*. Paneth cells as a site of origin for intestinal inflammation. Nature, 2013. 503(7475): p. 272-6.
15. Perminow G, Beisner J, Koslowski M, Lyckander LG, Stange E, Vatn MH *et al*. Defective paneth cell-mediated host defense in pediatric ileal Crohn's disease. Am J Gastroenterol, 2010. 105(2): p. 452-9.
16. Liu TC, Gurram B, Baldridge MT, Head R, Lam V, Luo C *et al*. Paneth cell defects in Crohn's disease patients promote dysbiosis. JCI Insight, 2016. 1(8): p. e86907.
17. Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, Lennerz JK *et al*. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature, 2008. 456(7219): p. 259-63.

18. Tschurtschenthaler M, Adolph TE, Ashcroft JW, Niederreiter L, Bharti R, Saveljeva S *et al*. Defective ATG16L1-mediated removal of IRE1α drives Crohn's disease-like ileitis. J Exp Med, 2017. 214(2): p. 401-422.
19. Deuring JJ, Fuhler GM, Konstantinov SR, Peppelenbosch MP, Kuipers EJ, de Haar C *et al*. Genomic ATG16L1 risk allele-restricted Paneth cell ER stress in quiescent Crohn's disease. Gut, 2014. 63(7): p. 1081-91.
20. VanDussen KL, Liu TC, Li D, Towfic F, Modiano N, Winter R *et al*. Genetic variants synthesize to produce paneth cell phenotypes that define subtypes of Crohn's disease. Gastroenterology, 2014. 146(1): p. 200-9.
21. Schneider CA, Rasband WS and Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods, 2012. 9(7): p. 671-5.
22. Nuij V, Peppelenbosch MP, van der Woude CJ and Fuhler GM. Genetic polymorphism in ATG16L1 gene is associated with adalimumab use in inflammatory bowel disease. J Transl Med, 2017. 15(1): p. 248.
23. Somasundaram R, Deuring JJ, van der Woude CJ, Peppelenbosch MP and Fuhler GM. Linking risk conferring mutations in NCF4 to functional consequences in Crohn's disease. Gut, 2012. 61(7): p. 1097; author reply 1097-8.
24. Zhang M, Zhou L, Wang Y, Dorfman RG, Tang D, Xu L *et al*. Faecalibacterium prausnitzii produces butyrate to decrease c-Myc-related metabolism and Th17 differentiation by inhibiting histone deacetylase 3. Int Immunol, 2019. 31(8): p. 499-514.
25. Aguirre de Cárcer D, Cuív PO, Wang T, Kang S, Worthley D, Whitehall V *et al*. Numerical ecology validates a biogeographical distribution and gender-based effect on mucosa-associated bacteria along the human colon. Isme J, 2011. 5(5): p. 801-9.
26. Ahmed S, Macfarlane GT, Fite A, McBain AJ, Gilbert P, and Macfarlane S. Mucosa-associated bacterial diversity in relation to human terminal ileum and colonic biopsy samples. Appl Environ Microbiol, 2007. 73(22): p. 7435-42.
27. Wang RF, Cao WW and Cerniglia CE. PCR detection and quantitation of predominant anaerobic bacteria in human and animal fecal samples. Appl Environ Microbiol, 1996. 62(4): p. 1242-7.
28. Galtier M, De Sordi L, Sivignon A, de Vallée A, Maura D, Neut C *et al*. Bacteriophages Targeting Adherent Invasive Escherichia coli Strains as a Promising New Treatment for Crohn's Disease. J Crohns Colitis, 2017. 11(7): p. 840-847.
29. Bretin A, Lucas C, Larabi A, Dalmasso G, Billard E, Barnich N *et al*. AIEC infection triggers modification of gut microbiota composition in genetically predisposed mice, contributing to intestinal inflammation. Sci Rep, 2018. 8(1): p. 12301.
30. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A *et al*. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol, 2002. 3(7): p. RESEARCH0034.
31. Germain A, Guéant RM, Chamaillard M, Allen PB, Bresler L, Guéant JL *et al*. NOD2 gene variant is a risk factor for postoperative complications in patients with Crohn's disease: A genetic association study. Surgery, 2016. 160(1): p. 74-80.
32. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY *et al*. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature, 2012. 491(7422): p. 119-24.
33. Farin HF, Karthaus WR, Kujala P, Rakhshandehroo M, Schwank G, Vries RG *et al*. Paneth cell extrusion and release of antimicrobial products is directly controlled by immune cell-derived IFN-. J Exp Med, 2014. 211(7): p. 1393-405.

34. McPhee JB, Small CL, Reid-Yu SA, Brannon JR, Le Moual H, and Coombes BK. Host defense peptide resistance contributes to colonization and maximal intestinal pathology by Crohn's disease-associated adherent-invasive Escherichia coli. Infect Immun, 2014. 82(8): p. 3383-93.

35. Sokol H, Pigneur B, Watterlot L, Lakhdari O, Bermúdez-Humarán LG, Gratadoux JJ *et al*. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proc Natl Acad Sci U S A, 2008. 105(43): p. 16731-6.

36. Darfeuille-Michaud A, Boudeau J, Bulois P, Neut C, Glasser AL, Barnich N *et al*. High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. Gastroenterology, 2004. 127(2): p. 412-21.

37. Darfeuille-Michaud A, Neut C, Barnich N, Lederman E, Di Martino P, Desreumaux P *et al*. Presence of adherent Escherichia coli strains in ileal mucosa of patients with Crohn's disease. Gastroenterology, 1998. 115(6): p. 1405-13.

38. Barnich N, Buisson A, Auzolle C, Rodrigues M, Stefanescu C, Nancey S *et al*. P772 The presence of adherent-invasive Escherichia coli strains on the surgical specimen is a predictor of severe endoscopic postoperative recurrence in Crohn's disease. Journal of Crohn's and Colitis, 2017. 11(suppl_1): p. S475-S476.

39. Li E, Zhang Y, Tian X, Wang X, Gathungu G, Wolber A *et al*. Influence of Crohn's disease related polymorphisms in innate immune function on ileal microbiome. PLoS One, 2019. 14(2): p. e0213108.

40. Aschard H, Laville V, Tchetgen ET, Knights D, Imhann F, Seksik P *et al*. Genetic effects on the commensal microbiota in inflammatory bowel disease patients. PLoS Genet, 2019. 15(3): p. e1008018.

41. Lavoie S, Conway KL, Lassen KG, Jijon HB, Pan H, Chun E *et al*. The Crohn's disease polymorphism, ATG16L1 T300A, alters the gut microbiota and enhances the local Th1/Th17 response. Elife, 2019. 8.

42. Kline BP, Weaver T, Brinton DL, Jr., Harris L, Yochum GS, Berg AS *et al*. Clinical and Genetic Factors Impact Time to Surgical Recurrence After Ileocolectomy for Crohn Disease. Ann Surg, 2019.

43. Fowler SA, Ananthakrishnan AN, Gardet A, Stevens CR, Korzenik JR, Sands BE *et al*. SMAD3 gene variant is a risk factor for recurrent surgery in patients with Crohn's disease. J Crohns Colitis, 2014. 8(8): p. 845-51.

44. Germain A, Guéant RM, Chamaillard M, Bresler L, Guéant JL, and Peyrin-Biroulet L. CARD8 gene variant is a risk factor for recurrent surgery in patients with Crohn's disease. Dig Liver Dis, 2015. 47(11): p. 938-42.

45. Ellinghaus D, Bethune J, Petersen BS and Franke A. The genetics of Crohn's disease and ulcerative colitis--status quo and beyond. Scand J Gastroenterol, 2015. 50(1): p. 13-23.

46. Imhann F, Vich Vila A, Bonder MJ, Fu J, Gevers D, Visschedijk MC *et al*. Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. Gut, 2018. 67(1): p. 108-119.

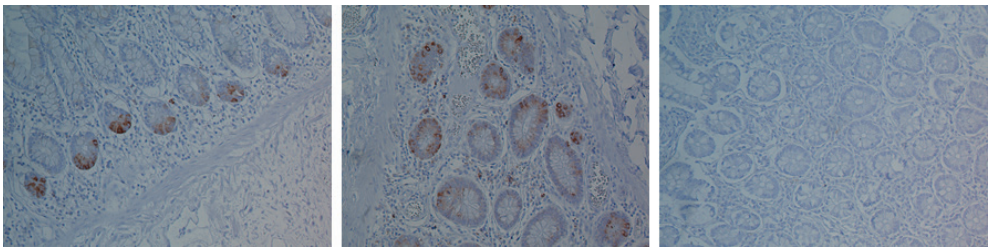
47. Gathungu G, Zhang Y, Tian X, Bonkowski E, Rowehl L, Krumsiek J *et al*. Impaired granulocyte-macrophage colony-stimulating factor bioactivity accelerates surgical recurrence in ileal Crohn's disease. World J Gastroenterol, 2018. 24(5): p. 623-630.

48. Denson LA, Jurickova I, Karns R, Shaw KA, Cutler DJ, Okou D *et al*. Genetic and Transcriptomic Variation Linked to Neutrophil Granulocyte-Macrophage Colony-Stimulating Factor Signaling in Pediatric Crohn's Disease. Inflamm Bowel Dis, 2019. 25(3): p. 547-560.

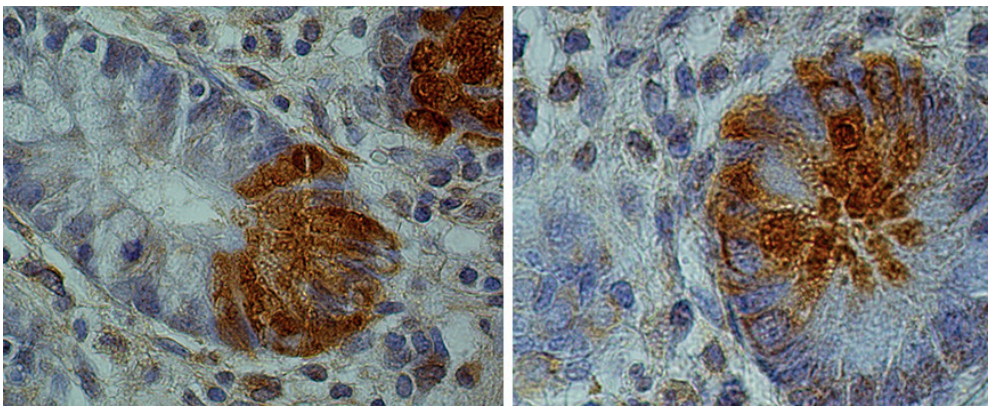
49. Cleynen I, Mahachie John JM, Henckaerts L, Van Moerkercke W, Rutgeerts P, Van Steen K *et al*. Molecular reclassification of Crohn's disease by cluster analysis of genetic variants. PLoS One, 2010. 5(9): p. e12952.

50. Bressenot A, Chevaux J-B, Williet N, Oussalah A, Germain A, Gauchotte G *et al*. Submucosal Plexitis as a Predictor of Postoperative Surgical Recurrence in Crohn's Disease. Inflammatory Bowel Diseases, 2013. 19(8): p. 1654-1661.

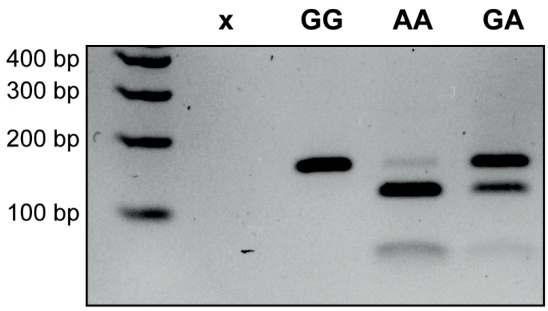
Supplementary information



▲**Supplementary Figure S1 |** ER-stress positive crypts (left and middle) and ER-stress negative crypts (right) assessed using GRP78 staining



▲**Supplementary Figure S2 |** Lysozyme staining under light microscopy (40x)



▲**Supplementary Figure S3 |** Gel electrophoresis following polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Representative examples of *ATG16L1* genotype status GG, AA GA.

10

CHAPTER 10

Toll-like receptor 1 locus re-examined in a genome-wide association study update on anti-*Helicobacter pylori* IgG titers

S.Y. Lam*[1], M.C. Mommersteeg*[1], L. Broer[2], M.C.W. Spaander[1], F. Frost[3], S. Weiss[3,4], H. Völzke[5], M.M. Lerch[3], B. Schöttker[6,7], Y. Zhang[6], H. Stocker[6,7], H. Brenner[6,7], D. Levy[8,9], S.J. Hwang[8,9], A.C. Wood[10], S.S. Rich[11], J.I. Rotter[12], K.D. Taylor[12], R.P. Tracy[13], E.K. Kabagambe[14], M. Leja[15], J. Klovins[16], R. Peculis[16], D. Rudzite[15], L. Nikitina-Zake[17], G. Skenders[15], V. Rovite[16], A. Uitterlinden[2], E.J. Kuipers[1], G.M. Fuhler[1], G. Homuth[4], M.P. Peppelenbosch[1]
* These authors contributed equally

[1] Department of Gastroenterology & Hepatology, Erasmus University Medical Center Rotterdam, The Netherlands; [2] Department of Internal Medicine, Erasmus University Medical Center Rotterdam, The Netherlands; [3] Department of Medicine A, University Medicine Greifswald, Greifswald, Germany; [4] Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; [5] Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; [6] Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; [7] Network Aging Research, University of Heidelberg, Heidelberg, Germany; [8] The Framingham Heart Study, National Heart, Lung, and Blood Institute's, Framingham, MA, USA; [9] The Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, Bethesda, MD, USA; [10] USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA; [11] Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; [12] The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at HarborUCLA Medical Center, Torrance, CA USA; [13] Laboratory for Clinical Biochemistry Research, The University of Vermont, College of Medicine, Colchester VT, USA;[14] Division of Academics, Ochsner Health, New Orleans, LA, USA; [15] Institute of Clinical and Preventive Medicine & Faculty of Medicine University of Latvia, Riga, Latvia; [16] Latvian Biomedical Research and Study Centre, Riga, Latvia; [17] Rīga Stradiņš University, Riga, Latvia.

Submitted

Abstract

Background: A genome-wide significant association between anti-*Helicobacter pylori* (*H. pylori*) IgG titers and the *Toll-like receptor 1* (*TLR1*) locus on *4p14* was demonstrated for individuals of European ancestry, but not uniformly replicated. We re-investigate this association in an updated genome-wide association study (GWAS) meta-analysis, address potential causes of heterogeneity across cohorts and explore functional implications of genetic variation at the *TLR1* locus.

Methods: The dichotomous GWAS (25% individuals exhibiting highest anti-*H. pylori* IgG titers *versus* remaining 75%) included a discovery sample of n=15,685 and a replication sample of n=9,676, all of European ancestry. Longitudinal analysis of serological data was performed on *H. pylori*-eradicated subjects (n=132) and patients under surveillance for premalignant gastric lesions (n=107). TLR1 surface expression, *TLR1* messenger RNA (mRNA) and cytokine levels were measured in leukocyte subsets of healthy subjects (n=26) genotyped for *TLR1* variants.

Results: The association of the *TLR1* locus with anti-*H. pylori* IgG titers (rs12233670; β=-0.267 SE±0.034; P=4.42x10^{-15}) presented with high heterogeneity and failed replication. Anti-*H. pylori* IgG titers declined within 2-4 years following eradication treatment (P=0.004), and decreased over time in patients with premalignant gastric lesions (P<0.001). Variation at the *TLR1* locus affected TLR1-mediated cytokine production and TLR1-surface expression on monocytes (P=0.016) and neutrophils (P=0.030), but not mRNA levels.

Conclusion: The association between anti-*H. pylori* IgG titers and *TLR1* locus was not replicated across cohorts, possibly due to dependency of anti-*H. pylori* IgG titers on therapy, clearance and antibody decay. *H. pylori*-mediated immune cell activation is partly mediated via TLR1 signaling, which in turn is affected by genetic variation.

Keywords:

Polymorphism, Single Nucleotide; serology; immunity; bacteria.

