

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Integrated Molecular Epidemiology of Colorectal Cancer: An Exploration of Chronic Inflammation and the Gut Microbiome

Citation

Mehta, Raaj S. 2016. Integrated Molecular Epidemiology of Colorectal Cancer: An Exploration of Chronic Inflammation and the Gut Microbiome. Doctoral dissertation, Harvard Medical School.

Link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:27007748

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

Table of Contents

Abstract……..…… 3
Summary…… 4
Introduction …… 5
MIC-1 and Risk of Colorectal Cancer
Methods ……..... 13
Results ……..... 16
Commentary …… 18
MIC-1 and Colorectal Cancer Survival
Methods …… 21
Results ……..... 24
Commentary ……..... 26
Diet and Risk of CRC According to *F. nucleatum* enrichment
Methods …… 30
Results ……..... 33
Commentary …… 35
Genomic and Functional Stability of the Gut Microbiome
Methods …… 38
Results …….. 42
Commentary …… 47
Discussion and Afterward …… 50
Abbreviations ……... 53
Acknowledgments …… 54
References …… 55
Tables and Figures
MIC-1 and Risk of CRC ……..... 63
MIC-1 and CRC Survival …… 69
Diet and Risk of CRC According to F. nucleatum enrichment ……..... 75
Genomic and Functional Stability of the Gut Microbiome …… 83

Abstract

Background: Colorectal cancer (CRC) is a heterogeneous disease, arising from a complex environment of host inflammation and gut microbes. We investigated whether the novel circulating inflammatory molecule, Macrophage Inhibitory Cytokine-1 (MIC-1), was associated with risk of CRC and CRC-associated mortality. In addition, to gain insights into the role of the gut microbiome in CRC development, we explored how a particular microbial species, *Fusobacterium nucleatum,* may mediate the association between diet and CRC incidence. Finally, to evaluate the fecal microbiome as a biomarker for CRC, we measured the temporal dynamics and stability of gut microbiome communities, metagenomes, and metatranscriptomes in a large prospective cohort.

Methods: Data for these prospective analyses were collected from two ongoing cohorts, the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). In both cohorts, participants have returned questionnaires every 2 years to provide information about lifestyle and dietary factors, medication use, and diagnoses of CRC and other diseases. We measured plasma MIC-1 in pre-diagnostic blood samples (1990 in the NHS; 1994 in the HPFS) drawn from 618 participants with incident CRC and 950 matched controls. We also assessed the relationship between levels of MIC1 and prostaglandin-endoperoxide synthase 2 expression (PTGS2 or cyclooxygenase-2), measured in 245 tumor samples by immunohistochemistry. In addition, we calculated Western & Prudent dietary pattern scores among NHS and HPFS participants and classified 990 CRC cases into either *F. nucleatum*-positive tumors [Fuso(+)] or *F. nucleatum*-negative tumors [Fuso(-)] tumors according to levels of the *nusG* gene sequence unique to *F nucleatum*. Finally, we performed taxonomic and metagenomic profiling on 929 samples collected from 308 participants and metatranscriptomic profiling on 382 samples collected from 96 participants. Unconditional logistic regression models and Cox proportional hazards models were used to examine the association between MIC-1 and risk of CRC as well as CRC-specific mortality. Cox proportional hazards models were also used to examine the association between diet and CRC according to *F. nucleatum* status. Finally, we calculated Bray Curtis scores to estimate stability of the gut microbiome, and we used GLMs to associate putative factors with microbial stability.

Results: Plasma MIC-1 was associated with a higher risk of incident CRC (Ptrend = .004). Comparing extreme quintiles, the multivariate relative risk (RR) for incident CRC was 1.93 (95% confidence interval [CI] = 1.27 to 2.94). Among individuals with established CRC, plasma MIC-1 was associated with CRC-specific death (Ptrend =0.009). Comparing extreme quartiles, the multivariate hazards ratio (HR) for CRC-specific death was 2.40 (95% confidence interval: 1.33- 4.34). In addition, we found that compared to those in the lowest quartile of the Western dietary pattern, those in the highest quartile had a multivariable-adjusted HR for CRC of 1.30 (95% CI, 1.14-1.48; ptrend, 0.0001). Conversely, those in the highest quartile of Prudent dietary pattern had a RR of 0.87 (95% CI, 0.78-0.97; ptrend = 0.03). The association between Prudent diet and risk of CRC differed significantly according to *F. nucleatum* enrichment (P_{het} = 0.005). Comparing extreme quartiles, Prudent diet was associated with a multivariate RR of 0.41 (95% CI = 0.24 to 0.70, P_{trend} = .003) for Fuso(+) tumors. In contrast, the corresponding RR was 0.98, 95% CI = 0.79 to 1.21, P_{trend} = .73) for Fuso(-) tumors. This differential association appeared to be driven by greater intakes of vegetables (Phet = 0.02) and/or dietary fiber (Phet=0.02). Finally, we found evidence that human gut microbiome organismal composition and metagenomic profiles, but not metatranscriptomic data, remain highly personalized over time. Metagenomic stability accounted for 70% of metatranscriptomic stability. Congruently, we found that 80% of the pathways identified as differentially regulated were consistently over- or underexpressed over three or more time points.

Conclusions: Our results support an association of higher levels of circulating MIC-1 (GDF15) with CRC incidence and CRC-associated mortality. In addition, we found that higher Prudent dietary pattern scores as well as greater intakes of vegetables and dietary fiber are associated with lower risk for Fuso(+) CRC but not Fuso(-) tumors, suggesting a potential role of the microbiome in mediating the association of diet with CRC. Finally, we found that the vast majority of gut microbial species and metagenomes can potentially be used as indicators of long-term exposure for future biomarker studies of CRC.

Summary:

Colorectal cancer (CRC) is the second leading cause of cancer-related death in the United States. Traditional epidemiological evidence has found that modifiable risk factors may contribute to over 75% of all CRC. Despite these advances, the prevention and treatment of CRC remain challenging. We still do not understand the mechanisms by which these risk factors drive CRC, and new data suggests that CRC is a complex group of diseases characterized by distinct molecular subtypes rather than a single disease. Thus, compared to traditional epidemiological approaches, integrated molecular epidemiology offers several advantages: 1) a mechanistic understanding of CRC carcinogenesis, 2) the discovery of new predictive and prognostic biomarkers, and 3) identification of personalized preventative and therapeutic interventions. In this thesis, we explore molecular data obtained from plasma, colorectal tumor tissue, and stool in order to characterize the development and progression of CRC. In the first two parts, we examine the association of a novel circulating marker, Macrophage Inhibitory Cytokine-1 (MIC-1), with the risk of CRC as well as with mortality from CRC among patients with established CRC. In the third part, we examine the relationship between Prudent and Western diets with the risk of CRC, and explore if the association differs according to intratumoral levels of a potentially oncogenic microbe, *Fusobacterium nucleatum*. Finally, in hopes to leverage the gut microbiome as a biomarker for future prospective studies of CRC, we explore the genomic and functional stability of these 100 trillion microorganisms. Our results indicate that MIC-1 measured several years before the time of diagnosis is associated with a nearly 2-fold increased risk of CRC and a nearly 2.5-fold increased risk of death from CRC. In addition, we found that Western diets were associated with a higher risk of CRC and that conversely, Prudent diets were associated with a lower risk of CRC. This apparent risk reduction appeared to be confined to colorectal tumors enriched with *F. nucleatum.* Finally, we found that human gut microbiome species and DNA profiles, but not RNA data, remain highly personalized over time and that DNA stability accounted for 70% of RNA stability. Taken together, these findings highlight the importance of taking an integrated molecular epidemiological approach to CRC. In addition, they provide evidence for the role of chronic inflammation in the development and progression of CRC. Finally, they underscore the need for future prospective studies examining the association of the gut microbiome as characterized by stool specimens to understand the complex intersection of diet, the intestinal microbiota, and colorectal carcinogenesis.

Introduction

Colorectal cancer (CRC) is the second leading cause of cancer-related death in the United States (Siegel 2015). Among developed nations, the lifetime risk of developing CRC is approximately 5%, and the lifetime risk of developing an adenoma, a cancerous precursor to CRC, is 20% (Siegel 2015). Notably, with increasing development and urbanization in low- and middle-income countries, the incidence of CRC is rising and is becoming a major global health problem (Kanavos 2006). When the disease is confined to the primary site or has spread to regional lymph nodes, cure rates range from 70%–90%; however, advanced CRC has much higher mortality rates (Howlader 2015).

The vast majority of CRC occurs sporadically, rather than in an inherited fashion (Markowitz 2009). A recent landmark twin study found that CRC has the lowest heritability estimates – between 14 to 15% – relative to the other 14 most common malignancies (Mucci 2016). Correspondingly, over the last fifty years, epidemiological evidence has suggested there are specific modifiable risk factors which contribute to over 75% of all CRC (Wu 2016). These include lifestyle and behavioral risk factors such as consumption of red and processed meat, central adiposity, and smoking as well as protective factors such as regular aspirin use, dietary fiber, and physical activity (Chan 2010). As such, even in the absence of early detection through endoscopic screening, CRC is one of the major cancers for which a large proportion of cases may be theoretically preventable (Haggar 2009).

Despite these notable epidemiological advances, the prevention and treatment of CRC remain challenging. The mechanisms by which previously identified epidemiological risk factors influence the development of colorectal neoplasia still remain speculative (Ogino 2011). In addition, across different population-based cohort studies and in randomized clinical trials, there appears to be considerable variability in the association between exposures and the development of CRC (Song 2015). Some of this heterogeneity may be explained by accumulating data suggesting that CRC is not a single disease (Yamamuchi 2012). Instead, CRC appears to be a diverse group of diseases, emerging from a complex tumor microenvironment, with different sets of genetic and epigenetic alterations (Markowitz 2009). As such, recent data is beginning to suggest there are differential associations between lifestyle behaviors and tumor subtypes – such as physical activity and the risk of CTNNB1-negative CRC – or interactions between host

genotypes and environmental exposures – such as aspirin and the risk of CRC according to specific single nucleotide polymorphisms (Nan 2015, Morikawa 2011).

Therefore, an integrative molecular epidemiological approach may offer new approaches to mechanistically understand the relationship between exposures and development of colorectal neoplasia as well as identify new markers for susceptibility to CRC. Specifically, there is an unmet need to: 1) identify new, reliable molecular biomarkers for the prediction and prognostication of CRC; 2) understand the mechanisms by which plausible etiologic factors influence the carcinogenic process; and 3) develop personalized dietary and lifestyle recommendations for individuals with susceptibility for specific tumor subtypes.

Therefore, this thesis will take an integrated molecular epidemiological approach to CRC. Analyzing decades of prospectively collected dietary and lifestyle information from the Nurses' Health Study (NHS) and Health Professionals' Follow-up Study (HPFS) along with molecular data obtained from plasma, CRC tumor tissue, and stool specimens, we specifically explore two domains within the burgeoning field of molecular epidemiology for CRC: chronic inflammation and the intestinal microbiome.

Chronic inflammation, MIC-1, and CRC

Considerable evidence suggests that chronic inflammation plays an important role in the development of colorectal cancer (CRC). Individuals with long-standing colonic inflammation due to inflammatory bowel disease (IBD) have an increased susceptibility to CRC (Bernstein 2001). Furthermore, non-steroidal anti-inflammatory drugs (NSAIDs), particularly aspirin, are associated with a lower risk of CRC and improved clinical outcomes after diagnosis (Chan 2009, Chan 2008, Chan 2005). Additionally, recent data suggests that chronic inflammation may partially mediate the link between obesity and CRC (Belkina 2012).