1 Introduction

The discovery of *Helicobacter pylori* (*H. pylori*) at the epithelial surface of the human stomach as late as 1983 represented a major breakthrough in gastric microbiology [1]. This flagellated bacterium has since been implicated in the etiology of atrophic gastritis and gastroduodenal ulcerative disease [2], identified as a class 1 carcinogen for gastric cancer [3-5] and ranked as the most important contributor to infection-attributable cancers in 2018 [6]. With estimates indicating that more than half of the world's population is colonized by *H. pylori*, the size of this global health problem is further emphasized [7]. Since *H. pylori* gastric presence has been linked to early stages of gastric carcinogenesis according to the Correa model [8], eradication strategies have been implemented to prevent gastric cancer development [9-12]. However, global resistance of *H. pylori* to antibiotics is reaching alarming levels [13], which puts further pressure on the *H. pylori*-related health burden and warrants new strategies to prevent colonization and infection-related consequences. It is generally accepted that *H. pylori* infection is acquired during early childhood [14-17], but the overall rate of infection is reported to be much higher in developing countries [18], Although socioeconomic and environmental factors likely explain the wide variation in *H. pylori* prevalence between regions and countries [7], genetic predisposition also needs to be considered. It has been shown that the same rearing environment contributes to a familial tendency to acquire *H. pylori* infection, but higher similarities in monozygotic than dizygotic twin pairs indicates that genetic factors account for a large part of the variation [19]. Some individuals are never infected by *H. pylori*, while others are able to clear the infection spontaneously when colonized [14, 16]. Moreover, only a small proportion of the *H. pylori* colonized population will develop gastric cancer [20], indicating that host-specific factors governing the host-pathogen interactions are involved in disease risk [21]. Since the host genetic background is suggested to be involved in the clinical outcome of *H. pylori* infection [22], a better understanding of the genetic contributions to the interaction between host and *H. pylori* may improve our insight into this complex relationship.

An increasing number of genomic association studies (GWAS) have linked single nucleotide polymorphisms (SNPs) to gastroduodenal ulcer disease [23, 24], gastric premalignant lesions [25-29], as well as gastric cancer [24, 27-39]. Interestingly, some of the associations found in these studies seem to be influenced by the presence of *H. pylori* infection [25, 29-33], suggesting that genomic variants might be involved in *H. pylori* colonization as well. The first and largest GWAS on *H. pylori* to date has combined data of Dutch and German population-based cohorts in a meta-analysis of anti-*H. pylori* IgG titers using a dichotomic study design that compared the 25% individuals exhibiting the highest anti-*H. pylori* IgG

titers *versus* the remaining 75% [40]. Two loci, namely the Toll-like receptor 1 (*TLR1*) locus on *4p14* (lead SNP rs10004195) and the Fc gamma receptor 2A (*FCGR2A*) locus on *1q23.3* (lead SNP rs368433), were identified to be associated with increased anti-*H. pylori* IgG titers [40]. A GWAS among Finnish male smokers (n=1402) confirmed the lead SNP rs10004195 to be associated with the height of IgG titer rather than a seropositive status itself [41]. However, no further replication of these findings have been reported so far, and in contrast to these European studies, no genome-wide significant associations of *H. pylori* serology with any loci were found in a Mexican-American population (n=1931) [42]. Since the main findings of the first GWAS study have not been uniformly confirmed to date [40], this study aimed to update the original meta-analysis using a larger sample size and to investigate the functional relevance of variation at the *TLR1* locus in *H. pylori* colonization.

2 Material and methods

2.1 Study cohorts

The discovery GWAS was conducted in subjects of European ancestry from population-based cohorts in Europe and the United States to re-investigate the previous association between *TLR1* and anti-*H. pylori* IgG levels, and to explore the possibility of new genetic associations. A total of seven cohorts were included and consisted of 15,685 participants (**Supplementary Table S1**). The replication was conducted in two independent European cohorts with a total of 9,676 participants with GWAS or *de novo* genotyping data. In all cohorts, serological measurements of anti-*H. pylori* IgG were performed by either a commercial or customized enzyme linked immunosorbent assay (ELISA) (**Supplementary Table S1**). As in the initial study, subjects with the highest 25% anti-*H. pylori* IgG values were compared with the remaining 75% in a dichotomous study design [40]. Informed consent for participation was obtained for all study subjects and approval was given by the Institutional Review Boards. More details concerning individual cohorts are described in the **Supplementary methods**.

2.2 Discovery

Genome-wide genotyping, imputation to 1kgP1v3 and genome wide association analyses were conducted separately by the discovery cohorts (**Supplementary methods**). EasyQC using standard settings was applied for quality control of individual cohort summary statistics [43]. The inverse-variance weighted fixed-effects model approach was employed for meta-analysis using METAL [44]. A quantile-quantile plot of observed compared to expected –log10 (*P*-value) was computed to investigate genome-wide distribution of

P-values and a Manhattan plot to illustrate genome-wide *P*-values. A regional plot was generated to show the genomic regions within 100kb of top hits. Genome wide significance was set at a threshold with *P*-value $<5.0 \times 10^{-8}$.

2.3 Replication

Eight top SNPs with the lowest association *P*-values from the discovery phase were selected for replication, in particular rs12233670 within the *TLR1* locus. The ESTHER cohort achieved *in silico* replication of seven out of eight SNPs (excluding rs147174426) and the Latvian cohort performed *de novo* genotyping for four individual SNPs (rs12233670, rs147174426, rs6107461, rs147900026) (**Supplementary methods**). Replication was considered successful with a *P*-value <0.05 for individual cohorts and a *P*-value $<5.0 \times 10^{-8}$ for the combined analysis.

2.4 Longitudinal analysis of serological data from H. pylori positive subsets

To determine whether the timing of anti-*H. pylori* IgG testing may influence serological outcomes relevant for genetic association studies, two different serological data subsets were analyzed. The first subset consisted of RS cohort study participants with pharmacy records of *H. pylori* eradication treatment prior to serology (n=132), allowing analysis of anti-*H. pylori* IgG titers in relation to time following eradication. The second subset consisted of patients from an ongoing prospective study aimed at the surveillance of atrophic gastritis, intestinal metaplasia (IM) and dysplasia in the Netherlands and Norway [45]. Anti-*H. pylori* antibodies were measured as part of the GastroPanel test (Biohit, Helsinki, Finland) using serum samples collected during clinical follow-up. Patients with elevated anti-*H. pylori* IgG levels (>30 enzyme immunoassay units) at baseline in addition to consecutive serum measurements (n=107) were included to explore fluctuation of the titers over time.

2.5 Restriction fragment length polymorphism polymerase chain reaction (RFLP-PCR) assay

Human genomic DNA was isolated from EDTA whole blood using the Kleargene Blood DNA isolation kit (LGC Ltd, Teddington, UK). RFLP-PCR could not be designed for rs10004195, but *TLR* variant rs28393318 is in complete linkage disequilibrium (LD) (r^2=1 among CEU) and was therefore used as proxy. For genotyping of rs28393318, 35 cycles of PCR amplification were performed with custom designed primers (forward: 5'-TAGCTCAGTGTAGGTGGTCT-3', reverse: 5'-ATGATTAGTGACCTTGGGGC-3') at an annealing temperature of 53˚C. PCR

products were confirmed on 2% Tris-Borate-EDTA agarose gel and 10µl of amplicons were subjected to 5 international units of Hin1II restriction enzyme (Thermofischer, Waltham, MA) for 2.5 hours at 37˚C. After 20 minutes of enzyme inactivation at 80˚C, digestion products were visualized on agarose gel, showing one band of 433 base pair (bp) for genotype GG, two bands of 311 and 122 bp for AA and three bands for GA.

2.6 Functional analysis

2.6.1 Flow cytometry

Whole blood samples from non-*H. pylori*-infected individuals without significant comorbidities (n=26) taken after informed consent were treated with eBioscience 1-step Fix/ Lyse Solution (Thermofischer) to lyse red blood cells. Monocytes and polymorphonuclear cells (PMNs) were incubated with antibodies specific for CD14 (APC-cy7, cat # A15453), CD66B (APC, cat# 17-0666-42), and TLR1 (PE, Cat #12-9911-42) (all from Thermofischer), respectively, or mouse IgG1 isotype control (PE) (BD Biosciences, Franklin Lakes, NJ, Cat # 554121) for 15 minutes on ice. Since the genes encoding TLR6 (PE, Cat # MA5-16177) and TLR10 (PE, Cat #12-2909-42) reside within the same genetic locus with as rs28393318, their surface expression was also measured. Flow cytometry was performed on the MACSQuant Flow Cytometer (Miltenyi biotec, Gladbach, Germany) and analysis conducted with FlowJo V10 (BD Biosciences). Monocytes and PMNs were identified on the basis of the forward/ sideward scatter and further gating on CD14 and CD66b, respectively. TLR positivity was measured with gating based on the isotype control of the same sample.

2.6.2 Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis of TLR1 transcript levels

Total RNA was isolated from peripheral blood mononuclear cells (PBMCs) of non-*H. pylori*-infected individuals without significant comorbidities (n=22) with the column-based NucleoSpin RNA kit (Macherey-Nagel GmbH & Co., Düren, Germany) and reversely transcribed into complementary DNA (cDNA) using PrimeScript RT (Takara, Kusastsu, Shiga, Japan). A qPCR assay of 40 cycles was performed on the StepOnePlus Real-Time PCR system (Thermofischer) using SYBR Select Master Mix (Thermofischer) and custom designed *TLR1* gene primers (forward: 5'-TGCCAAATGGAACAGACAAGCAG-3', reverse: 5'-ACAGATTCCTTTTGTAGGGG TGCC-3') and *RP2* housekeeping gene primers (forward: 5'-AAGCTGAGGATGCTCA AAGG-3', reverse: 5'-CCCATTAAACTCCAAGGCAA-3'). The annealing temperature was 61˚C for both primer sets. The delta-delta cycle threshold (ΔΔCt) method was applied for data analysis.

2.6.3 Enzyme linked immunosorbent assay (ELISA) for cytokine analysis upon TLR1 stimulation

PBMCs from non-*H. pylori*-infected individuals (n=22) were isolated from heparinized blood as described [46]. In brief, phosphate buffered saline (PBS) diluted blood was layered onto Ficoll (Amersham, Upsala, Sweden) and PBMCs harvested after centrifugation, washed in PBS and plated in Roswell Park Memorial Institute medium (Lonza, Basel, Switzerland) containing 10% fetal bovine serum (FBS) and penicillin/streptomycin (Lonza). Adherent PBMCs were stimulated with 1e6 CFU of heat-killed *H. pylori* (strain ATCC-43504 [cagA$^+$, vacA$^{(s1/m1)}$, iceA$^+$, babA2$^+$], Manassas, VA) grown on Trypticase Soy Agar (Oxoid, Hampshire, UK) that was supplemented with 5% defibrinated sheep blood (VWR, Radnor, PA) and DENT selective medium (Oxoid). Other stimuli used were TLR1 inhibitor CU-CPT-22 (5μM) (Tocris Bioscience, Bristol, UK) and TLR1 agonist Pam3Cys4 (300ng/ml) (InvivoGen, San Diego, CA) [47]. Supernatants were harvested after eight hours of stimulation for ELISA experiments unless otherwise specified to measure tumor necrosis factor alpha (TNFα), interleukin 8 (IL-8) and IL-10 (eBioscience, San Diego, CA) as described previously [48]. All samples were tested in duplicate.

2.6.4 Statistical analysis of serological and functional data

Statistical differences among three groups were determined with the one-way analysis of variance or Kruskal-Wallis tests for unpaired data and repeated measures analysis of variance or Friedman tests for paired data and was followed by post-hoc analysis for selected pairs with adjustment for multiple testing. The two-sample t-test or Mann-Whitney test were applied to compare two groups with unpaired data. GraphPad Prism software version 5.01 (GraphPad Software Inc, San Diego, CA) were applied for calculations and graphical representation.

3 Results

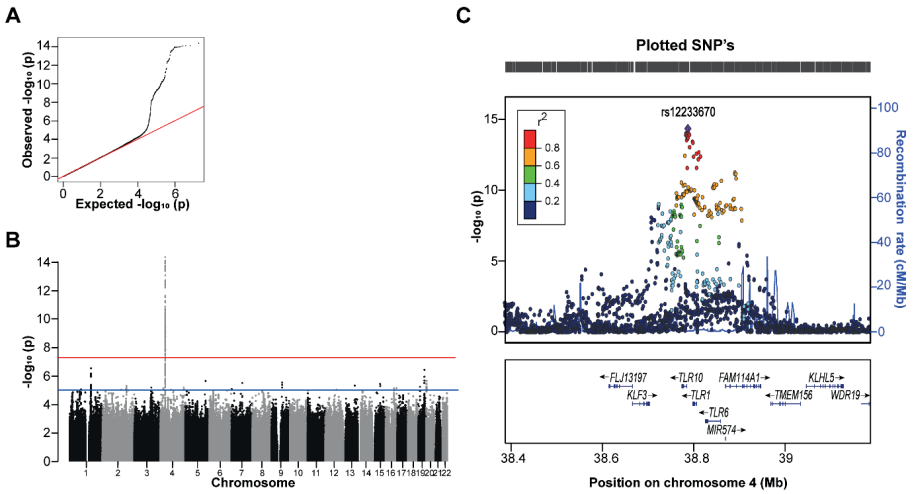
3.1 TLR1 SNP and other genomic variants associated with anti-H. pylori IgG titers in an updated GWAS

We performed a GWAS meta-analysis based on seven independent European epidemiological cohorts. The quantile-quantile (QQ) plot showed a clear deviation from the null-distribution at the tail (**Figure 1A**). A genome-wide significant association for the *TLR1* locus on chromosome *4p14* was found with top SNP rs12233670 carrying the lowest *P*-value (β=-0.267 SE ±0.034 for minor allele T; $P=4.42x10^{-15}$; MAF=25%) (**Figure 1B,C**), albeit with statistical heterogeneity (**Table 1**). The association between top SNP rs12233670 and

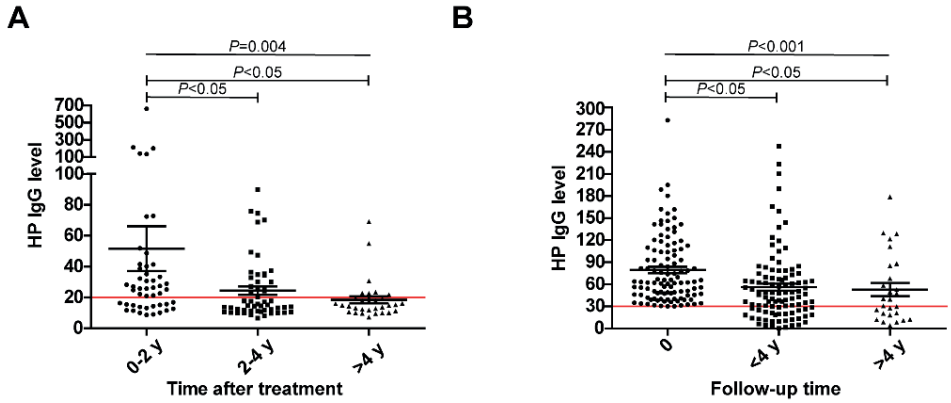
Table 1 | Summary of single nucleotide polymorphism (SNP) at the *TLR1 locus* in discovery, replication and combined meta-analysis

rs12233670				Discovery					Replication					Combined				
Chr	**Position**	**Gene** [a]	**A½** [b]	**EAF**	**Beta** [c]	**SE**	***P*-value**	**I²**	**EAF**	**Beta** [c]	**SE**	***P*-value**	**I²**	**Beta** [c]	**SE**	**Dir** [d]	***P*-value**	**I²**
4	38787216	*TLR10*	T/C	0.25	-0.267	0.034	$4.42x10^{-15}$	79.8	0.20	0.034	0.042	0.417	0.0	-0.149	0.027	-++	$1.97x10^{-08}$	30.9

[a] Gene nearest to SNP
[b] A½, effect allele 1 and the other allele 2;
[c] effect size is relative to the allele 1
[d] Direction of beta of respectively the discovery, ESTHER and Latvia cohorts: positive (+) or negative (-) or not determined (?)
Abbreviations: Chr, chromosome; EAF, effect allele frequency; I², measure of heterogeneity; SE, standard error; SNP; +/-.



◂Figure 1 | Results of genome-wide association meta-analysis for anti-*Helicobacter pylori* (*H. pylori*) IgG titer. (A) Quantile-quantile plot with all single nucleotide polymorphisms (SNPs) displayed by black dots and the red line corresponding to the null-hypothesis of no true association. (**B)** Manhattan plot of the genome wide association meta-analysis with the chromosome position on the *X*-axis and $-\log_{10}$ *P*-values on the *Y*-axis. Each dot indicates a SNP, the blue line marks the threshold of $P=1.0x10^{-5}$ and red horizontal line represents the genome wide significant threshold of $P=5.0x10^{-8}$. **(C)** Regional plot showing locus *4p14* with rs12233670 and other flanking region markers. The *P*-values of SNPs associated with *H. pylori* are plotted against their chromosome position on the *X*-axis. Each SNP is represented by a coloured dot which indicates their correlation (Linkage disequilibrium) with the top ranked SNP (purple diamond). The *Y*-axis at the right represents the recombination rates. The bottom part depicts the annotated genes at the locus and their transcriptional direction.



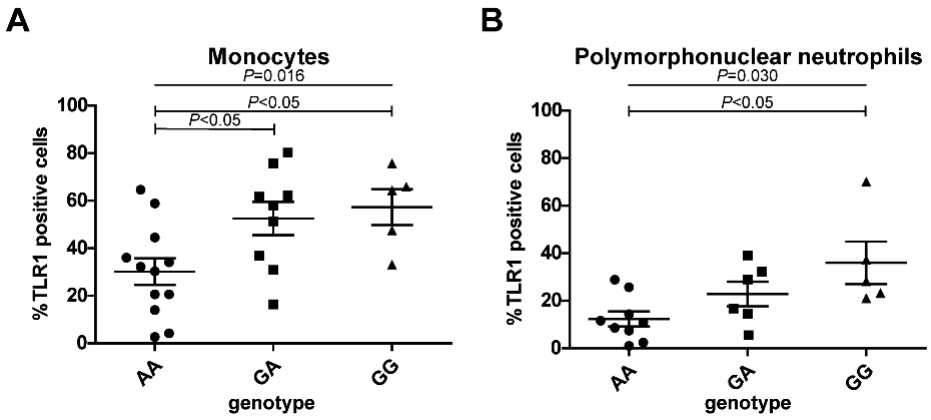
▲**Figure 2 | Anti-*Helicobacter pylori* (*H. pylori*) IgG levels over time in two subsets. (A)** IgG levels of Rotterdam Study subjects (n=132) that received eradication therapy prior to serological testing. The measurements of subjects are divided in three groups based on the time between treatment and *H. pylori* serology: 0-2 (n=48), 2-4 (n=53) and >4 (n=31) years (y). **(B)** IgG levels of *H. pylori*-positive patients with premalignant gastric lesions (n=107) at baseline (time point 0 with IgG titers >30 enzyme immunoassay units) and during serological follow-up at <4 (n=104) and/or >4 (n=25) years (y). In both plots, the mean ±SEM and the manufacturer's test cut-off (red horizontal line) are shown.

anti-*H. pylori* IgG titers was not significant in either ESTHER (β=0.041 SE ± 0.050 for the minor allele; *P*=0.41) or Latvia (β=0.017 SE ±0.079 for the minor allele; *P*=0.83) cohorts, resulting in a failed replication (β=0.034 SE ± 0.042 for minor allele; *P*=0.42) (**Table 1**). Consequently, the level of genome-wide significance decreased in the combined analysis (β=-0.149 SE ± 0.027 for the minor allele; $P=1.97 \times 10^{-08}$) (**Table 1**).

Seven other promising SNPs were identified, but did not reach genome wide significance, including the *FCGR* locus (1q23.3; top-ranked SNP rs147174426, β=0.480 SE ±0.094 for major allele A; $P=2.89 \times 10^{-07}$; MAF=7%). Similar results were obtained using a sensitivity model including adult participants only (data not shown). None of these selected SNPs reached genome-wide significance in the combined analysis with discovery and replication cohorts (**Supplementary Table S2**).

3.2 Anti-H. pylori antibody decay in H. pylori-infected subjects

We considered that antibody decay and timing of sampling for *H. pylori* serology may contribute to cohort heterogeneity. To investigate the serological course of *H. pylori*-infected subjects, anti-*H. pylori* IgG data were studied in two settings. In a subset of RS study participants that received *H. pylori* eradication treatment at some point prior to the measurement of IgG antibodies (n=132), titers were significantly higher in individuals tested

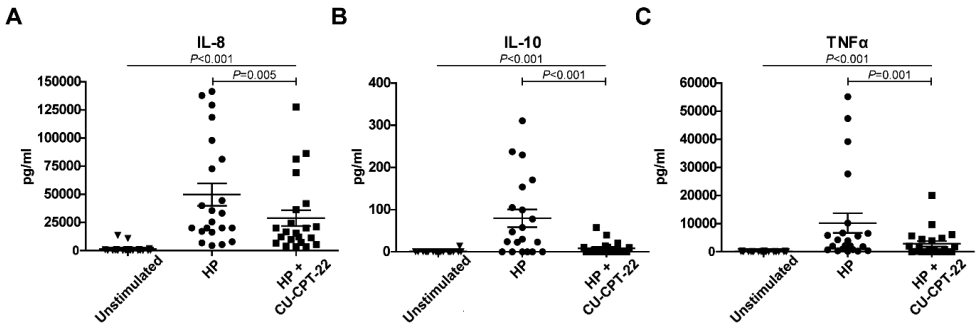


▲Figure 3 | Measurement of Toll-like receptor 1 (TLR1) positive cells by flow cytometry. Dot plots illustrating the percentage of TLR1 positive cells in healthy subjects genotyped for rs28393318. The mean ±SEM and statistical significance between three genotypes is shown. **(A)** TLR1 positivity of monocytes in AA (n=12), GA (n=9) and GG (n=5) carriers. **(B)** TLR1 positivity of polymorphonuclear neutrophils in AA (n=9), GA (n=6) and GG (n=5) carriers.

within 0-2 (n=48) years after receiving eradication therapy than in those tested 2-4 (n=53) or >4 (n=31) years (*P*=0.004) post-treatment (**Figure 2A**). When analyzing sequential anti-*H. pylori* IgG titers from patients with gastric premalignant lesions with positive *H. pylori* serology (n=107), a significant decline between baseline measurement (time point 0) and retesting at <4 (n=104) or >4 (n=25) years of medical follow-up (*P*<0.001) was seen (**Figure 2B**). Together these data indicate anti-*H. pylori* antibody decay occurs within 2 years after treatment or clearance of *H. pylori*.

3.3 Higher TLR1 surface protein, but not intracellular mRNA expression levels in leukocytes of G allele carriers of rs28393318

To investigate potential functional consequences of variation at the *4q14* locus, expression of the TLR encoding genes within this locus was investigated in healthy subjects (n=26) genotyped for rs28393318. A significant difference in the percentage of TLR1-positive monocytes (*P*=0.016) and PMNs (*P*=0.030) was observed between AA, GA and GG genotype carriers (**Figure 3A-B**). Post-hoc analysis revealed significantly higher TLR1 surface expression on monocytes for carriers of the minor rs28393318 allele (G) and on PMNs in subjects homozygous for the G allele compared to A allele carriers. Variation at rs28393318 did not influence TLR6 and TLR10 surface expression on either monocytes or PMNs (**Supplementary Figure S1**). Unlike previous RNAseq-based findings of reduced *TLR1* mRNA levels for minor rs10004195 A allele carriers [40], our RT-qPCR findings



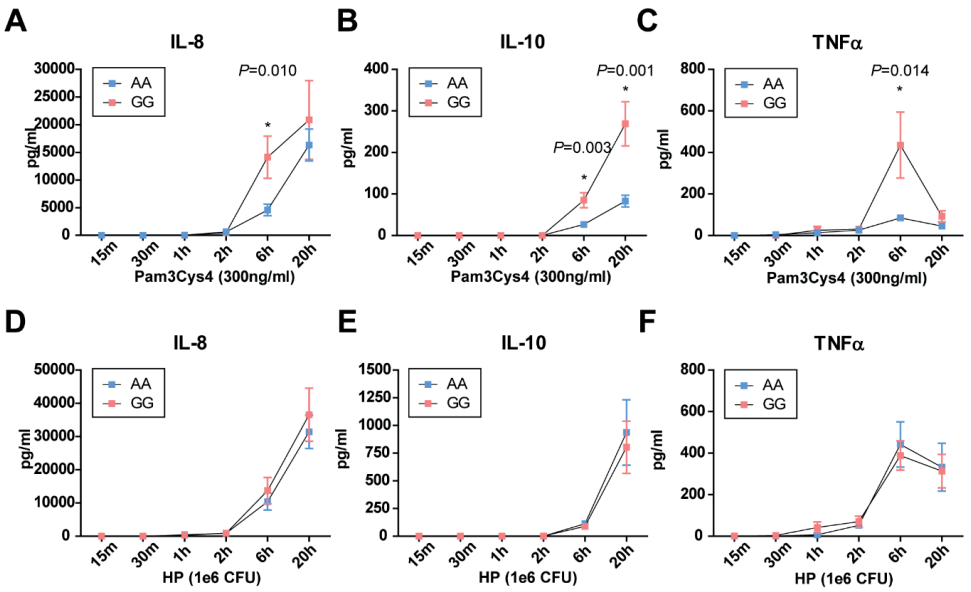
▲Figure 4 | Cytokine production by peripheral blood mononuclear cells (PBMCs) upon stimulation with *Helicobacter pylori (H. pylori).* (A-C) Interleukin 8 (IL-8) (A), interleukin 10 (IL-10) (B) and tumor necrosis factor alfa (TNFα) (C) levels in picogram per milliliter (pg/ml) by PMCBs of healthy subjects (n=22). Cytokine levels were measured at baseline without stimulation and following *H. pylori* exposure in the absence and presence of TLR1 inhibitor CU-CPT-22.

showed no differences in *TLR1* transcript levels in PBMCs between healthy subjects (n=22) with different genotypes of rs28393318 (**Supplementary Figure S2**), which is in line with previous reports demonstrating no differences in mRNA and total cellular TLR1 protein levels among other TLR1 variants (in high LD with rs28393318 and rs10004195 among CEU) tested [49-52].

3.4 TLR1 rs28393318 affects immune cell cytokine production

To first confirm TLR1 involvement in *H. pylori* pathogenesis, PBMCs from healthy subjects (n=22) were treated with *H. pylori* in the presence of selective TLR1 inhibitor CU-CPT-22 or vehicle control. *H. pylori* significantly stimulated IL-8, IL-10 and TNFα production (all *P*<0.001), which was significantly but not fully abrogated upon treatment of cells with the TLR1 inhibitor CU-CPT-22 (*P*=0.005 for IL8, *P*<0.001 for IL10 and *P*=0.001 for TNFα) (**Figure 4**). These findings suggest that *H. pylori*-related cytokine signalling is partly mediated *via* TLR1.

We next explored TLR1-mediated differences in cytokine production between healthy subjects with rs28393318 genotypes AA (n=12) and GG (n=8). PBMCs were stimulated with the specific TLR1 agonist Pam3Cys4 or with *H. pylori*, where after cytokine levels were measured at different time points (¼, ½, 1, 2, 6 and 20 hour(s)). Pam3Cys4 stimulation resulted in higher IL-8 (*P*=0.010), IL-10 (*P*=0.003) and TNFα (*P*=0.014) production at 6 hours and also higher IL-10 levels at 20 hours (*P*=0.001) in GG carriers compared to AA carriers. Cytokine production upon *H. pylori* treatment was considerably higher than upon Pam3Cys4 stimulation, but no differences among genotypes were observed (**Figure 5**).



▲Figure 5 | Time course of cytokine production by peripheral blood mononuclear cells (PBMCs) upon stimulation with TLR1 ligand Pam3Cys4 or *Helicobacter pylori* (*H. pylori*) in AA and GG carriers of rs28393318. (A-F) Interleukin 8 (IL-8), interleukin 10 (IL-10) and tumor necrosis factor alpha (TNFα) levels in picogram per milliter (pg/ml) measured at different time points following stimulation with TLR1 agonist Pam3Cys4 (A-C) or *H. pylori* (D-F). The results are stratified for AA (n=12) and GG (n=8) carriers of rs28393318. The mean ±SEM and statistical significance between genotypes is shown. Abbreviations: CFU, colony forming unit; h, hour; m, minutes.

4 Discussion

This study aimed to better understand the genome-wide association between the *TLR1* locus and *H. pylori*. We extended on the original work of Mayerle *et al.* [40] by the inclusion of an additional 4,747 subjects of European ancestry in an updated GWAS meta-analysis. An association between anti-*H. pylori* IgG titers and the *TLR1* locus with top SNP rs12233670 was demonstrated in the discovery phase, but replication proved to be challenging. Significant heterogeneity for our top association was observed across cohorts, with SHIP-TREND and both replication cohorts showing association in the opposite direction. The interpretation of these findings remains complex, but might be partially explained by methodological differences that are inherent in the inclusion of longitudinally population cohorts, including time of recruitment (e.g. SHIP *versus* SHIP-TREND) and employment of non-uniform serological assays. The concession of accounting for false positive assignment of cases by using the 25% highest *versus* 75% lowest of IgG distribution might be another explanation. The allocation of truly *H. pylori*-infected subjects into the control group could

have possibly limited the detection and replication of promising SNPs, in particular for high endemic regions such as the republic of Latvia [53]. Furthermore, studying two different serological datasets we demonstrated that anti-*H. pylori* IgG antibody decay over time occurs relatively quickly, as was also previously observed upon *H. pylori* eradication treatment [54]. Knowing that antibody decay takes place, it is imperative to know the time of collection in relation to *H. pylori* infection. Different rates of spontaneous clearance, re-infection and *H. pylori* eradication might have contributed to study heterogeneity, but this information was unfortunately not routinely collected. Lastly, various *H. pylori* strains with varying virulence may interact differently with their human host, influencing the clinical outcome [55].

Despite technical challenges preventing a clear replication, several studies do point towards a role of the *TLR1* locus in *H. pylori* infectivity [40, 41]. Since the data of the discovery phase showed an association for *H. pylori* in the same direction as the original report, the relevance of the *TLR1* locus in relation to *H. pylori* pathology was indicated. Our functional experiments demonstrate that variation at the *TLR1* locus indeed has functional implications as shown by a higher TLR1 surface expression and higher cytokine production in minor allele G carriers of rs28393318 (which is in complete LD with our top SNP rs12233670 and rs10004195 among CEU). This might be explained by two non-synonymous *TLR1* variants affecting intracellular-to-cell surface trafficking (rs5743618; r^2=0.86 with rs28393318 among CEU) and transportation of the receptor to the cell membrane (rs4833095; r^2=0.97 with rs28393318 among CEU) [49-52, 56, 57]. Minor allele carriers of these *TLR1* genetic variants displayed higher cytokine responses upon targeted TLR1 stimulation (i.e. Pam3Cys4) which was attributed to increased TLR1 surface expression rather than changes in total protein or mRNA levels measured in cells [49-52], which is in line with our findings. While *H. pylori* mediates IL-8, IL-10 and TNFα production at least partially *via* TLR1 signaling in PBMCs, *H. pylori*-mediated cytokine production was not affected by rs28393318 status of carriers. It is likely that rs28393318 variation effects are masked by other components of the host immune system triggered by *H. pylori* [58]. Similarly, the effect of rs28393318 variation on serological titers induced by *H. pylori* infection may be masked by additional *H. pylori*-induced host-specific immune responses.

This study has two major limitations that need to be addressed. First, the identification of new genetic variants and replication of promising candidates for anti-*H. pylori* IgG titers in this study may have been hampered by the chosen definition of seropositivity for *H. pylori* antibodies. Future studies might have to reconsider the phenotype definition as the interpretation of *H. pylori* serological determination is not straightforward. Anti-*H.*

pylori IgG levels are more likely to represent a combination of the host ability to mount an immunological response to infection as well as antibody clearance, than actual *H. pylori* incidence. It should also be noted that *H. pylori* infection involves an interplay of factors (host, bacterial and environmental). Many genetic variants have been identified for *H. pylori*-related conditions such as ulcerative and (pre)malignant gastric lesions in different ethnic populations [23-25, 27-32, 38], and therefore it seems plausible that other genetic variants are relevant in *H. pylori* pathogenesis besides the *TLR* locus. Secondly, we have tested a selective cytokine panel in our functional analysis as a proof of concept. To better understand the immune response with regard to *H. pylori* susceptibility, future experiments with different *H. pylori* strains in different ethnic populations would be of interest.

In summary, the previously observed association between the *TLR1* locus and *H. pylori* was not uniformly confirmed across cohorts in this study. The interpretation of *H. pylori* serology is complex and subject to alterations in response to therapy and over time. While variation at the *TLR1* locus regulates surface expression and cytokine production upon stimulation, further efforts are required to better understand the clinical relevance of *TLR1* variants and other loci in their complex interaction with *H. pylori*.