The precise mechanism by which chronic inflammation promotes carcinogenesis is unclear, but several hypotheses include induction of DNA damage, modulation of the immune system, suppression of DNA repair, promotion of angiogenesis and cell proliferation, epigenetic changes, and inhibition of apoptosis (Elkinav 2013). One proposed pathway through which inflammation promotes carcinogenesis is the overexpression of PTGS2 (prostaglandin-endoperoxide synthase 2, also known as cyclooxygenase-2 [COX-2]) (Wang 2010, Chan 2007).

PTGS2-positive colorectal tumors appear to respond more favorably to aspirin and NSAIDs than PTGS2 negative tumors (Chan 2009).

Despite the abundance of data relating chronic inflammation to the development of CRC, there is limited and inconclusive molecular epidemiological data behind the association between systemic inflammatory markers and colorectal neoplasia. Circulating C-reactive protein (CRP), an acute phase reactant protein which is elevated in chronic inflammatory conditions, has been inconsistently associated with an increased risk of CRC, with only half of prospective studies demonstrating a significant association (Song 2013). Other inflammatory markers, such as interleukin-6 (also known as IL6), and tumor necrosis factor (TNF)-α have presented similarly equivocal data (Chan 2011, Song 2013). Of note, urinary PGE-M has recently been associated with an increased risk of advanced adenoma, and the use of aspirin and NSAIDS seem to reduce adenoma risk among women with high, but not low PGE-M (Bezawada 2014). These findings have not yet been replicated and have not been related to the development of CRC.

A novel circulating inflammatory marker is macrophage inhibitory cytokine-1 (MIC-1), also known as growth differentiation factor 15, which may play an important role in mediating the systemic inflammatory response (GDF15) (Wang 2013). MIC-1 is a product of activated macrophages and rises in the circulation in response to injury or inflammation (Breit 2011). Elevated levels of MIC-1 have been previously associated with an increased risk of diseases hypothesized to result from chronic inflammation, such as atherosclerosis and inflammatory arthritis (Breit 2011, Brown 2002). MIC-1 has also been linked to the development of cancers, including those of the prostate, thyroid, pancreas, and colon (Wang 2013, Breit 2011, Bauskin 2006). Recently, prediagnostic MIC-1 levels were associated with the development of advanced adenoma, and within the Polyp Prevention Trial, MIC-1 levels were associated with an increased risk of recurrent adenoma among patients with a history of prior adenoma (Brown 2012, Song 2016). Furthermore, accumulating data shows that high serum levels of MIC1 measured after diagnosis among patients with CRC, ovarian cancer, prostate cancer, and glioblastoma have been associated with shorter survival (Shnaper 2009, Staff 2010, Brown 2012, Brown 2003, Brown 2009). Additionally, in experimental models, MIC-1, as a member of the human transforming growth factor-β (TGFβ1) superfamily (Breit 2011), has been shown to play a specific role in carcinogenesis (Wang 2013, Zimmers 2010, Wang 2011, Baek 2006). For example, knock-out of

the MIC-1 gene in murine models abolishes the antitumoral effects of aspirin and NSAIDs (Zimmers 2010).

Therefore, in the first two substudies within this thesis, we examined the association between prediagnostic levels of MIC-1 and the risk of CRC as well as the association between prediagnostic MIC-1 levels with CRC-specific and overall survival among CRC patients. Given the important role of PTGS2 in inflammation-associated colorectal carcinogenesis, we also explored the association between MIC-1 levels and CRC according to the expression status of intratumoral PTGS2. Finally, we considered if aspirin/NSAID use may be differentially associated with CRC according to prediagnostic levels of MIC-1.

Diet, the gut microbiome, and CRC

Another nascent field of investigation within molecular epidemiology of CRC is the gut microbiome. Increasing evidence suggests that the human gut microbiome is linked to colorectal cancer (CRC) development (Ahn 2013; Sears 2014; Nakatsu 2015). These approximately 100 trillion microorganisms play important roles in epithelial cell proliferation and differentiation, intestinal immunity, nutrient processing, and prevention of overgrowth by pathogenic organisms (Zhu 2013). Recent clinical studies have also closely associated CRC with specific changes in gut microbiome communities. These include an increased abundance of gut sulfur-reducing bacteria (SRB); a depletion of *Firmicutes* and fiber-fementing *Clostridia*; an increased presence of bile-tolerant microorganisms; and a greater abundance of proinflammatory genera including *Porphyromonas* (Ahn 2013). It remains unclear, though, how these bacteria mechanistically influence carcinogenesis.

Many have postulated that the gut microbiome may drive colorectal carcinogenesis by mediating the effects of dietary components on colonic epithelium (Song 2015, Louis 2014). As mentioned above, it is well-established that diet plays a critical role in the development of CRC. Western dietary patterns – dominated by red and processed meats and saturated fat – have also been strongly linked with colorectal carcinogenesis (Fung 2003; Magalhaes 2012; Dixon 2004; Wu 2004). In contrast, Prudent patterns – rich in fruits, vegetables, fish, and poultry– have been associated with a lower risk of CRC and colorectal adenoma (Fung 2003, Terry 2001, Maghales 2012, Cottett 2005). Mechanisms underlying this diet-cancer association remain unclear. There is increasing recognition that the composition and metabolic activity of the gut microbiome may be