References

1. Warren JR and Marshall B. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet, 1983. 1(8336): p. 1273-5.
2. Marshall BJ and Warren JR. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet, 1984. 1(8390): p. 1311-5.
3. Schistosomes, liver flukes and Helicobacter pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Lyon, 7-14 June 1994. IARC Monogr Eval Carcinog Risks Hum, 1994. 61: p. 1-241.
4. Kuipers EJ, Uyterlinde AM, Pena AS, Roosendaal R, Pals G, Nelis GF *et al.* Long-term sequelae of Helicobacter pylori gastritis. Lancet, 1995. 345(8964): p. 1525-8.
5. Helicobacter and Cancer Collaborative G. Gastric cancer and Helicobacter pylori: a combined analysis of 12 case control studies nested within prospective cohorts. Gut, 2001. 49(3): p. 347-53.
6. de Martel C, Georges D, Bray F, Ferlay J and Clifford GM. Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis. Lancet Glob Health, 2020. 8(2): p. e180-e190.
7. Hooi JKY, Lai WY, Ng WK, Suen MMY, Underwood FE, Tanyingoh D *et al.* Global Prevalence of Helicobacter pylori Infection: Systematic Review and Meta-Analysis. Gastroenterology, 2017. 153(2): p. 420-429.
8. Correa P. Human gastric carcinogenesis: a multistep and multifactorial process--First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res, 1992. 52(24): p. 6735-40.
9. Wong BC, Lam SK, Wong WM, Chen JS, Zheng TT, Feng RE *et al.* Helicobacter pylori eradication to prevent gastric cancer in a high-risk region of China: a randomized controlled trial. Jama, 2004. 291(2): p. 187-94.
10. Ford AC, Forman D, Hunt RH, Yuan Y and Moayyedi P. Helicobacter pylori eradication therapy to prevent gastric cancer in healthy asymptomatic infected individuals: systematic review and meta-analysis of randomised controlled trials. Bmj, 2014. 348: p. g3174.
11. Lee YC, Chiang TH, Chou CK, Tu YK, Liao WC, Wu MS *et al.* Association Between Helicobacter pylori Eradication and Gastric Cancer Incidence: A Systematic Review and Meta-analysis. Gastroenterology, 2016. 150(5): p. 1113-1124 e5.
12. Leung WK, Wong IOL, Cheung KS, Yeung KF, Chan EW, Wong AYS *et al.* Effects of Helicobacter pylori Treatment on Incidence of Gastric Cancer in Older Individuals. Gastroenterology, 2018. 155(1): p. 67-75.
13. Savoldi A, Carrara E, Graham DY, Conti M and Tacconelli E. Prevalence of Antibiotic Resistance in Helicobacter pylori: A Systematic Review and Meta-Analysis in World Health Organization Regions. Gastroenterology, 2018.
14. Kuipers EJ, Pena AS, van Kamp G, Uyterlinde AM, Pals G, Pels NF *et al.* Seroconversion for Helicobacter pylori. Lancet, 1993. 342(8867): p. 328-31.
15. Malaty HM, El-Kasabany A, Graham DY, Miller CC, Reddy SG, Srinivasan SR *et al.* Age at acquisition of Helicobacter pylori infection: a follow-up study from infancy to adulthood. Lancet, 2002. 359(9310): p. 931-5.
16. Perez-Perez GI, Sack RB, Reid R, Santosham M, Croll J, and Blaser MJ. Transient and persistent Helicobacter pylori colonization in Native American children. J Clin Microbiol, 2003. 41(6): p. 2401-7.
17. Rowland M, Daly L, Vaughan M, Higgins A, Bourke B, and Drumm B. Age-specific incidence of Helicobacter pylori. Gastroenterology, 2006. 130(1): p. 65-72; quiz 211.

18. Bardhan PK. Epidemiological features of Helicobacter pylori infection in developing countries. Clin Infect Dis, 1997. 25(5): p. 973-8.
19. Malaty HM, Engstrand L, Pedersen NL and Graham DY. Helicobacter pylori infection: genetic and environmental influences. A study of twins. Ann Intern Med, 1994. 120(12): p. 982-6.
20. Kuipers EJ. Review article: exploring the link between Helicobacter pylori and gastric cancer. Aliment Pharmacol Ther, 1999. 13 Suppl 1: p. 3-11.
21. Polk DB and Peek RM, Jr. Helicobacter pylori: gastric cancer and beyond. Nat Rev Cancer, 2010. 10(6): p. 403-14.
22. Mommersteeg MC, Yu J, Peppelenbosch MP and Fuhler GM. Genetic host factors in Helicobacter pylori-induced carcinogenesis: Emerging new paradigms. Biochim Biophys Acta Rev Cancer, 2018. 1869(1): p. 42-52.
23. Tanikawa C, Urabe Y, Matsuo K, Kubo M, Takahashi A, Ito H *et al*. A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. Nat Genet, 2012. 44(4): p. 430-4, S1-2.
24. Garcia-Gonzalez MA, Bujanda L, Quintero E, Santolaria S, Benito R, Strunk M *et al*. Association of PSCA rs2294008 gene variants with poor prognosis and increased susceptibility to gastric cancer and decreased risk of duodenal ulcer disease. Int J Cancer, 2015. 137(6): p. 1362-73.
25. Jiang J, Jia ZF, Kong F, Jin MS, Wang YP, Tian S *et al*. Association of polymorphism of PTPN 11 encoding SHP-2 with gastric atrophy but not gastric cancer in Helicobacter pylori seropositive Chinese population. BMC Gastroenterol, 2012. 12: p. 89.
26. Tang FB, Li ZX, Wang YM, Zhang L, Ma JL, Zhou T *et al*. Toll-like receptor 1 and 10 polymorphisms, Helicobacter pylori susceptibility and risk of gastric lesions in a high-risk Chinese population. Infect Genet Evol, 2015. 31: p. 263-9.
27. Rizzato C, Kato I, Plummer M, Munoz N and Canzian F. Genetic variation in PSCA and risk of gastric advanced preneoplastic lesions and cancer in relation to Helicobacter pylori infection. PLoS One, 2013. 8(9): p. e73100.
28. He CY, Sun LP, Xu Q, Liu JW, Jiang JY, Dong NN *et al*. PGC TagSNP and its interaction with H. pylori and relation with gene expression in susceptibility to gastric carcinogenesis. PLoS One, 2014. 9(12): p. e115955.
29. Cai M, Dai S, Chen W, Xia C, Lu L, Dai S *et al*. Environmental factors, seven GWAS-identified susceptibility loci, and risk of gastric cancer and its precursors in a Chinese population. Cancer Med, 2017. 6(3): p. 708-720.
30. He BS, Sun HL, Xu T, Pan YQ, Lin K, Gao TY *et al*. Association of Genetic Polymorphisms in the LncRNAs with Gastric Cancer Risk in a Chinese Population. J Cancer, 2017. 8(4): p. 531-536.
31. Sala N, Munoz X, Travier N, Agudo A, Duell EJ, Moreno V *et al*. Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: results from the EPIC-EURGAST study. Int J Cancer, 2012. 130(10): p. 2417-27.
32. Li M, Huang L, Qiu H, Fu Q, Li W, Yu Q *et al*. Helicobacter pylori infection synergizes with three inflammation-related genetic variants in the GWASs to increase risk of gastric cancer in a Chinese population. PLoS One, 2013. 8(9): p. e74976.
33. Ravishankar Ram M, Goh KL, Leow AH, Poh BH, Loke MF, Harrison R *et al*. Polymorphisms at Locus 4p14 of Toll-Like Receptors TLR-1 and TLR-10 Confer Susceptibility to Gastric Carcinoma in Helicobacter pylori Infection. PLoS One, 2015. 10(11): p. e0141865.

34. Study Group of Millennium Genome Project for C, Sakamoto H, Yoshimura K, Saeki N, Katai H, Shimoda T *et al*. Genetic variation in PSCA is associated with susceptibility to diffuse-type gastric cancer. Nat Genet, 2008. 40(6): p. 730-40.
35. Abnet CC, Freedman ND, Hu N, Wang Z, Yu K, Shu XO *et al*. A shared susceptibility locus in PLCE1 at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. Nat Genet, 2010. 42(9): p. 764-7.
36. Saeki N, Saito A, Choi IJ, Matsuo K, Ohnami S, Totsuka H *et al*. A functional single nucleotide polymorphism in mucin 1, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. Gastroenterology, 2011. 140(3): p. 892-902.
37. Shi Y, Hu Z, Wu C, Dai J, Li H, Dong J *et al*. A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. Nat Genet, 2011. 43(12): p. 1215-8.
38. Companioni O, Bonet C, Munoz X, Weiderpass E, Panico S, Tumino R *et al*. Polymorphisms of Helicobacter pylori signaling pathway genes and gastric cancer risk in the European Prospective Investigation into Cancer-Eurgast cohort. Int J Cancer, 2014. 134(1): p. 92-101.
39. Dargiene G, Streleckiene G, Skieceviciene J, Leja M, Link A, Wex T *et al*. TLR1 and PRKAA1 Gene Polymorphisms in the Development of Atrophic Gastritis and Gastric Cancer. J Gastrointestin Liver Dis, 2018. 27(4): p. 363-369.
40. Mayerle J, den Hoed CM, Schurmann C, Stolk L, Homuth G, Peters MJ *et al*. Identification of genetic loci associated with Helicobacter pylori serologic status. Jama, 2013. 309(18): p. 1912-20.
41. Sung H, Camargo MC, Yu K, Weinstein SJ, Morgan DR, Albanes D *et al*. Association of 4p14 TLR locus with antibodies to Helicobacter pylori. Genes Immun, 2015. 16(8): p. 567-70.
42. Rubicz R, Yolken R, Drigalenko E, Carless MA, Dyer TD, Kent J, Jr. *et al*. Genome-wide genetic investigation of serological measures of common infections. Eur J Hum Genet, 2015. 23(11): p. 1544-8.
43. Winkler TW, Day FR, Croteau-Chonka DC, Wood AR, Locke AE, Mägi R *et al*. Quality control and conduct of genome-wide association meta-analyses. Nat Protoc, 2014. 9(5): p. 1192-212.
44. Willer CJ, Li Y and Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics, 2010. 26(17): p. 2190-1.
45. den Hollander WJ, Holster IL, den Hoed CM, Capelle LG, Tang TJ, Anten MP *et al*. Surveillance of premalignant gastric lesions: a multicentre prospective cohort study from low incidence regions. Gut, 2019. 68(4): p. 585-593.
46. Somasundaram R, Fernandes S, Deuring JJ, de Haar C, Kuipers EJ, Vogelaar L *et al*. Analysis of SHIP1 expression and activity in Crohn's disease patients. PLoS One, 2017. 12(8): p. e0182308.
47. Cheng K, Wang X, Zhang S and Yin H. Discovery of small-molecule inhibitors of the TLR1/TLR2 complex. Angew Chem Int Ed Engl, 2012. 51(49): p. 12246-9.
48. Lie M, van der Giessen J, Fuhler GM, de Lima A, Peppelenbosch MP, van der Ent C *et al*. Low dose Naltrexone for induction of remission in inflammatory bowel disease patients. J Transl Med, 2018. 16(1): p. 55.
49. Johnson CM, Lyle EA, Omueti KO, Stepensky VA, Yegin O, Alpsoy E *et al*. Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. J Immunol, 2007. 178(12): p. 7520-4.
50. Hawn TR, Misch EA, Dunstan SJ, Thwaites GE, Lan NT, Quy HT *et al*. A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. Eur J Immunol, 2007. 37(8): p. 2280-9.

51. Wurfel MM, Gordon AC, Holden TD, Radella F, Strout J, Kajikawa O *et al*. Toll-like receptor 1 polymorphisms affect innate immune responses and outcomes in sepsis. Am J Respir Crit Care Med, 2008. 178(7): p. 710-20.

52. Uciechowski P, Imhoff H, Lange C, Meyer CG, Browne EN, Kirsten DK *et al*. Susceptibility to tuberculosis is associated with TLR1 polymorphisms resulting in a lack of TLR1 cell surface expression. J Leukoc Biol, 2011. 90(2): p. 377-88.

53. Leja M, Cine E, Rudzite D, Vilkoite I, Huttunen T, Daugule I *et al*. Prevalence of Helicobacter pylori infection and atrophic gastritis in Latvia. Eur J Gastroenterol Hepatol, 2012. 24(12): p. 1410-7.

54. Feldman M, Cryer B, Lee E and Peterson WL. Role of seroconversion in confirming cure of Helicobacter pylori infection. Jama, 1998. 280(4): p. 363-5.

55. Correa P and Piazuelo MB. Evolutionary History of the Helicobacter pylori Genome: Implications for Gastric Carcinogenesis. Gut Liver, 2012. 6(1): p. 21-8.

56. Schumann RR and Tapping RI. Genomic variants of TLR1--it takes (TLR-)two to tango. Eur J Immunol, 2007. 37(8): p. 2059-62.

57. Mikacenic C, Schneider A, Radella F, Buckner JH and Wurfel MM. Cutting edge: Genetic variation in TLR1 is associated with Pam3CSK4-induced effector T cell resistance to regulatory T cell suppression. J Immunol, 2014. 193(12): p. 5786-90.

58. Peek RM, Jr., Fiske C and Wilson KT. Role of innate immunity in Helicobacter pylori-induced gastric malignancy. Physiol Rev, 2010. 90(3): p. 831-58.

Supplementary information

Supplementary methods

1 Discovery cohorts

The genome-wide association study (GWAS) of Mayerle *et al.* was comprised of data from the Rotterdam Study (RS) and the Study of Health in Pomerania (SHIP) with a total of 10,938 participants [1]. For our current meta-analysis, a total of 15,685 participants of European ancestry with GWAS data and *H. pylori* serology measurements were included from RS-I (n=4771), RS-II (n=2112), SHIP (n=3830), SHIP-TREND (n=983), Framingham Heart Study (FHS) (n=3141), Multi-Ethnic Study of Atherosclerosis (MESA) (n=447) and Generation R (GenR) study (n=401) cohorts.

RS-I and RS-II

The RS study is an ongoing prospective Dutch study in which participants of the well-defined district Ommoord in Rotterdam were enrolled in sequential cohorts over time to investigate a variety of diseases commonly in the elderly [2-4]. Baseline recruitment of the first three cohorts (RS-I, RS II, RS-III) dated from respectively 1990-1993 (n=7983), 2000-2001 (n=3011) and 2006-2008 (n=3932), and was extended with a fourth cohort that started in 2016 (RS-IV). All participants were interviewed and underwent an extensive set of examinations focusing on possible causes of invalidating diseases in the elderly. Examinations were repeated every 3-4 years in sequential examination cycles, with the emphasis on imaging and the collection of biological specimens.

SHIP and SHIP-TREND

SHIP is a population-based study which was set up in the North-East of Germany to investigate a wide range of health-related conditions [5, 6]. The main objectives of the study were to assess the prevalence and incidence of common risk factors, subclinical disorders and clinical diseases, and to investigate the complex associations among them. It comprised of two independent cohorts including 20-79 year old participants during 1997-2001 (n=6265) and 2008-2012 (n=8016) for SHIP and SHIP-TREND, respectively. Baseline data collection were obtained from a health-related interview, a health- and risk questionnaire and a medical examination during which biological samples were taken. Further information was collected during follow-up, including morbidity and mortality data.

FHS

FHS is a longitudinal multigenerational study which was originally set up to investigate the epidemiology of cardiovascular diseases in the residents of Framingham, as described

previously [7-11]. After completion of the baseline recruitment of the Original cohort in 1953 (n=5209), the process was repeated for the Offspring and Spouses cohort in 1971-1975 (n=5124), the Third Generation cohort in 2002-2005 (n=4095) and the Offspring Spouses cohort during 2003-2005 (n=103). While these subjects were predominantly of Western-European descent, the Omni 1 (n=507) and 2 (n=410) cohorts represent the multi-ethnic society in Framingham [11]. Examinations cycles were performed every 2-6 years and included the collection of biological samples.

MESA

MESA is large population-based study in the United States designed to study factors influencing the conversion of subclinical cardiovascular disease to overt disease in four different ethnic groups (Caucasian-, African-, Hispanic-, and Asian- American) [12-14]. Study subjects were 48-84 years old and free of overt signs of cardiovascular disease at baseline enrolment in 2000-2002 (n=6814) and recruited from six field centers across the United States. At baseline, extensive data was obtained from comprehensive questionnaires, the collection of biological samples and physical assessments incorporating various imaging technologies. Follow-up examinations were conducted at ~18 month intervals in MESA.

GenR Study

The Gen R Study is a population-based prospective cohort study from fetal life to young adulthood and was conducted in the city of Rotterdam, the Netherlands [15-18]. Gen R was designed for the long-term follow-up of a prenatally recruited multi-ethnic birth cohort to identify early environmental and genetic causes and also pathways leading to normal and abnormal growth, development and health in children The first 'Generation R' cohort is still ongoing after recruitment in 2001 (n=9749) and 'Generation R *Next'* has been launched in 2017 as the second cohort of the Gen R study.

2 Replication cohorts

The replication was performed in participants of European ancestry from the epidemiological investigations on chances of preventing recognizing early and optimally treating chronic diseases in an elderly population (ESTHER) study and the Latvia cohort with respectively 6112 and 3564 subjects.

ESTHER

The German ESTHER study is a population-based cohort study aimed at the prevention, early detection and optimized treatment of chronic diseases [19-21]. This ongoing study in the federal State of Saarland was conducted for the longitudinal follow up of residents

which were recruited at the age of 50-74 years during medical check-up in 2000-2002 (n=9940). A health check-up documentation was provided by the general practitioner and a standardized questionnaire regarding sociodemographic, lifestyle factors and medical history was collected from the study participant. Biological samples were obtained at baseline and follow-ups after 5, 8, 11, 14 and 17 years.

Latvia cohort

The Latvia cohort consists of a cross-sectional subset of adults from the National Latvian population in the period of 2008-2009 to study cardiovascular risk factors [22, 23]. A total of 4198 subjects agreed on participation, 4022 came for a visit, and 3807 completed the interview and allowed blood collection for biochemical and genetic analysis. Phenotypic data was stored in the Institute of Cardiology at the University of Latvia and genotyping data at the Latvian Biomedical Research and Study Centre (BMC). The latter maintains the Genome Database of the Latvian Population (LDGB), a national biobank that provides biologic specimens and associated phenotypic and clinical data for genetic and biomolecular research purposes [24].

3 Ethics

Informed consent for participation was obtained for all study subjects and approval was given by the Institutional Review Boards of the Erasmus University Medical Center Rotterdam (RS-I, RS-II, Gen R and healthy controls), University Medicine Greifswald (SHIP, SHIP-TREND), Boston University Medical Center (FHS), Johns Hopkins University (MESA), Northwestern University (MESA), Wake Forest University (MESA), University of California at Los Angeles (MESA), Columbia University (MESA), University of Minnesota (MESA), University of Heidelberg (ESTHER), Medical Association of Saarland (ESTHER) and the University of Latvia (Latvia cohort).

4 Stage 1: discovery

Different platforms for genome-wide genotyping were employed by discovery cohorts using standard procedures of the manufacturer. The Illumina platform was used by RS-I (HumanHap 550K (V.3) single and duo arrays), RS-II (HumanHap 550K (V.3) duo and Human 610K Quad arrays), SHIP-TREND (HumanOmni2.5-Quad array) and the GenR study (HumanHap 610K and 660K Quad arrays). The Affymetrix platform was employed by SHIP and MESA (both Genome-Wide Human SNP array 6.0) in addition to FHS (Human Mapping 500K plus 50K supplemental arrays). To aid meta-analysis, all datasets were imputed to the 1000 genome (1KG) dataset version 1v3 (with 30 million resulting SNP genotypes).

Genome-wide association analyses were performed in individual cohorts with adjustment for sex, age and study specific covariates.

5 Stage 2: replication

Genotyping data of the ESTHER cohort was generated on the Illumina platform (Infinium OncoArray-500K BeadChip) and imputed with 1000 genome (1KG) dataset version 1v3. *In silico* data for replication in the ESTHER cohort was available for seven out of eight SNPs including rs12233670, rs12985060, rs6107461, rs79710468, rs138776142, rs147900026, rs3905275, but not rs147174426. Individual SNP genotyping for four SNPs was applied in the Latvian cohort using a TagMan probe-based assay (Life Technologies, Carlsbad, CA). A total of four SNPs were available for replication (rs12233670, rs147174426, rs6107461 and rs147900026). The analyses were adjusted for age, sex and study specific covariates, and the replication results were included in the combined meta-analysis.

Supplementary Table S1 | Characteristics of discovery and replication cohorts

Study [a]	Country	Cohort	Total [b]	*H. pylori* IgG serology with cut-off [c]	
RSI	The Netherlands	Discovery	4771	Pyloriset EIA-G III[1]	≥20 U/mL
RSII	The Netherlands	Discovery	2112	Pyloriset EIA-G III[1]	≥20 U/mL
SHIP	Germany	Discovery	3830	Pyloriset EIA-G III[1]	≥20 U/mL
SHIP-TREND	Germany	Discovery	983	Pyloriset EIA-G III[1]	≥20 U/mL
FHS	United States	Discovery	3141	HM-CAP EIA[2]	EV > 2.2
MESA	United states	Discovery	447	*Is-H.pylori* IgG EIA[3]	IV ≥1.1
Gen R	The Netherlands	Discovery	401	Customized EIA[4]	ODR ≥1
ESTHER	Germany	Replication	6112	*H.pylori* Screening EIA[5]	>7.5 U
Latvia	Latvia	Replication	3564	*recom*Well Helicobacter IgG EIA[6]	>24 U/mL

[a] Study cohorts: Rotterdam Study I and II (RSI and II); Study of Health in Pomerania that consist of SHIP and SHIP-TREND; Framingham Heart Study (FHS); Multi-Ethnic Study of Atherosclerosis (MESA); Generation R study (GenR); Epidemiological investigations on chances of preventing recognizing early and optimally treating chronic diseases in an elderly population (ESTHER); Latvia cohort.

[b] Number of participants with *Helicobacter pylori* serology data combined with either GWAS or genotyping data

[c] Enzyme-linked immunosorbent assay used to measure IgG antibodies against *Helicobacter pylori*

[1] Orion Diagnostica, Espoo, Finland; expressed as units per milliliter (U/mL)

[2] Enteric Products Inc., Wesbury, NY; expressed as EIA value (EV)

[3] Diamedix Corporation, Miami, FL; expressed as index value (IV)

[4] Customized; expressed as optimal density ratio (ODR)

[5] Ravo Diagnostika, Freiburg, Germany, expressed as units (U)

[6] Mikrogen Diagnostik, Neuried, Germany, expressed in units per milliliter (U/mL)

Supplementary Table S2 | Summary of seven other single nucleotide polymorphisms (SNPs) in discovery, replication and combined meta-analysis

SNP					Discovery					Replication					Combined				
ID	**Chr**	**Position**	**Gene** [a]	**A½** [b]	**EAF**	**Beta** [c]	**SE**	***P*-value**	**I²**	**EAF**	**Beta** [c]	**SE**	***P*-value**	**I²**	**Beta** [c]	**SE**	**Dir** [d]	***P*-value**	**I²**
rs147174426	1	161528012	*FCGR3A*	A/T	0.93	0.480	0.094	$2.89x10^{-07}$	23.5	0.95	-0.146	0.123	0.237	0.0	0.250	0.075	+?-	$8.33x10^{-04}$	16.3
rs12985060	19	47145658	*DACT3*	T/C	0.28	0.182	0.036	$3.69x10^{-07}$	42.2	0.19	-0.148	0.074	0.044	0.0	0.118	0.032	+-?	$2.56x10^{-04}$	16.2
rs6107461	20	4412588	*ADRA1D*	T/C	0.04	0.338	0.071	$2.23x10^{-06}$	3.5	0.03	-0.102	0.101	0.313	26.5	0.193	0.058	+--	$9.09x10^{-04}$	14.0
rs79710468	5	155366973	*SGCD*	A/C	0.97	-0.519	0.110	$2.25x10^{-06}$	0.0	0.99	-0.158	0.270	0.558	0.0	-0.468	0.102	--?	$4.42x10^{-06}$	1.5
rs138776142	9	82093055	*TLE4*	A/T	0.01	0.677	0.145	$2.88x10^{-06}$	4.1	0.01	-0.202	0.228	0.375	0.0	0.424	0.122	+-?	$5.33x10^{-04}$	10.6
rs147900026	7	82468050	*PCLO*	T/G	0.99	-0.756	0.162	$3.10x10^{-06}$	0.0	0.99	0.381	0.286	0.183	0.0	-0.480	0.141	-++	$6.68x10^{-04}$	12.0
rs3905275	15	60833019	*RORA*	A/C	0.82	-0.170	0.037	$3.72x10^{-06}$	0.0	0.84	0.102	0.063	0.102	0.0	-0.100	0.032	-+?	$1.78x10^{-03}$	14.0

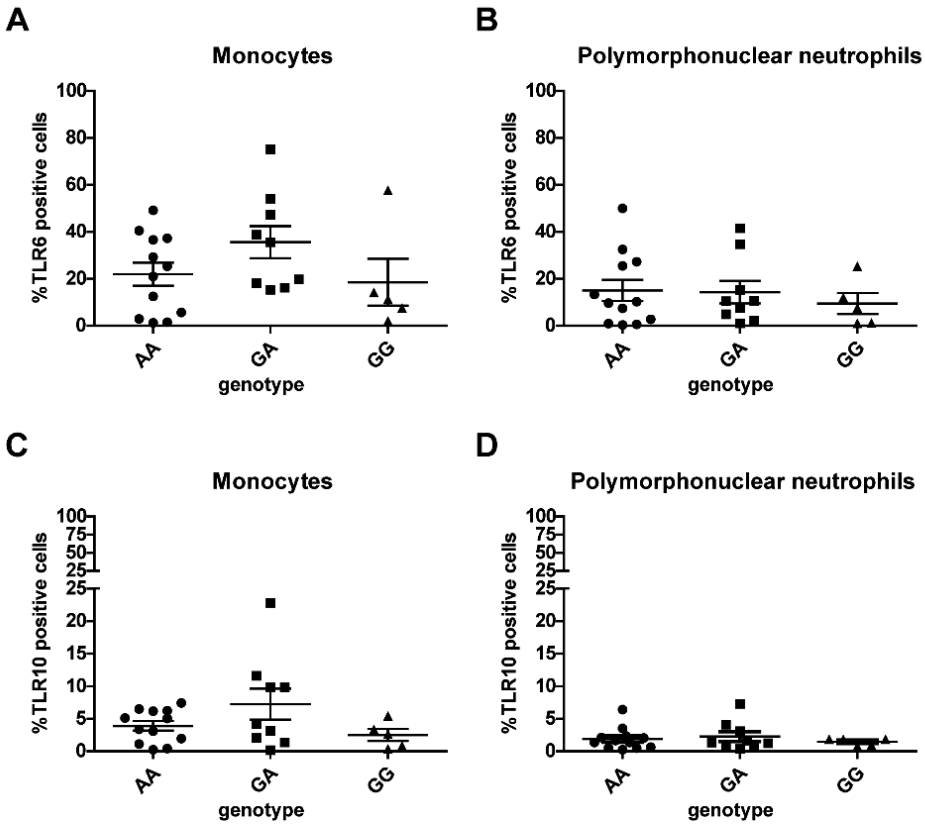
[a] Gene nearest to SNP.

[b] A½, effect allele 1 and the other allele 2.

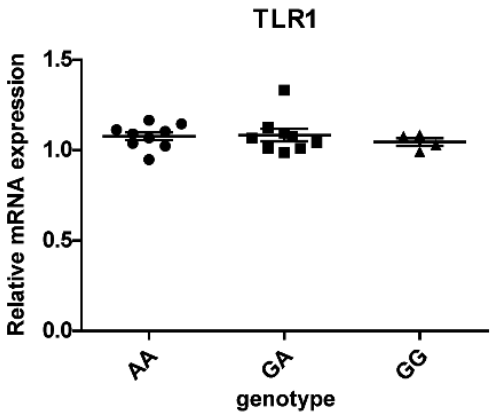
[c] effect size is relative to the allele 1.

[d] Direction of beta of respectively the discovery, ESTHER and Latvia cohorts: positive (+) or negative (-) or not determined (?).

Abbreviations: Chr, chromosome; EAF, effect allele frequency, I², measure of heterogeneity; SE, standard error.



▲Supplementary Figure S1 | Measurement of Toll-like receptor 6 (TLR6)- and TLR10- positive cells by flow cytometry. (A-D) Dot plots illustrating the percentage of monocytes and polymorphonuclear neutrophils positive for TLR6 (A-B) or TLR10 (C-D) in healthy subjects (n=26) genotyped for rs28393318: AA (n=12), GA (n=9) and GG (n=5). The mean ±SEM is shown. Dots indicate individual measurements.



▲Supplementary Figure S2 | Measurement of Toll-like receptor 6 (TLR6)- and TLR10- positive cells by flow cytometry. (A-D) Dot plots illustrating the percentage of monocytes and polymorphonuclear neutrophils positive for TLR6 (A-B) or TLR10 (C-D) in healthy subjects (n=26) genotyped for rs28393318: AA (n=12), GA (n=9) and GG (n=5). The mean ±SEM is shown. Dots indicate individual measurements.

Acknowledgment

The Rotterdam Study (RS) authors are grateful to Prof. dr. Bruno H Stricker and dr. Frank J van Rooij from the department of epidemiology at the Erasmus University Medical Center for providing pharmacological and serological data of RS study participants. The RS authors also acknowledge Wouter J. Den Hollander for his contribution at the initiation of the study.

The MESA authors thank the other investigators, the staff, and the participants of the MESA study for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

The Latvia cohort authors acknowledge the Genome Database of Latvian Population, the Latvian Biomedical Research and Study Centre for providing data and DNA samples.

Funding / Support

Cohort	Funding/support
RS-I RS-II	Netherlands Organisation of Scientific Research NWO Investments (nr.175.010.2005.011, 911-03-012) Institute Research Institute for Diseases in the Elderly (014-93-015; RIDE2) Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) (project nr. 050-060-810) Erasmus Medical Center Rotterdam Erasmus University Rotterdam Netherlands Organization for the Health Research and Development (ZonMw) Research Institute for Diseases in the Elderly (RIDE) Ministry of Education, Culture and Science Ministry for Health, Welfare and Sports European Commission (DG XII) Municipality of Rotterdam
GenR	Erasmus Medical Center Rotterdam Erasmus University Rotterdam Netherlands Organization for the Health Research and Development (ZonMw), including additional grant (ZonMw 907.00303, ZonMw 916.10159) Netherlands Organisation for Scientific Research (NWO) Ministry for Health, Welfare and Sports
SHIP SHIP-TREND	Deutsche Krebshilfe/ Dr. Mildred-Scheel-Stiftung (109102) Deutsche Forschungsgemeinschaft (DFG GRK840-D2/E3/E4, MA 4115/1-2/3) Federal Ministry of Education and Research (BMBF GANI-MED 03IS2061A and BMBF 0314107, 01ZZ9603, 01ZZ0103, 01ZZ0403, 03ZIK012) European Union (EU-FP-7: EPCTM and EU-FP7-REGPOT-2010-1) AstraZeneca (unrestricted grant) Federal Ministry of Education and Research Siemens Healthcare Federal State of Mecklenburg- West Pomerania University of Greifswald
FHS	The Framingham Heart Study is funded by National Institutes of Health contract N01-HC-25195, HHSN268201500001I, and 75N92019D00031 (Boston University). This project was funded in part by the Division of Intramural Research, National Heart, Lung, and Blood Institute (NHLBI), National Institutes of Health (NIH), Bethesda, MD. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the U.S. Department of Health and Human Services.
MESA	MESA and the MESA SHARe projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01HC95159, 75N92020D00005, N01HC95160, 75N92020D00002, N01HC95161, 75N92020D00003, N01HC95162, 75N92020D00006, N01HC95163, 75N92020D00004, N01HC95164, 75N92020D00007, N01 HC95165, N01HC95166, N01HC95167, N01HC95168, N01HC95169, UL1TR000040, UL1TR001079, and UL1TR001420. Funding for SHARe genotyping was provided by NHLBI Contract N02HL64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix GenomeWide Human SNP Array 6.0. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center.
ESTHER	State Ministry of Science, Research and Arts of Baden-Württemberg Federal Ministry of Education and Research Federal Ministry of Family Affairs, Senior Citizens, Women and Youth (Berlin, Germany)
Latvia	European Regional Development Fund (ERDF) (009/0220/1DP/1.1.1.2.0/09/ APIA/VIAA/016) National Program for Research in Latvia Ministry of Health (6-1396-2016) Fundamental and Applied Research Projects (FLPP) Programm in Latvia, project nr. lzp2018/10135