altered by a Western diet, and thus the gut microbiome may be the unidentified link in the diet-cancer relationship. For example, recent data suggests that increased meat intake and lower fiber consumption may be associated with altered bacterial and metagenomic richness as well as an inflammatory phenotype (Sonnenberg 2016, Cotillard 2013, Chatelier 2013, Koeth 2013). In addition, high-fat and high-meat diets in experimental models as well as in human studies have been shown to alter microbial community structures in the gut (Hildebrandt 2009, Turnbaugh 2009, De Fillipo 2010, Wu 2011, David 2014). Experimental studies have also shown that red meat, high in sulfur-containing amino acids, may promote the growth of SRB species, leading to the production of the carcinogenic hydrogen sulfide (H_2S), which may be important in stimulating T cell proliferation (Attene-Ramos 2007, Vipperla 2013, Miller 2012). Recently, L-carnitine, a prominent component of red meat, was found to be digested by gut bacteria into trimethylamine-*N*-oxide (TMAO) which may promote atherosclerosis and inflammation (Koeth 2013). Further still, heavy meat intake has been associated with decreased production of short-chain fatty acids (SCFA), especially butyrate, which are essential for epithelial colonocyte functioning; mitigation of DNA-damage; and the production of regulatory T cells, IL-10, and Foxp3 (Conlon 2012, Smith 2013). Lastly, dietary animal fat appears to promote the transformation by gut microbes of primary bile acids (BAs) into carcinogenic secondary BAs (David 2014, Ridlon 2006).

Among several possible microbial protagonists of CRC, one notable and emerging species is *Fusobacterium nucleatum.* Through 16S profiling, metagenomics, and RNA-sequencing, *F. nucleatum* has consistently been found to be enriched in CRC tumor samples relative to normal adjacent colonic tissue and in CRC cases compared to cancer- free controls (Kostic 2012, Castellarin 2012, McCoy 2013, Warren 2013, Mima 2015, Ahn 2013). Experimental data also suggest that *F. nucleatum* may play a mechanistic role in CRC tumorigenesis through modulation of host immunity and activation of pathways associated with cell proliferation (Kostic 2013, Mima 2015, Gur 2015). Furthermore, *F. nucleatum* has been linked to shorter survival; proximal tumor location; and CIMP and MSI mutations (Ito 2015, Mima 2015, Tahara 2014). Of note, recently, in a dietary intervention study, stool *F. nucleatum* levels were found to have markedly increased after participants were switched from a high-fiber, low-fat diet to a Western-style low-fiber, high-fat diet (O'Keefe 2015).

Based on these findings, in the third substudy of this thesis, we hypothesized that the association between diet and risk of CRC differ according to intratumoral levels of *F. nucleatum*.

To test this hypothesis and extend our previous analyses of dietary patterns and incidence of CRC (Fung 2003, Wu 2004, Mehta in progress), we examined if Western dietary patterns were associated with a higher risk of *F. nucleatum*-positive [Fuso(+)] tumors when compared to *F. nucleatum*-negative [Fuso(-)] tumors within two nationwide cohorts. Similarly, we investigated if Prudent dietary patterns were more strongly associated with a lower risk of Fuso(+) tumors compared to Fuso(-) tumors. Our two cohort studies – the NHS and the HFPS – offered a unique opportunity to integrate long-term dietary intake data with microbial features in CRC tissue over longitudinal follow-up.

Evaluating stool biomarkers for CRC: Genomic and Transcriptomic stability of the gut microbiome

As the role of gut microbiome in the development of CRC becomes more evident (Garrett 2015) and as high throughput sequencing improves in conducting a census of our intestinal bacteria, there is a need to understand which microbes drive colonocyte DNA damage ("bacterial drivers") and which proliferate as a result of intestinal tumorigenesis ("bacterial passengers") (Tjalsma 2012). Thus, it is essential to prospectively relate the gut microbiome to the development of CRC, prior to its occurrence. For technical reasons and because human biopsy samples are most often collected after bowel preparation, most large-scale human health-related studies focus on the stool microbiota, which recently has been shown to be highly representative of the colonic lumen and mucosa (Yasusada 2015).

Before we can exploit the gut microbiome as a predictive biomarker in prospective cohort studies or eventually as a diagnostic tool in clinical settings, it is critical to be able to discriminate between normal versus pathologic variation over time (Caporaso 2011). In addition, understanding the temporal dynamics of the gut microbiome is integral in leveraging these nearly 100 trillion microorganisms to promote human health. Large-scale changes in microbial community composition have been associated with host health conditions (Hsiao 2013, Mazmanian 2008, Blumberg 2012). Furthermore, instability of the gut microbiome within individuals has been noted to be a strong indicator of disease status (Giongo 2011).

Prior efforts have attempted to characterize the stability of microorganismal composition in the adult gut microbiome (Franzosa 2015, Caporaso 2011, Faith 2013, Flores 2014, Schlossing 2013, Jeffery 2016, Ding 2014, David 2014, Faust 2015, Rajilic 2012, Zoetendal 1998, Costello

2009). In spite of daily variability in species' relative abundance, bacterial communities in the gut microbiome have been observed to be generally consistent over a time scale of years or even decades (Caporaso 2011, Rajilic 2012, Faith 2013). This stability appears to persist even on a strain level (Faith 2013, Schlossing 2013). Moreover, specific individual differences in bacterial community structures appear to be preserved over the long-term (HMP1), allowing an individual's gut microbiome to be uniquely distinguished from that of others, serving as a fecal microbial fingerprint (Franzosa 2015). Nonetheless, despite the relative stability of the community profile over the long-term, recent studies have shown that host behavior or experiences such as a sudden change in diet, initiation of antibiotics, or acquisition of pathogenic species can lead to profound disruptions in the microbiome (David 2014, David 2014, Dethlefsen 2011). When such environmental perturbations are eliminated, the host's gut microbiome generally recovers to its original state.