Supplementary references

1. Mayerle J, den Hoed CM, Schurmann C, Stolk L, Homuth G, Peters MJ *et al*. Identification of genetic loci associated with Helicobacter pylori serologic status. Jama, 2013. 309(18): p. 1912-20.
2. Hofman A, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A, Ikram MA *et al*. The Rotterdam Study: 2014 objectives and design update. Eur J Epidemiol, 2013. 28(11): p. 889-926.
3. Hofman A, Brusselle GG, Darwish Murad S, van Duijn CM, Franco OH, Goedegebure A *et al*. The Rotterdam Study: 2016 objectives and design update. Eur J Epidemiol, 2015. 30(8): p. 661-708.
4. Ikram MA, Brusselle GGO, Murad SD, van Duijn CM, Franco OH, Goedegebure A *et al*. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol, 2017. 32(9): p. 807-850.
5. John U, Greiner B, Hensel E, Ludemann J, Piek M, Sauer S *et al*. Study of Health In Pomerania (SHIP): a health examination survey in an east German region: objectives and design. Soz Praventivmed, 2001. 46(3): p. 186-94.
6. Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N *et al*. Cohort profile: the study of health in Pomerania. Int J Epidemiol, 2011. 40(2): p. 294-307.
7. Dawber TR, Meadors GF and Moore FE, Jr. Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health, 1951. 41(3): p. 279-81.
8. Feinleib M, Kannel WB, Garrison RJ, McNamara PM and Castelli WP. The Framingham Offspring Study. Design and preliminary data. Prev Med, 1975. 4(4): p. 518-25.
9. Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ *et al*. The Third Generation Cohort of the National Heart, Lung, and Blood Institute's Framingham Heart Study: design, recruitment, and initial examination. Am J Epidemiol, 2007. 165(11): p. 1328-35.
10. Mahmood SS, Levy D, Vasan RS and Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet, 2014. 383(9921): p. 999-1008.
11. Tsao CW and Vasan RS. Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. Int J Epidemiol, 2015. 44(6): p. 1800-13.
12. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR *et al*. Multi-Ethnic Study of Atherosclerosis: objectives and design. Am J Epidemiol, 2002. 156(9): p. 871-81.
13. Olson JL, Bild DE, Kronmal RA and Burke GL. Legacy of MESA. Glob Heart, 2016. 11(3): p. 269-274.
14. Burke G, Lima J, Wong ND and Narula J. The Multiethnic Study of Atherosclerosis. Glob Heart, 2016. 11(3): p. 267-268.
15. Jaddoe VW, van Duijn CM, van der Heijden AJ, Mackenbach JP, Moll HA, Steegers EA *et al*. The Generation R Study: design and cohort update 2010. Eur J Epidemiol, 2010. 25(11): p. 823-41.
16. Jaddoe VW, van Duijn CM, Franco OH, van der Heijden AJ, van Iizendoorn MH, de Jongste JC *et al*. The Generation R Study: design and cohort update 2012. Eur J Epidemiol, 2012. 27(9): p. 739-56.
17. Kruithof CJ, Kooijman MN, van Duijn CM, Franco OH, de Jongste JC, Klaver CC *et al*. The Generation R Study: Biobank update 2015. Eur J Epidemiol, 2014. 29(12): p. 911-27.
18. Kooijman MN, Kruithof CJ, van Duijn CM, Duijts L, Franco OH, van IMH *et al*. The Generation R Study: design and cohort update 2017. Eur J Epidemiol, 2016. 31(12): p. 1243-1264.
19. Low M, Stegmaier C, Ziegler H, Rothenbacher D, Brenner H, and study E. [Epidemiological investigations of the chances of preventing, recognizing early and optimally treating chronic diseases in an elderly population (ESTHER study)]Epidemiologische Studie zu Chancen der Verhutung, Fruherkennung und optimierten Therapie chronischer Erkrankungen in der alteren Bevolkerung (ESTHER-Studie). Dtsch Med Wochenschr, 2004. 129(49): p. 2643-7.

20. Raum E, Rothenbacher D, Low M, Stegmaier C, Ziegler H, and Brenner H. Changes of cardiovascular risk factors and their implications in subsequent birth cohorts of older adults in Germany: a life course approach. Eur J Cardiovasc Prev Rehabil, 2007. 14(6): p. 809-14.
21. Schottker B, Saum KU, Perna L, Ordonez-Mena JM, Holleczek B, and Brenner H. Is vitamin D deficiency a cause of increased morbidity and mortality at older age or simply an indicator of poor health? Eur J Epidemiol, 2014. 29(3): p. 199-210.
22. Erglis A, Dzerve V, Pahomova-Strautina J, Narbute I, Jegere S, Mintale I *et al*. A population-based cross-sectional study of cardiovascular risk factor in Latvia. Medicina (Kaunas), 2012. 48(6): p. 310-6.
23. Leja M, Shums Z, Nikitina-Zake L, Gavars M, Kikuste I, Milo J *et al*. Prevalence estimation of celiac disease in the general adult population of Latvia using serology and HLA genotyping. United European Gastroenterol J, 2015. 3(2): p. 190-9.
24. Rovite V, Wolff-Sagi Y, Zaharenko L, Nikitina-Zake L, Grens E, and Klovins J. Genome Database of the Latvian Population (LGDB): Design, Goals, and Primary Results. J Epidemiol, 2018. 28(8): p. 353-360.

11

CHAPTER 11

Summary and discussion

This thesis covers different topics regarding the gastro-intestinal (GI) microbiota in human health and disease. We have provided an overview of the bacteria in health and its role in oncogenesis in Chapter 2, underscoring the fine balance of the highly complex system of microbes. In contrast to the identification of *Helicobacter pylori* (*H. pylori*) as a well-known singe bacterial carcinogen for gastric cancer, multiple bacterial candidates have been proposed for other GI cancers in accumulating reports. Nevertheless, the International Cancer Microbiome Consortium stated that there is no direct proof for the human commensal microbiota as key drivers in the etiology of cancer in 2019 [1], besides the causal association between *H. pylori* and gastric cancer. The development of cancer and other diseases is indeed complex and multifactorial, but more insight into the microbial communities might bring us one step closer to unravelling their role in disease pathogenesis and to improvement of current diagnostic and treatment strategies. To comprehend the significance of microbial disturbances (dysbiosis) observed in GI diseases, it is paramount to map the microbial dynamics in the healthy state for comparison. In contrast to previous microbiome studies that collectively contributed to our knowledge of the GI microbiota within specific niches, our aim of Chapter 3 was to characterize the mucosal microbiome along the entire human GI tract within the same subjects. Patients undergoing double-balloon enteroscopy provided the best opportunity to gain concurrent access to nine different GI sites for downstream molecular analysis. We confirmed that the mucosal bacterial diversity and composition change along the GI tract and demonstrated a distinct bacterial profile in a subject diagnosed with a cecum tumor, reflecting the presence of bacterial dysbiosis in colorectal cancer (CRC). While disease status can significantly impact the GI microbial composition, also other factors (e.g. medication, diet, smoking habits) need to be considered as and collected as part of the metadata.

Mucosal biopsies are considered to be most representative for the microbiota at the human epithelium, but other resources might also be valuable for microbial research. In Chapters 4-5, we performed bacterial analyses using formalin-fixed paraffin embedded (FFPE) tissues from pathology archives and fecal immunochemical tests (FITs) from the CRC screening program, respectively. Despite these samples having a relatively low bacterial biomass and are more likely to be affected by bacterial contamination, the detection of specific bacterial markers was feasible using quantitative polymerase chain reaction (qPCR) analysis. On the other hand, we experienced that 16S rRNA amplicon sequencing of FFPE tissues remains difficult due to the extraction of deoxyribonucleic acid (DNA) of questionable quality in addition to the presence of bacterial contaminants. While a recent study declared that microbial analysis of FFPE tissues could be accomplished

by shotgun metagenomics, it should be underscored that microbial read counts were low after discarding human genomic reads [2]. Moreover, phylum Proteobacteria made up a substantial part of the bacterial microbiota (>25%) detected in FFPE colonic tissues [2], which was also noted in our 16S rRNA amplicon sequencing study. Since both Firmicutes and Bacteroidetes generally dominate the adult gut, these observations together suggest that the formalin fixation and embedding process might influence the microbiome profile of tissues [3]. Formalin causes crosslinking and chemical modifications of nucleic acids that limits the use of archived paraffin embedded tissue specimens for downstream molecular analysis in medical practice and research. On the other hand, it was reported that the application of formalin-free fixatives (i.e. PAXgene Tissue system, a non-crosslinking alcohol based tissue fixation and stabilization system) combined with storage at lower temperatures (4˚C or -20˚C) improves the retrieval and amplification of human genomic DNA and ribonucleic acids (RNA) from tissues that were paraffin embedded and stored up to nine years, and that results closely resembled those achieved with cryo-preserved samples irrespective from the amplicon length [4]. Histomorphology features of PAXgene fixed paraffin embedded (PFPE) tissues were essentially comparable to FFPE tissues [5, 6], suggesting that diagnostic assessment can still be performed without carcinogenic formalin for tissue fixation. The adaptation of current preservation and storage operations in pathology archives might not be logistically feasible in the short term, but high throughput sequencing projects including microbiome studies might still benefit from prospective PFPE tissue biobanks. Nevertheless, microbial DNA comprises only a small fraction of the total DNA extracted from paraffin embedded tissues and therefore we believe that molecular analysis of these and other low bacterial biomass sources requires critical appraisal of laboratory procedures and the inclusion of positive and negative controls for quality control.

The investigation of affordable and disease-specific biomarkers is vital to recognize (early) changes that influence our health, but the success of their application in clinical practice depends on the accuracy of the test as well as the ease to collect and analyze bio-samples. Stool-based testing of bacterial microbes is feasible using FITs from the CRC screening program (Chapter 5) and seems promising for the prediction of radiation enteropathy in future efforts [7] as described (Chapter 6). However, it should be noted that non-bacterial microbes have gained more attention in these recent years. With regard to CRC research, fecal shotgun metagenomics has provided new insights showing that tumors are characterized by a discriminating virome profile that also differs between early and late stages [8], a distinct mycobiome profile with higher Basidiomycota :

Ascomycota ratio (i.e. fugal dysbiosis) [9], as well as an altered gut archaea profile with the enrichment of halophiles and depletion of methanogens [10]. The diagnostic potential of the newly identified fecal metagenomic biomarkers is promising for the detection of CRC, but the overall performance to discriminate precancerous colorectal adenomas from healthy controls was less accurate. Although the diagnostic ability of the bacterial marker *Lachnoclostridium M3* was previously shown to be more convincing for colorectal adenoma detection [11], future studies might combine both bacterial and non-bacterial biomarkers to select the best panel for the targeted analysis of (pre)malignant lesions. Since stool samples have been replaced in current screening strategies, additional efforts are needed to validate the utility of candidate microbial biomarkers in FITs from subjects with different ethnic backgrounds. With regard to CRC screening in inflammatory bowel disease (IBD), the measurement of methylation levels of specific loci in stool was proposed by Kisiel *et al.* to identify patients at risk of cancer and high-grade dysplasia [12], as briefly discussed in Chapter 7. Perhaps the application of complementary biomarkers will aid the screening of specific patient groups at risk of cancer. While stool-based tests are more commonly employed to assess different aspects of GI health, it depends on accurate sampling of stool by patients, which could possibly influence microbial analysis. The fecal content within stool and in particular FIT containers represents only a minor part of a whole stool specimen, nevertheless it was previously shown that the degree of homogenization of stool did not alter the microbiome community composition [13]. A previous 16S rRNA amplicon study using the microbial content within FIT containers also determined that the bacterial community structure and membership is recapitulated when compared to stool [14]. Whether other lower abundant non-bacterial biomarkers are readily detectable in FIT fluids, remains to be investigated.

Meanwhile, the characterization of the GI microbiome and the search for biomarkers also continues for other diseases. 16S rRNA amplicon sequencing may show a glimpse of the individual members within bacterial communities, despite its accuracy of relying on the variable region of the *16S rRNA* gene amplified, the sequencing depth and the bioinformatic analysis workflow (e.g. pre-processing of data, operational taxonomic unit (OTU) picking, the reference database). In Chapter 8, we have demonstrated some interesting fecal bacterial microbes that were differentially abundant in the chronic inflammatory skin disease hidradenitis suppurativa (HS) using 16S rRNA amplicon sequencing. The enrichment of genus *Robinsoniella* in HS patients was an interesting finding, which may be relatively easily validated in a larger cohort using specific primers described in literature [15]. With only one small study reporting on the fecal microbiota in HS to date [16], future efforts are necessary

to investigate the role of the gut microbiota in HS and the possibility of a gut-skin axis in more depth. While HS could be part of the extra-intestinal manifestations observed in IBD, the common denominator attributing to these chronic inflammatory conditions is not fully understood. Whether the gut microbiota and/or byproducts trigger the immune system to cause inflammation or *vice versa* is an important question that is left to be answered. Microbial dysbiosis has been reported in IBD, including the depletion of *Faecalibacterium prausnitzii* and the enrichment of adherent-invasive *Escherichia coli* (AIEC) [17], but were not apparent in HS. In Chapter 9, we demonstrated that these bacterial markers as well as Paneth cell-related histological and genetic (i.e. *ATG16L1* single nucleotide polymorphism) markers were not able to distinguish Crohn's disease patients at risk of surgical recurrence following ileocecal resection. With the significant need to monitor disease activity and assist treatment strategies in the individual IBD patient [18], there is an urgent need for additional non-invasive biomarkers besides the commonly used C-reactive protein and fecal calprotectin. The selection of blood markers based on transcriptomic and proteomic analyses is promising, as shown by the development of a whole blood qPCR based classifier to predict the prognosis in newly diagnosed IBD patients [19], and the identification of protein markers associated with mid/long-term relapse in Crohn's disease patients following infliximab withdrawal [20]. However, none of these have made it to clinical practice yet. In the course of time, microbiome-based biomarkers might also serve as prediction tools for specific IBD subgroups.

In Chapter 10, we return to the most investigated microbe of the human GI tract, *H. pylori*. The association between the *Toll-like receptor 1* (*TLR1*) locus and anti-*H. pylori* IgG titers was re-assessed in an updated genome-wide association study (GWAS), but not uniformly replicated across the individual cohorts included in this study. Although the serological interpretation of IgG levels is complex and their correlation with GWAS results is not easy to unravel, we do believe that *H. pylori*-mediated immune cell activation is at least partly mediated via TLR1, the surface expression of which is influenced by genetic variation. While it is conceivable that other host genetic factors are involved in the biological processes triggered when *H. pylori* is encountered, much is yet to be learned about this bacterium, which is notorious for escaping the human immune system. Whether other, non-*H. pylori* residents of the stomach are involved in the initiation and progression of *H. pylori*-related diseases is a topic that has gained interest. As detailed in Chapter 2, microbial dysbiosis was previously described across different stages of gastric carcinogenesis (i.e. gastritis, intestinal metaplasia, gastric cancer) [21, 22], but a recent report also indicated considerable differences in microbial diversity and composition among gastric cancer microhabitats

(tumor, peritumor and normal tissues) [23]. While the higher presence of *H. pylori* significantly influenced the overall population structure within normal and peritumoral microhabitats, there is accumulating evidence that this bacterium disappears from the tumorigenic environment [23, 24]. These findings suggest that *H. pylori* seems to be involved in early pathogenic processes, potentially enabling other microbial candidates to thrive and take over in the changing gastric microenvironment. Meanwhile, the role of *H. pylori* in health seems controversial with paradoxical effects observed in extra-gastric diseases [25, 26]. *H. pylori* colonization was shown to reduce the risk of esophageal cancer [27-29], and additionally a negative association has been observed between *H. pylori* infection and both Crohn's disease and ulcerative colitis [30-32]. Although Helicobacteraceae species are detected in non-gastric parts of the GI tract as demonstrated in Chapter 3, it is thought that the protective effect in IBD is due to the immune tolerance-inducing properties of *H. pylori*, which favours gastric mucosal colonization and regulates systemic immune responses by inducing dendritic cells with tolerogenic phenotype and immunosuppressive regulatory T cells (Tregs) [33]. When taking the potential beneficial effects of this bacterium in mind, it was suggested that *H. pylori* eradication therapy on population-wide scale should be carefully reconsidered [30, 33].

Technological innovations in biomedical research have and will continue to provide us the opportunity to learn more about the significance of the GI microbiota in human health and disease. During the establishment of this thesis, scientific literature has been supplemented with numerous publications describing microbial signatures related to changes in our health. With the large amount of data available from high throughput sequencing, there is a need to translate this information for usage in medical practice. While relative abundances are generally reported in microbiome studies, quantifying absolute abundances of the microbiota enables the detection of potential drivers of changes in microbial dynamics over time [34]. Alpha- and β diversity metrics are also commonly computed to describe the microbial diversity within- and between samples, but are less practical to apply in diagnostic, prediction and monitoring strategies for individual patients. Biomarker discovery enables the identification of metagenomic markers that are possibly relevant in disease pathogenesis and offers an opportunity to establish a distinctive panel of metagenomic markers to measure with less complex technologies. A (q)PCR based approach could ease the concurrent assessment of samples at large scale, but the clinical cut-off needs to be carefully defined for individual candidate biomarkers. A multiplex (q)PCR assay might be convenient to detect several microbial targets within one single run, but the amplification efficacy and the reproducibility needs to be considered when choosing targets of varying

microbial abundance. Although microbial dysbiosis is commonly characterized by both the enrichment and depletion of individual microbes, it is still uncertain whether repeated targeted analysis of these biomarkers allow monitoring of health changes over time. The collection of longitudinal samples and functional experiments remain necessary to proof causality in disease pathogenesis and to unravel the molecular mechanisms involved, although functional profiles could be predicted using shotgun metagenomics. With the trend to simplify the sample collection method to comply for patient convenience, it could be more challenging to conduct comprehensive analyses with less material as exemplified by the substitution of stool for FIT specimens.

While the restoration of microbial dysbiosis seems an attractive strategy to change the disease course in patients, it will be challenging to set up personalized treatments in future therapeutic efforts. The administration of probiotics to supplement of depleted microbiota might be helpful to restore homeostasis, but also a change in dietary habits and the use of prebiotics might be helpful to promote the growth of beneficial microbiota. The targeted elimination of potentially harmful agents is more complex with anti-microbial treatments such as antibiotics that are not selective. Fecal microbiota transplantation (FMT) with healthy donor stool has been applied in patients with *Clostridium difficile* infection, but it remains to be determined whether this therapy is beneficial for other GI conditions. Its potential role in IBD treatment is currently much explored, but mixed results are obtained, and may depend on stool quality, disease entity and severity of disease [35]. Since the efficacy of FMT seems inherent to the ability of the donor's stool to restore the state of dysbiosis, the adjustment of therapy regimens for the individual patient has been advocated. With the scientific community so highly committed to unravel the role of microbiota in disease pathogenesis, it is likely that patients will benefit from these efforts in the future practice by personalized medicine. The administration of stool from the most compatible donor or the treatment with selected microbiota might be possible for the individual patient. For future perspectives, the integration of data from metagenomics, metabolomics, transcriptomics and proteomics is promising to develop new diagnostic, prediction, treatment and prevention strategies. A close collaboration between different disciplines together with the implementation of artificial intelligence might overcome the challenges in health care for the best interest of the patient. The identification of specific biological patterns based on big data profiles enables the recognition of patients at risk for disease development or progression and guide treatment strategies more efficiently. Moreover, the application of genetic data could allow the identification of genetically predisposed subjects. Despite the complexity of big data driven science, it is a very important tool to aid personalized

medicine. The investigation and incorporation of microbial profiles might contribute to individualized strategies for clinical conditions such as metabolic syndrome [36], but the inclusion of different data types and points might be even more promising. For instance, personized therapy (diet) based on an machine-learning algorithm has previously shown to improve multiple aspects of glucose metabolism by integrating different parameters including blood, dietary habits, anthropometrics besides the gut microbiota [37]. With the growing collection of personalized data on large scale, the interpretation and translation of this valuable information will eventually allow to the best care for the individual patient. Since big data is making its way into clinical practice, physicians and biomedical researchers have to work more closely together in the near future.

References

1. Scott AJ, Alexander JL, Merrifield CA, Cunningham D, Jobin C, Brown R *et al*. International Cancer Microbiome Consortium consensus statement on the role of the human microbiome in carcinogenesis. Gut, 2019. 68(9): p. 1624-1632.
2. Debesa-Tur G, Pérez-Brocal V, Ruiz-Ruiz S, Castillejo A, Latorre A, Soto JL *et al*. Metagenomic analysis of formalin-fixed paraffin-embedded tumor and normal mucosa reveals differences in the microbiome of colorectal cancer patients. Sci Rep, 2021. 11(1): p. 391.
3. Pinto-Ribeiro I, Ferreira RM, Pereira-Marques J, Pinto V, Macedo G, Carneiro F *et al*. Evaluation of the Use of Formalin-Fixed and Paraffin-Embedded Archive Gastric Tissues for Microbiota Characterization Using Next-Generation Sequencing. Int J Mol Sci, 2020. 21(3).
4. Groelz D, Viertler C, Pabst D, Dettmann N and Zatloukal K. Impact of storage conditions on the quality of nucleic acids in paraffin embedded tissues. PLoS One, 2018. 13(9): p. e0203608.
5. Kap M, Smedts F, Oosterhuis W, Winther R, Christensen N, Reischauer B *et al*. Histological assessment of PAXgene tissue fixation and stabilization reagents. PLoS One, 2011. 6(11): p. e27704.
6. Groelz D, Sobin L, Branton P, Compton C, Wyrich R, and Rainen L. Non-formalin fixative versus formalin-fixed tissue: a comparison of histology and RNA quality. Exp Mol Pathol, 2013. 94(1): p. 188-94.
7. Reis Ferreira M, Andreyev HJN, Mohammed K, Truelove L, Gowan SM, Li J *et al*. Microbiota- and Radiotherapy-Induced Gastrointestinal Side-Effects (MARS) Study: A Large Pilot Study of the Microbiome in Acute and Late-Radiation Enteropathy. Clin Cancer Res, 2019. 25(21): p. 6487-6500.
8. Nakatsu G, Zhou H, Wu WKK, Wong SH, Coker OO, Dai Z *et al*. Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. Gastroenterology, 2018. 155(2): p. 529-541 e5.
9. Coker OO, Nakatsu G, Dai RZ, Wu WKK, Wong SH, Ng SC *et al*. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. Gut, 2019. 68(4): p. 654-662.
10. Coker OO, Wu WKK, Wong SH, Sung JJY and Yu J. Altered Gut Archaea Composition and Interaction With Bacteria Are Associated With Colorectal Cancer. Gastroenterology, 2020. 159(4): p. 1459-1470 e5.
11. Liang JQ, Li T, Nakatsu G, Chen YX, Yau TO, Chu E *et al*. A novel faecal Lachnoclostridium marker for the non-invasive diagnosis of colorectal adenoma and cancer. Gut, 2020. 69(7): p. 1248-1257.
12. Kisiel JB, Klepp P, Allawi HT, Taylor WR, Giakoumopoulos M, Sander T *et al*. Analysis of DNA Methylation at Specific Loci in Stool Samples Detects Colorectal Cancer and High-Grade Dysplasia in Patients With Inflammatory Bowel Disease. Clin Gastroenterol Hepatol, 2019. 17(5): p. 914-921 e5.
13. Santiago A, Panda S, Mengels G, Martinez X, Azpiroz F, Dore J *et al*. Processing faecal samples: a step forward for standards in microbial community analysis. BMC Microbiol, 2014. 14: p. 112.
14. Baxter NT, Koumpouras CC, Rogers MA, Ruffin MTt and Schloss PD. DNA from fecal immunochemical test can replace stool for detection of colonic lesions using a microbiota-based model. Microbiome, 2016. 4(1): p. 59.
15. Whitehead TR, Anoma C and McLaughlin RW. Rapid identification of Robinsoniella peoriensis using specific 16S rRNA gene PCR primers. Anaerobe, 2017. 43: p. 39-42.
16. Kam S, Collard M, Lam J and Alani RM. Gut Microbiome Perturbations in Patients with Hidradenitis Suppurativa: A Case Series. J Invest Dermatol, 2021. 141(1): p. 225-228 e2.
17. Kostic AD, Xavier RJ and Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. Gastroenterology, 2014. 146(6): p. 1489-99.
18. Sands BE. Biomarkers of Inflammation in Inflammatory Bowel Disease. Gastroenterology, 2015. 149(5): p. 1275-1285 e2.

19. Biasci D, Lee JC, Noor NM, Pombal DR, Hou M, Lewis N *et al*. A blood-based prognostic biomarker in IBD. Gut, 2019. 68(8): p. 1386-1395.
20. Pierre N, Baiwir D, Huynh-Thu VA, Mazzucchelli G, Smargiasso N, De Pauw E *et al*. Discovery of biomarker candidates associated with the risk of short-term and mid/long-term relapse after infliximab withdrawal in Crohn's patients: a proteomics-based study. Gut, 2020.
21. Coker OO, Dai Z, Nie Y, Zhao G, Cao L, Nakatsu G *et al*. Mucosal microbiome dysbiosis in gastric carcinogenesis. Gut, 2018. 67(6): p. 1024-1032.
22. Aviles-Jimenez F, Vazquez-Jimenez F, Medrano-Guzman R, Mantilla A and Torres J. Stomach microbiota composition varies between patients with non-atrophic gastritis and patients with intestinal type of gastric cancer. Sci Rep, 2014. 4: p. 4202.
23. 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: p. 336-348.
24. Seo I, Jha BK, Suh S-I, Suh M-H and Baek W-K. Microbial Profile of the Stomach: Comparison between Normal Mucosa and Cancer Tissue in the Same Patient. jbv, 2014. 44(2): p. 162-169.
25. Bravo D, Hoare A, Soto C, Valenzuela MA and Quest AF. Helicobacter pylori in human health and disease: Mechanisms for local gastric and systemic effects. World J Gastroenterol, 2018. 24(28): p. 3071-3089.
26. Gravina AG, Priadko K, Ciamarra P, Granata L, Facchiano A, Miranda A *et al*. Extra-Gastric Manifestations of Helicobacter pylori Infection. J Clin Med, 2020. 9(12).
27. Anderson LA, Murphy SJ, Johnston BT, Watson RG, Ferguson HR, Bamford KB *et al*. Relationship between Helicobacter pylori infection and gastric atrophy and the stages of the oesophageal inflammation, metaplasia, adenocarcinoma sequence: results from the FINBAR case-control study. Gut, 2008. 57(6): p. 734-9.
28. Xie FJ, Zhang YP, Zheng QQ, Jin HC, Wang FL, Chen M *et al*. Helicobacter pylori infection and esophageal cancer risk: an updated meta-analysis. World J Gastroenterol, 2013. 19(36): p. 6098-107.
29. Vohlonen IJ, Hakama M, Härkönen M, Malila N, Pukkala E, Koistinen V *et al*. Oesophageal cancer incidence in 20-year follow-up in a population-based sample of 12 000 middle-age men with or without Helicobacter pylori infection in Finland. Gut, 2018. 67(6): p. 1201-1202.
30. Castaño-Rodríguez N, Kaakoush NO, Lee WS and Mitchell HM. Dual role of Helicobacter and Campylobacter species in IBD: a systematic review and meta-analysis. Gut, 2017. 66(2): p. 235-249.
31. Shirzad-Aski H, Besharat S, Kienesberger S, Sohrabi A, Roshandel G, Amiriani T *et al*. Association Between Helicobacter pylori Colonization and Inflammatory Bowel Disease: A Systematic Review and Meta-Analysis. J Clin Gastroenterol, 2020.
32. Ding ZH, Xu XP, Wang TR, Liang X, Ran ZH, and Lu H. The prevalence of Helicobacter pylori infection in inflammatory bowel disease in China: A case-control study. PLoS One, 2021. 16(3): p. e0248427.
33. Yu Y, Zhu S, Li P, Min L and Zhang S. Helicobacter pylori infection and inflammatory bowel disease: a crosstalk between upper and lower digestive tract. Cell Death Dis, 2018. 9(10): p. 961.
34. Rao C, Coyte KZ, Bainter W, Geha RS, Martin CR, and Rakoff-Nahoum S. Multi-kingdom ecological drivers of microbiota assembly in preterm infants. Nature, 2021.
35. Wilson BC, Vatanen T, Cutfield WS and O'Sullivan JM. The Super-Donor Phenomenon in Fecal Microbiota Transplantation. Front Cell Infect Microbiol, 2019. 9: p. 2.
36. Shapiro H, Suez J and Elinav E. Personalized microbiome-based approaches to metabolic syndrome management and prevention. J Diabetes, 2017. 9(3): p. 226-236.
37. Zeevi D, Korem T, Zmora N, Israeli D, Rothschild D, Weinberger A *et al*. Personalized Nutrition by Prediction of Glycemic Responses. Cell, 2015. 163(5): p. 1079-1094.

11

12

CHAPTER 12

Dutch summary

Nederlandse samenvatting

Technologische ontwikkelingen binnen het microbiologisch onderzoek hebben het mogelijk gemaakt om de gastro-intestinale (GI) micro-organismen (microbiota) en hun genoom (microbioom) te onderzoeken in ziekte en gezondheid. Hoewel virussen, schimmels en archaea onderdeel zijn van het complexe netwerk van microben, vormen de bacteriën de grootste groep microben binnen het maag-darm kanaal en staan derhalve centraal in veel studies. Metagenoom en 16S rRNA amplicon sequencing hebben geleid toch meer inzicht in de microbiële samenstelling van het maag-darm kanaal, maar de betekenis van microbiële verstoringen (dysbiosis) dient verder te worden onderzocht. Hoewel een samenspel van microben aannemelijk is in de verschillende fasen van de pathogenese, dient de individuele bijdrage te worden vastgesteld. Aangezien de GI microbiota potentieel betrokken zijn bij processen buiten het maag-darm kanaal, mogelijk indirect via microbiële producten, is verdere evaluatie uiterst interessant.

Voor het onderzoek naar de GI microbiota wordt veelal gebruik gemaakt van verse of ingevroren biopten en feces monsters. Hoewel de mucosale microben een nauwere interactie hebben met humane epitheelcellen, is feces makkelijker te verkrijgen dan biopten, en derhalve biedt feces bij uitstek een mogelijkheid om niet-invasieve microbiële biomarkers te identificeren. Feces immunochemische testen (FITs) worden heden toegepast in het bevolkingsonderzoek naar darmkanker, maar niet alle individuen met (pre)maligne afwijkingen worden hiermee opgespoord. De toepassing van het microbioom als additionele non-invasieve biomarker is veelbelovend, maar het gebruik van hoogwaardige sequencing technieken is arbeidsintensief en niet kosteneffectief voor screening op grote schaal. Daarentegen is het doelgericht analyseren van selecte metagenoom markers in feces monsters technisch eenvoudiger uitvoerbaar, maar is het de vraag of dit op termijn kan plaats vinden in FIT monsters in plaats van feces. De zoektocht naar de meest geschikte microbiële taxonomische markers voor de verschillende stadia van ziekte is gaande en behoeft validatie in studiepopulaties die verschillen in etnische achtergrond en comorbiditeit. Aangezien omgevingsinvloeden en genetische factoren mede bijdragen aan de vatbaarheid voor ziekten, zullen ook deze componenten relevant zijn voor toekomstige onderzoeksprojecten