Despite these advances, we lack a clear understanding of the determinants of stability and which aspects inherent to the community promote its stability (Coyte 2015). In addition, our knowledge of the variability of the functional activity of the human gut microbiome over time is limited. The Human Microbiome Project found in a subset of its participants that metabolic potential – determined by metagenomic levels – was unique and stable among individuals across time (HMP1). However, few studies have focused on the metatranscriptome, a more accurate indicator of functional activity (Franzosa 2014, Lozupone 2012). Previous research in experimental settings has revealed that dietary and xenobiotic interventions may substantially alter gut microbial gene-expression profiles (McNulty 2011, Maurice 2013). However, no study has examined the temporal stability of the metatranscriptome. Specifically, the magnitude of changes in microbial gene expression that occur over various time intervals or the factors that drive variation in gene expression over time are unclear.

To address these knowledge gaps, in the fourth substudy of this thesis, we deeply sequenced the gut microbiome among 308 individuals enrolled in the Men's Lifestyle Validation Style, nested within the ongoing HPFS. This cohort provided an unparalleled opportunity to apply insights from community ecology and epidemiology to characterize the stability of the gut microbiome structure and function over time. This study will provide a foundation for leveraging the gut microbiome as a biomarker for epidemiological studies and provide insight into how metagenomics and metatranscriptomics relate to each other over time.

Thesis aims

In the four substudies comprising this thesis, we embrace an integrated molecular epidemiological approach toward CRC using data from two prospective United States cohorts, the Nurses' Health Study and the Health Professionals' Follow-up Study. This thesis elucidates reliable and robust biomarkers for CRC, investigates the causal mechanisms by which diet promotes and prevents CRC development, and identifies markers by which we can stratify participants for personalized interventions. To review, in the first two sections of this thesis, we examine the association between prediagnostic levels of MIC-1 and the risk of CRC as well as prediagnostic MIC-1 levels in relation to CRC-specific and overall survival. In the third section, we explore if Western dietary patterns are associated with a higher risk of *F. nucleatum*-positive [Fuso(+)] tumors when compared to *F. nucleatum*-negative [Fuso(-)] tumors. In the fourth section, we characterize the stability of the gut microbiome structure and function over time so that we can leverage the gut microbiome as a biomarker for integrative molecular epidemiological studies for CRC. Finally, we conclude with a summary of our findings in a general discussion and afterword for future studies.

MIC-1 and Risk of Colorectal Cancer

Methods

Study populations

Data for this analysis were drawn from two ongoing cohorts, the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). The NHS began in 1976 among 121,700 U.S. female registered nurses aged 30 to 55 at enrollment. The HPFS began in 1986 among 51,529 U.S. male podiatrists, dentists, osteopathic physicians, veterinarians, pharmacists, and optometrists aged 40 to 75 at enrollment. In both cohorts, participants have biennially returned questionnaires with >90% follow-up to provide information about lifestyle and dietary factors, medication use, and diagnoses of CRC and other diseases. The Committee on Human Studies of the Brigham and Women's Hospital and Harvard University approved this study. Completion of self-administered questionnaires was considered to imply informed consent.

Blood collection

We collected blood samples by mailing phlebotomy kits to participants. From 1989-90, 32,826 NHS participants, and from 1993-95, 18,225 HPFS participants returned blood samples on ice packs by overnight courier. Upon receipt, blood samples were immediately centrifuged, aliquoted into plasma, and stored in continuously monitored liquid nitrogen freezers (-130 °C or below). More than 95% of the blood samples arrived in our laboratory within 26 hours of phlebotomy. Further details regarding blood collection, transportation of samples, and storage within these two cohorts has been previously described (Hankinson 1995, Wei 2006).

Selection of colorectal cancer cases and control participants

Eligible men and women for this study provided a blood specimen, completed the baseline questionnaire, and did not have a history of inflammatory bowel disease or cancer (except non-melanoma skin cancer). Incident cases of CRC through 2008 follow-up were initially self-reported and then confirmed with hospital records or pathology reports. We identified deaths through next-of-kin and the National Death Index. For all deaths, we sought information to determine the cause, including death certificates and medical records. A study physician, blinded to exposure information, reviewed all records to confirm cases, as well as to extract data on histological type, anatomic location, and stage of the cancer. Then, using risk-set sampling, we

randomly selected up to two controls from among participants who had not developed cancer by the age of the corresponding case, matching on sex (cohort), year of birth, and month of blood draw. We excluded 33 cases and their matched controls that failed laboratory assays. After these exclusions, we included 618 incident CRC case and 950 control participants for this analysis.

Laboratory Assays

In a core laboratory facility, personnel blinded to case-control status used sandwich enzyme-linked immunosorbent assays (ELISA) (R&D Systems, Minneapolis, MN) to measure MIC-1 (GDF15) levels in the archived pre-diagnostic plasma specimens. In HPFS, we assayed cases and controls identified from 1993-2008 in a single batch. In NHS, we assayed cases and controls identified from 1990-2004 follow-up in one batch and cases and controls identified from 2006-2008 in a second batch. All samples from case patients and each of their matched control participants were analyzed in the same batch. Based on quality controls randomly interspersed among the case-control samples, the coefficients of variation were 9.0% in HPFS, 7% in NHS (1990-2004), and 11% in NHS (2006-2008). In these cohorts, we have previously described our measurements of other inflammatory biomarkers, including high-sensitivity CRP, IL6, the TNFRSF1B, and adiponectin (Chan 2011, Wei 2005).

Assessment of PTGS2 (COX-2) Expression

Among cases of CRC in NHS and HPFS with available tumor specimens, we conducted immunohistochemical staining for PTGS2 as previously described (Chan 2007). A pathologist (S.O.) who was unaware of any data concerning the participants scored the tumor epithelial PTGS2 expression according to the intensity of staining compared to surrounding normal epithelium using a standardized grading scheme. As a result, tumors with moderate or strong staining were classified as PTGS2 positive, whereas tumors with weak or absent staining were classified as PTGS2 negative. For an agreement study, a random selection of 124 cases was examined by a second observer (T.M.) unaware of other data. The concordance between the 2 observers ($P<0.001$) was kappa=0.69, indicating substantial agreement.