Het belang van de GI microbiota in ziekte en gezondheid is in de afgelopen jaren evident geworden en heden zijn er veel verschillende aandoeningen geassocieerd met veranderingen in de microbiële samenstelling. In hoofdstuk 2 geven wij een uitgebreide beschrijving van de GI microbiota in relatie tot gezondheid en haar rol in oncogenese. Ons literatuuroverzicht laat zien dat verschillende studies tezamen een bijdrage hebben

geleverd aan onze kennis omtrent de bacteriële microben in de gezonde populatie, maar veelal richtten studies zich slechts op één segment van het maag-darmkanaal. In hoofdstuk 3 onderzoeken wij de bacteriële dynamiek tussen negen mucosale locaties binnen het maag-darm kanaal van veertien individuen middels 16S rRNA amplicon sequencing van biopten die verkregen waren bij dubbel ballon endoscopie. Wij tonen aan dat in het onderste deel van het maag-darm kanaal zowel de bacteriële dichtheid als microbiële diversiteit per locatie hoger is dan in het bovenste deel, en dat de bacteriële compositie verschilt van het bovenste deel van het maag-darmkanaal. Hoewel vers of ingevroren materiaal (biopten en feces) doorgaans gebruikt wordt om microbe-gerelateerde gezondheidsvraagstukken te beantwoorden, kan het pathologisch-anatomisch archief een uitkomst bieden voor onderzoek dat specifieke anatomische locaties, zeldzame ziektebeelden of longitudinale follow-up vereist. Het gebruik van formaline-gefixeerd en in paraffine ingebed (FFPE) weefsel is niet eenvoudig doordat het desoxyribonucleïnezuur (DNA) in dit materiaal doorgaans van slechtere kwaliteit is en doordat de bacteriële biomassa relatief laag is in dit type monsters. Echter, in sommige gevallen is dit het enig beschikbare materiaal voor microbieel onderzoek. In hoofdstuk 4 demonstreren wij dat bacteriële analyse van FFPE monsters onder specifieke omstandigheden mogelijk is, maar dat 16S rRNA amplicon sequencing voor alsnog uitdagend blijft. Onze aanbevelingen om de kwaliteit te waarborgen en bacteriële contaminatie te onderkennen kunnen van belang zijn voor toekomstig onderzoek met lage biomassa monsters. In hoofdstuk 5 tonen wij aan dat specifieke bacteriële markers ook betrouwbaar gemeten kunnen worden in FITs afkomstig uit het bevolkingsonderzoek naar darmkanker, en dat bacteriële contaminatie minimaal lijkt te zijn zoals blijkt uit kwantitatieve polymerase ketting reactie (PCR) bepalingen. Daarmee is de toepassing van FITs voor het analyseren van het GI microbioom en detectie van (pre)maligne afwijkingen veelbelovend. Daarnaast is de klinische toepassing van het microbioom tevens hoopgevend in radiatie-enteritis (RE) volgens de studie van Ferreira en collega's (Clin Cancer Res, 2019). In hoofdstuk 6 bespreken wij hun bevindingen in het licht van recente literatuur en beschrijven wij de parallellen met inflammatoire darmziekte (IBD), welke geassocieerd is met een veranderde microbiële compositie alsook een verhoogd risico op darmkanker. Naast de mogelijkheid om microbieel DNA uit feces te isoleren, blijkt ook humaan DNA bruikbaar voor de meting van specifieke genetische loci (zoals *BMP3* en *VAV3* genen). Deze ontwikkeling is gunstig voor de screening van hooggradige dysplasie en darmkanker in IBD patiënten, maar in hoofdstuk 7 bespreken wij dat het effect van medicatie op dergelijke diagnostische metingen in acht moet worden genomen. IBD is geassocieerd met extra-intestinale ongemakken, zoals psoriasis en hidradenitis suppurativa (HS), en recente hypotheses suggereren dat de GI microbiota

ook een bijdrage kunnen leveren aan processen buiten het maag-darm kanaal. Gezien de mogelijkheid dat de GI microbiota binnen het concept van een darm-huid-as (i.e. gut-skin axis) een rol zouden kunnen vervullen in dermatologische aandoeningen, onderzoeken wij als één van de eersten het fecale microbioom in HS en demonsteren wij in hoofdstuk 8 een aantal interessante bacteriële veranderingen in patiënten ten op zichte van gezonde individuen. In hoofdstuk 9 richten wij ons verder op het intestinale microbioom bij patiënten met IBD. Voor de IBD-subgroep met de ziekte van Crohn (CD) kan het verwijderen van de ontstoken darm middels een ileocoecaal resectie een succesvolle therapie zijn, maar bestaat er alsnog de kans op een recidief. In hoofdstuk 9 gebruiken wij FFPE weefsel resectie preparaten om te onderzoeken of microbiële biomarkers kunnen bijdragen aan de predictie van CD patiënten die risico lopen op een postoperatief recidief. Wij laten zien dat de bacteriële markers (*Faecalibacterium prausnitzii* en adherent invasieve *Escherichia coli* [AIEC]) niet in staat zijn om individuen met en zonder re-resectie retrospectief te onderscheiden. Genetische factoren (i.e. genetische varianten in het autofagie gen *ATG16L1*) kunnen middels hun effect op de Paneth cel functie de capaciteit van de gastheer om bacteriën te verwerken beïnvloeden. Desondanks waren zowel histopathologische markers voor Paneth cel functie als de *ATG16L1* status niet voorspellend voor een recidief en derhalve zal de zoektocht naar geschikte biomarkers verder moeten worden doorgezet. In hoofdstuk 10 verplaatsen wij onze aandacht naar de rol van genetische factoren in de interactie tussen het human afweersysteem en *Helicobacter pylori* (*H. pylori*), een welbekende risicofactor voor maagkanker. Wij onderzochten de associatie tussen de genetische variatie van het *Toll-Like receptor 1* (*TLR1*) gen en anti-*H. pylori* antilichaam spiegels, welke tot heden niet uniform bevestigd kon worden na de initiële rapportage door Mayerle en collega's (JAMA, 2013). Wij concluderen dat antilichaam verval en TLR1-onafhankelijke effecten van *H. pylori* mogelijk genetische associatie studies kunnen bemoeilijken. Samenvattend laten de studies in dit proefschrift zien dat de interactie tussen het humane lichaam en bacteriën mogelijk een rol kan spelen in het ontstaan van intra en extra-intestinale ziektebeelden, dat er verschillende biomaterialen zijn die mogelijk gebruikt kunnen worden in het onderzoek naar deze associaties, maar dat technische limitaties dit type onderzoek kunnen bemoeilijken. Hoewel standaardisering van onderzoeksmethoden dus essentieel is, blijft het interessant om de rol van microben in diagnostiek en behandeling van verschillende ziektebeelden verder te onderzoeken.

12

APPENDIX

Acknowledgments

Publications

PhD portfolio

Curriculum Vitae

Acknowledgments

To **professor Peppelenbosch** – bedankt dat u mij de kans hebt gegeven om onder uw supervisie op het laboratorium te mogen werken. Het was een leuke en interessante tijd met u als grote leermeester van onze onderzoeksgroep. Ik heb veel van u mogen leren tijdens de wekelijkse vergaderingen op maandag waarbij u met veel enthousiasme ons vragenvuur beantwoordde en ons voorzag van een nieuwe "to-do list" aan experimenten voor de rest van de week.

To **Gwenny** – bedankt dat je de afgelopen jaren mijn directe supervisor wilde zijn. Ik prijs jouw onuitputtelijke energie en enthousiasme om alle studenten binnen onze onderzoeksgroep te blijven motiveren en te helpen met uiteenlopende projecten. Ik waardeer dat de deuren van jouw kantoor altijd open stonden voor spontaan overleg en dat je altijd fanatiek was in het beantwoorden van e-mails, het meedenken met experimenten, het meepuzzelen met data, het vormgeven van nieuwe figuren, het corrigeren van manuscripten, het regelen van formulieren, het toesturen van literatuur, het bijwonen van ingewikkelde meetings enzovoorts. Ik zal nooit vergeten dat ik zo veel van je heb mogen leren en deel mocht uitmaken van "groep Gwenny".

To **Leonie** – bedankt voor alle steun, gezelligheid en alle leuke momenten in de afgelopen jaren. Ik waardeer je vriendelijkheid en laagdrempeligheid om mij met van alles en nog wat te helpen op het lab. Ook fijn dat je mij altijd nog extra tijd gaf om de primers te selecteren zodat ook mijn bestelling nog meekon met de wekelijkse ronde.

To **Zhouhong** – Thanks for your friendship. Although you belonged to another research group, you were always available to help and to support me at the lab. I'm pleased that I have met you and that we have spent some fantastic years together at work. Also, many thanks to Jerry for the nice time outside working times.

To **Alessandra** and **Sharida** – Thank you for your patience and support during your time at the lab. We have had a lot of fun together and our friendship will for sure continue beyond this journey.

To **Xiaopei** – my office and group buddy – Thank you for your kindness and help during the past years. You were always the most calm and patient person of our group. It was nice to hang out with you in and outside the lab and we will for sure keep in touch with each other while you're in the Netherlands, but also when you will return to China.

To **Changbo** – my travel buddy – Thank you for your help and support. You always "put your happy face on" to remind everyone to stay positive in life. It was nice to talk, to drink and share food with you during the train journeys. Many thanks and all the best in life!

To **Yang** – my office buddy – Thank you for being my neighbor. You were always prepared to help me around the lab and you never said "no" to anything. Thanks for the free lessons to upgrade my mandarin and the tasty food that you brought to the lab. I wish a you and your family a very bright future!

To **Thijmen** – kantoor maatje – Bedankt voor alles wat je voor mij hebt gedaan van het begin tot het eind. Je was altijd vrolijk en zorgde voor een leuke en gezellige sfeer op het lab, maar was vooral ook bereid om mee te denken tijdens ups-and-downs.

To **Jiaye** and **Pengfei** – Thank you both for your help and support. You were always available to help me out and to give me constructive advice. It was nice to hang around with you at the office. Best wishes to both of you!

To **Bingting, Celio**, **Shihao** – It was a pleasure to meet you. It was quite interesting to be around all three of you guys at the lab. We have laughed a lot in the many months together. Keep smiling and good luck with everything in life!

To **YingYing** and **Wen** – my office buddies – Thank you very much for your support during your time at the lab as senior PhD students. I could always ask you for help without any hesitation. Thanks for the nice time that we have spent together in and outside the lab. YingYing - also special thanks to your family for having me as guest at your home.

To **Michiel, Lauke** and **Vincent** – het hoog associatief denkende trio – bedankt voor jullie wijsheid en hulp tijdens de afgelopen jaren. Het was uiteraard ook gezellig met jullie op het lab en wellicht durven jullie in de toekomst nog een uitdagend potje "bierpong" aan, en in geval van Michiel ook nog potje patience/pesten/ jeu de boules.

To **Kateryna** – my group buddy – Thank you for being the happy smiling member of our group. It was a pleasure to spent time with you in and outside the lab. You always brought us to some nice places to hang out. Wish you all the good things in life!

To **Elisa** – my group buddy – Thank you for your kindness and generosity. I will never forget the many food parties that you organized, especially the special Italian dinner. Wish you a bright future in science and all the best to you and your family!

To **Lisanne** and **Patrick** – bedankt voor alle hulp en gezelligheid op het lab. Hoewel ik geen deel van jullie groep uitmaakte, heb ik een enorm leuke tijd met jullie gehad tijdens de vele feestjes en etentjes. Heel veel succes en geluk toegewenst in de toekomst!

To **Estella** – Thank you for your help and kindness in the past years. It was always nice to talk to you and to have your support as senior PhD student. All the best wishes to you and your family!

To **Sunrui**, **Shan**, **Zhijiang** – Thank you for support at the lab. You were always friendly and available to help me around. I also appreciate your generosity, because you always bring many gifts from China. I wish you good health and a very bright future!

To **Wenhui**, **Wanlu**, **Meng**, **Ling** – Most of you were already senior PhD students when I started, but thank you for your help and kindness. I wish you all the best in the future.

To **Junhong** – my office buddy – Thank you for all the brainstorm sessions and nice conversations at the office. I have learned some important lessons from you. Take care and all the best to you and your family!

To **Pauline** – Thank you for sharing your knowledge and skills with me when I just started at the lab. I was a pleasure to meet you. I wish you and your family all the good things in life.

To **Rebecca** – Bedankt voor je harde inzet tijdens de stage. Het was gezellig om samen met jou te experimenteren op het lab. Ik wens je het allerbeste in de toekomst!

To **Yoena** – Bedankt voor alle nuttige besprekingen en de gezelligheid op kantoor. Je werkte altijd keihard door en met jouw doorzettingsvermogen zal je heel ver komen.

To other students – **Mehjabeen, Marith, Michelle, Bradley, Darya, Leonie** – good luck with all of your careers!

To my other Chinese colleagues – **Buyun, Wenshi, Peifa, Yunlong, Ruyi, Pengyu, Shaojun, Yining, Shanshan** – Thank you for all your help in these past years. Good luck with everything in life!

To all others from the laboratory: **Rattana, Stefano, Rachid, Robbie, Gülce, Lucia, Amy, Valeska, Anthonie, Gertine, Ron, Marla, Aafke, Monique, Diahann, Jaap, Jan, Hanneke, Auke, Greta, Sonja, Frances, Buddy, Martine, Natasha, Kelly, Hugo, Marcel, André, Abdullah** and the group of surgery [**Ivo, Henk, Ruby, Floris, Jorke, Gilles, Eric, Petra, Luc, Monique**] – it was a pleasure to meet you all. Thank you, very much and best wishes, to all.

To **Nancy** – special thanks to a very kind, patient and hardworking collaborator. It was a pleasure to work with you. Good luck with finishing your PhD.

To **other co-authors** – Thank you very much for your efforts in our collaboration projects.

To **professor Yu, Steffi** and **Celia** – Thank you very much for your hospitality. I will never forget your warm welcomes during work-related and private visits to HK.

To Fanny, Evelien, Jeanine and other clinical PhD students – bedankt en het allerbeste in de toekomst toegewenst.

To **Marie-Christine / Aáron-Christiaan, Anesh, Charita, Kwani, Maartje, Nita, Thelma, Samina, Tamara, Nina, Inge, Diane e.a.** – bedankt voor jullie vriendschap en onuitputtelijke steun in de afgelopen jaren.

And special thanks to my family in the Netherlands and abroad – 最後感謝我的家人和在荷蘭, 日本, 香港及加拿大的親戚:是你們的鼓勵和支持給了我堅持下去的勇氣和信心。我想把這篇畢業論文獻給你們和在遙遠的祖父母和外祖父母!

List of publications

Lam SY, Yu J, Wong SH, Peppelenbosch MP and Fuhler GM. The gastrointestinal microbiota and its role in oncogenesis. Best Pract Res Clin Gastroenterol, 2017. 31(6): p. 607-618.

Vuik F, Dicksved J, Lam SY, Fuhler GM, van der Laan L, van de Winkel A *et al*. Composition of the mucosa-associated microbiota along the entire gastrointestinal tract of human individuals. United European Gastroenterol J, 2019. 7(7): p. 897-907.

Grobbee EJ, Lam SY, Fuhler GM, Blakaj B, Konstantinov SR, Bruno MJ *et al*. First steps towards combining faecal immunochemical testing with the gut microbiome in colorectal cancer screening. United European Gastroenterol J, 2020. 8(3): p. 293-302.

Lam SY, Peppelenbosch MP and Fuhler GM. Prediction and Treatment of Radiation Enteropathy: Can Intestinal Bugs Lead the Way? Clin Cancer Res, 2019. 25(21): p. 6280-6282.

Lam SY, Yu J and Peppelenbosch MP. Value of VAV3 Methylation in Stool DNA Might Be Restricted to Non-Thiopurine-Treated Inflammatory Bowel Disease Patients. Clin Gastroenterol Hepatol, 2020. 18(2): p. 520.

Lam SY, Radjabzadeh D, Eppinga H, Nossent YRA, van der Zee HH, Kraaij R *et al*. A microbiome study to explore the gut-skin axis in hidradenitis suppurativa. J Dermatol Sci. 2021 Mar;101(3):218-220.

PhD portfolio

Name:	Suk Yee Lam
Department:	Laboratory of Gastroenterology and Hepatology
PhD period:	2017 – 2021 (4 years)
Promotor:	Prof. dr. Maikel P. Peppelenbosch
Co-promotor:	dr. Gwenny M. Fuhler

General training and courses for academic and research skills		**ECTs**
2020	Weekly research group education	0.15
2020	Weekly seminar program in experimental Gastroenterology & Hepatology	0.30
2019	Basic course Rules and Organisation for Clinical researchers (BROK)	1.50
2019	Microsoft Excel: Advanced	0.40
2017	Microsoft Excel: Basic	0.30
2017	Scientific integrity	0.30
2017	SNP Course: SNPs and Human Diseases	2.00
2017	Microbiomics I	0.60
2017	Microbiome course (online)	1.00
2017	Molecular Diagnostics	1.00
2017	GraphPad Prism Version 7	0.30
2017	Photoshop and Illustrator CC	0.30
2017	Systematisch literatuur onderzoek Pubmed en andere databases	0.60
2017	EndNote workshop	0.20
2017-2019	Weekly research group education	3.30
2017-2019	Weekly seminar program in experimental Gastroenterology & Hepatology	6.75

Teaching and reviewing tasks		**ECTs**
2019-2020	Supervision Laboratory Student	8.50
2019	Supervision Junior Medicine students	1.40
2018	Reviewing for British Journal of Surgery	0.29
2018	Reviewing for Canadian Journal of Gastroenterology and Hepatology	0.29
2018	Reviewing for Scientific Reports	0.29

Presentations and conferences		**ECTs**
2019	International Digestive Disease Forum (poster)	0.85
2019	Molecular Medicine Day	0.30
2018	Molecular Medicine Day	0.30
2018	Collaboration meeting at Chinese University of Hong Kong CUHK	0.30
2018	Seminar in experimental Gastroenterology & Hepatology (presentation II)	0.55
2018	Seminar in experimental Gastroenterology & Hepatology (presentation I)	0.55
2017	Seminar in experimental Gastroenterology & Hepatology (presentation)	0.55

Curriculum vitae

Suk Yee was born in Amsterdam (1989), the Netherlands. She grew up with her beloved parents and siblings. After finishing the gymnasium program at high school in 2007, she started medical school at VU Amsterdam where she obtained her degree in January 2014. Thereafter, she gained clinical experience as a young medical doctor at the departments of internal medicine, gastroenterology, cardiology and pulmonary medicine. In 2017, she had the opportunity to start her PhD research under the supervision of prof. dr. Maikel P. Peppelenbosch and dr. Gwenny M. Fuhler at the laboratory of Gastroenterology and hepatology of the Erasmus MC University Medical Center Rotterdam. She is currently pursuing her aspirations to a career in clinical medicine.