Statistical Analysis

We calculated means, medians, and proportions for baseline characteristics of study participants, and then used parametric or non-parametric methods (i.e. Paired*t*test, χ^2, or Wilcoxon signed rank tests) to compare case and control participants. We calculated Spearman coefficients to estimate correlations between mean levels of inflammatory biomarkers, including MIC-1, and lifestyle factors among controls.

To examine the effect of MIC-1 levels on the risk of CRC, we categorized MIC-1 into quintiles based upon the distribution among the controls within each cohort. We estimated relative risks (RRs) for CRC with 95% confidence intervals (CIs) using logistic regression. We also tested for trend by assigning each participant to the median of their quintile and entering these values as continuous terms in the logistic regression model. Our results were similar when we employed conditional and unconditional (adjusting for matching factors) logistic regression; thus, we present the results of unconditional logistic regression models. To determine if there was a non-linear association between MIC-1 and CRC risk, we fit a restricted cubic spline function which has been described elsewhere (Chan 2011).

In multivariate analysis, we adjusted for matching factors as well as potential confounders including pack-years of smoking; race; alcohol consumption; body mass index (BMI); physical activity (in MET-hours/week); family history of CRC; regular use of aspirin or NSAIDs; regular use of multivitamins; post-menopausal hormone use (in women); red meat intake; energy-adjusted intake of calcium and folate; previous diagnosis of polyp; history of endoscopic screening; and the inflammatory markers CRP, IL6, TNFRSF1B , and adiponectin (as continuous measures). We also conducted stratified analyses by various CRC risk factors. To test for multiplicative interaction between stratification factors and biomarkers, we included cross-product terms between stratification factors and the continuous value of MIC-1 in our models.

To examine potential heterogeneity in the associations between MIC-1 and CRC according to cancer subsite (colon or rectum) and PTGS2 tumor status, we used a polytomous logistic regression model in which the association with MIC-1 was allowed to vary between the case groups. To calculate the statistical significance between case groups, we performed a likelihood ratio test comparing the model described above to a model in which the MIC-1 quintile variables were held constant. Finally, we tested for trend including MIC-1 as a continuous variable in the PTGS2 logistic regression models.

We used SAS version 9.2 (SAS Institute, Inc, Cary, NC) for all analyses with the exception of polytomous logistic regression models for which we used Stata version 11.0 (StataCorp, College Station, TX). All statistical tests were two-sided and $P< 0.05$ was considered statistically significant.

Results

Baseline cohort characteristics

Table 1.1 shows the baseline characteristics of the 618 CRC case participants and 950 matched control participants at the time of blood draw. CRC cases were more likely to have a higher BMI, while control participants were more likely to regularly use multivitamins and aspirin/NSAIDs and to consume more folate. The Spearman correlation coefficients between MIC-1, age, BMI, metabolic equivalent task score, and other inflammatory markers within sex-specific strata are shown in supplementary tables 1.1 and 1.1. Among both men and women, MIC-1 directly correlated with age and waist-to-hip ratio as well as plasma levels of TNFRSF1B, IL6, CRP, and total adiponectin.

Plasma MIC-1 and risk of colorectal cancer

We examined the association of MIC-1 with risk of CRC according to quintile categories determined by the distribution among controls (Table 1.2). Compared to men and women with levels of plasma MIC-1 in the lowest quintile (Q1), the multivariate RRs for CRC were 1.36 (95% CI, 0.95 – 1.95) for those with MIC-1 levels in the second quintile; 1.68 (95% CI,1.16 - 2.43) for the third quintile; 1.88 (95% 1.27 - 2.77) for the fourth quintile; and 1.93 (95% CI, 1.27–2.94) for the highest quintile after adjusting for known risk factors for CRC and plasma CRP, IL6, TNFRSF1B, and adiponectin (p for linear trend = 0.004). Although the association between MIC-1 and CRC appeared to be more evident among men compared with women, we did not observe any statistically significant heterogeneity ($p_{heterogeneity}$=0.13). We explored the shape of the relationship between MIC-1 levels and the risk of CRC using restricted cubic splines. A test for overall significance of the curve was statistically significant (p = 0.02; Figure 1.1).

To assess if there was effect modification by selected lifestyle risk factors for CRC, we performed stratified analyses of MIC-1 according to age at time of blood collection, smoking,

BMI, family history of CRC, regular aspirin/NSAID use, history of polyps, and history of endoscopic screening (Table 1.3). No statistically significant interactions were identified between MIC-1 levels and these CRC risk factors. We also did not observe material differences between MIC-1 and risk of cancers of the colon compared with rectum.

To address the possibility that occult or subclinical CRC could influence levels of MIC-1, we excluded cases diagnosed up to four years after blood draw from our analysis and observed similar results (multivariate RR for CRC 1.65, 95% CI, 1.10-2.47), comparing extreme quintiles of MIC-1. In stratified analysis of MIC-1 and risk of CRC, we also found similar results when comparing cases diagnosed within 4 years of blood draw (p for linear trend: 0.03) and cases diagnosed four years after blood draw (p for linear trend: 0.02). Finally, because cardiovascular disease and other cancers have been associated with higher levels of MIC-1, we also conducted analyses excluding cases of cardiovascular disease or any cancer diagnosed between the time of enrollment and within two years of after blood draw and observed similar results (multivariate RR for CRC 1.65, 95% CI, 1.06-2.58 comparing extreme quintiles of MIC-1, p for linear trend: 0.04).

Association between MIC-1 levels and colorectal cancer according to PTGS2 status

In an exploratory analysis, we considered the possibility that MIC-1 may be differentially associated with risk of CRC according to intratumoral PTGS2 expression among the 245 cases in which there was available tumor tissue for analysis. Compared to men and women with levels of plasma MIC-1 in the lowest quintile, the multivariate RRs for PTGS2 positive CRC were 1.65 (95% CI 0.93 – 2.93) for those with MIC-1 levels in the second quintile; 1.48 (95% CI, 0.79 - 2.77) for the third quintile; 1.92 (95% 1.02-3.64) for the fourth quintile; and 1.58 (95% CI, 0.78-3.20) for the highest quintile (p for linear trend = 0.18). In contrast, the corresponding RRs for PTGS2 negative CRC were 1.10 (95% CI 0.55-2.17) for those with MIC-1 levels in the second quintile; 0.66 (95% CI, 0.30-1.45) for the third quintile; 0.61 (95% 0.27-1.42) for the fourth quintile; and 0.80 (95% CI, 0.34-1.89) for the highest quintile (p for linear trend = 0.24). Nonetheless, a formal test of heterogeneity in the association of MIC-1 and CRC according to PTGS2 status was not statistically significant ($p_{heterogeneity}$=0.22).

Association of aspirin/NSAID use and risk of PTGS2-positive CRC according to MIC-1 levels

In an exploratory analysis, we examined if aspirin/NSAID use may be differentially associated with risk of the PTGS2 positive tumors according to baseline levels of MIC-1. Among individuals with high plasma MIC-1 levels (quintiles 2-5), compared to non-use, regular use of aspirin and NSAIDs was associated with a lower risk of PTGS2 positive CRC (multivariate RR=0.60, 95% CI: 0.41 - 0.88) but not PTGS2 negative CRC (multivariate RR=1.21, 95% CI: 0.71–2.07). In contrast, among individuals with low MIC-1 levels (quintile 1), aspirin and NSAID use was not associated with a lower risk of PTGS2 positive CRC (multivariate RR=0.57, 95% CI: 0.21-1.54) or PTGS2 negative CRC (multivariate RR=1.41, 95% CI: 0.47-4.23).

Commentary

In this prospective study, we observed that men and women in the highest quintile of MIC-1 had a nearly two-fold higher risk of incident CRC. Results were similar in men and women and persisted after adjusting for other known lifestyle risk factors for CRC, as well as other plasma inflammatory biomarkers, including CRP, IL6, TNFRSF1B, and adiponectin. The associations did not materially change according to subgroups defined by age, BMI, smoking status, regular use or non-use of aspirin/NSAIDs, family history of CRC, history of polyps, and history of screening. Our results, which included a large number of incident cases of CRC (n=618), are consistent with and extend upon a prior smaller cross-sectional investigation demonstrating an association between MIC-1 and CRC (n=58) (Brown 2003). Similarly, in a recent secondary analysis of the Polyp Prevention Trial, elevated levels of MIC-1 were associated with the presence of adenoma and an increased risk of recurrent colorectal adenoma which appeared particularly evident among users of aspirin/NSAIDs (Brown 2012).

Our findings support the role of chronic inflammation in the development of CRC. Moreover, because MIC-1 was associated with risk of CRC even after accounting for other inflammatory markers, our data suggest that MIC-1 may be a more specific predictor of CRC than these other biomarkers. This is supported by experimental evidence demonstrating a specific mechanistic role for MIC-1 in tumor development. MIC-1 is a member of the human transforming growth factor-β (TGFβ1) superfamily (Breit 2011). Like TGFβ1 itself, MIC-1 appears to have a pleiotropic effect, inhibiting neoplasia in early stages (e.g. tumor initiation) but perhaps promoting progression in later stages (e.g. metastasis) (Mimeault 2010). *In vitro*, MIC-1 is activated by the p53 gene product and is upregulated by anti-cancer compounds, such as those

in cruciferous vegetables (Wang 2013). In animal models, MIC-1 gene expression suppresses the formation of azoxymethane-induced colonic tumors and mediates the chemopreventive effect of NSAIDs (Wang 2013, Zimmers 2010, Wang 2011, Baek 2006). Under some circumstances, however, MIC-1 may play a pro-tumorigenic role. In mouse xenograft models, MIC-1 enhances tumor growth, stimulates cell proliferation, and promotes distant metastases (Wang 2013). Furthermore, in humans, a cross-sectional study that included 224 prevalent cases of CRC showed that MIC-1 levels were associated with increased tumor stage, presence of metastases, and earlier disease relapse (Brown 2003). Nonetheless, regardless of its specific role in either inhibiting or stimulating carcinogenesis, our findings provide evidence that circulating levels of MIC-1 are elevated prediagnostically and have potential as a biomarker for CRC risk.

Although we had a limited number of cases with available tissue for PTGS2 analysis, in exploratory analyses it appeared that the association between MIC-1 and CRC was more evident among PTGS2 positive tumors compared with PTGS2 negative tumors. Among the individuals with PTGS2 positive tumors and with elevated serum levels of MIC-1, we found evidence that participants taking aspirin/NSAIDs had a particularly lower risk of PTGS2 positive cancer. These findings provide additional mechanistic specificity to our results, suggesting that MIC-1 is primarily associated with cancers that arise through a proinflammatory milieu. Furthermore, they are consistent with findings in animal models that MIC-1 may mediate the anti-tumorigenic effect of aspirin/NSAID use. Similarly, aspirin/NSAID use reduced adenoma recurrence primarily among individuals with elevated MIC-1 levels in the Polyp Prevention Trial (Brown 2012).

There are several strengths of our study, including its large size, its prospective design, and high follow-up rate. By measuring MIC-1 levels before diagnosis of cancer, we minimized potential bias related to elevation of this marker by the cancer itself. Second, our matched controls were selected from the same cohort in which the cases developed, minimizing population stratification or selection bias. Third, our findings were consistent between two independent cohorts. Finally, we were able to examine associations after accounting for other inflammatory cytokines. There are also limitations to this study. First, a single measurement of MIC-1 in relation to subsequent CRC risk may underestimate associations due to regression dilution bias (Clarke 1999). Second, because our participants were all health professionals, our study may not be generalizable to the greater US population.

In summary, our results demonstrate a higher risk of CRC associated with elevated prediagnostic MIC-1 (GDF15) levels, even after accounting for other CRC risk factors and inflammatory cytokines. These findings suggest that MIC-1 may be a specific marker for risk of CRC, and provide additional evidence for the importance of chronic inflammation in the pathogenesis of CRC. Additional studies are needed to confirm these findings and further explore a mechanistic role for MIC-1 in the development and progression of CRC.

MIC-1 and Colorectal Cancer Survival

Methods

Study populations

We used data drawn from two ongoing cohorts, the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). The NHS began in 1976 among 121,700 U.S. female registered nurses aged 30 to 55 at enrollment. The HPFS began in 1986 among 51,529 U.S. male podiatrists, dentists, osteopathic physicians, veterinarians, pharmacists, and optometrists aged 40 to 75 at enrollment. In both cohorts, participants have returned questionnaires every two years with follow-up rates to 2010 of 95.4% in the NHS and 95.9% in the HPFS to provide information about lifestyle and dietary factors, medication (including aspirin) use, and diagnoses of CRC and other diseases. The Institutional Review Board at the Brigham and Women's Hospital and Harvard School of Public Health approved this study.

Blood collection

Blood samples were collected by mailing phlebotomy kits to participants. From 1989-90 and 1993-95, 32,826 NHS participants and 18,225 HPFS participants, respectively, returned a single blood sample on ice packs by overnight courier. When received, blood samples were immediately centrifuged, aliquoted into plasma, and stored in continuously monitored liquid nitrogen freezers (-130 °C or below). More than 95% of the blood samples arrived in our laboratory within 26 hours of phlebotomy. Additional details regarding blood collection, transportation of specimens, and processing and storage of plasma aliquots within these two cohorts has been previously described (Hankinson 1995, Wei 2006).

Selection of colorectal cancer participants and ascertainment of death

Eligible men and women for this study were individuals diagnosed with incident CRC after blood draw through 2010 follow-up who provided a prediagnostic blood specimen, completed the baseline questionnaire, and did not have a history of inflammatory bowel disease or other cancer (except non-melanoma skin cancer) prior to diagnosis of CRC. Baseline characteristics of participants with colorectal cancer whose blood we collected were not materially different from those of participants with CRC whose blood we did not collect (mean age, 68.8 vs. 67.3 years; former or current smoker, 58% vs. 63%; mean body-mass index [the

weight in kilograms divided by the square of the height in meters], 26.1 vs. 26.1 kg/m^2; mean metabolic equivalent [MET; i.e., exercise intensity] score per week, 20.7 vs. 18.5 hours/week; cancer of colon or rectum, 75% vs. 72%).

Cases of CRC were self-reported by participants or next-of-kin and after obtaining consent confirmed with medical records or pathology reports. We identified deaths through next-of-kin and the National Death Index. For all deaths, we sought information to determine the cause, including death certificates and medical records. A study physician, blinded to exposure information, reviewed all records to confirm cases, as well as to extract data on histological type, anatomic location, and stage of the cancer. We excluded 33 cases that failed laboratory assays, thereby leaving 618 CRC cases for the analysis.

Laboratory Assays

In a core laboratory facility, technicians used sandwich enzyme-linked immunosorbent assays (ELISA) (R&D Systems, Minneapolis, MN) to measure MIC1 (GDF15) levels in the archived pre-diagnostic plasma specimens. Based on quality controls randomly interspersed among the case–control samples, the coefficients of variation were 9.0% in HPFS, 7.0% in NHS (1990–2004), and 11.0% in NHS (2006–2008). Because we measured NHS samples in two different batches (cases and controls from 1990-2004 follow up; cases and controls from 2006-2008 follow-up), duplicate samples were included to assess laboratory drift for the NHS. The estimated coefficient of reliability between the two drift batches was 0.96 (95% CI 0.87-0.99), suggesting excellent reproducibility in the assay's performance. Furthermore, to assess the intra-individual variation of MIC1 levels over time, we also measured MIC1 among a subset of the controls enrolled in the NHS who also provided a second blood specimen in 2000-2001 (n=103). There appeared to be good correlation among MIC1 levels in these participants (Spearman's rho = 0.81, $p<0.0001$), supporting the reproducibility of MIC1 levels within individuals over a ten-year period. We have previously described our measurements of MIC1 and the other biomarkers of inflammation in these cohorts, including high-sensitivity CRP and IL6 (Wei 2005, Chan 2011).

Assessment of PTGS2 (COX-2) Expression

Among cases of CRC in NHS and HPFS with available tumor specimens, we conducted immunohistochemical staining for PTGS2 as previously described (Ogino 2006, Chan 2007). A pathologist (S.O.) who was unaware of any participant data scored the tumor epithelial PTGS2 expression according to the intensity of staining compared to surrounding normal epithelium using a standardized grading scheme. Tumors with moderate or strong staining intensity compared to normal colon were classified as PTGS2 positive, whereas tumors with weak or absent staining were classified as PTGS2 negative. For an agreement study, a random selection of 124 cases was examined by a second pathologist (T.M.) unaware of other data. There was good concordance between the 2 observers (kappa=0.69, $p<0.001$).

Statistical Analysis

We first examined heterogeneity for the association between MIC1 and CRC mortality among men in the HPFS compared with women in the NHS using the Cochran's Q test.[21] We found no significant heterogeneity (P for heterogeneity=0.41). Thus, we pooled data from the two cohorts and calculated means, medians, and proportions according to MIC1 levels for baseline characteristics of study participants for demographics and exposures at the time of blood draw as well as information on cancer stage, site, and grade at the time of diagnosis. To examine the association of MIC1 levels with overall and CRC-specific mortality, we categorized MIC1 into quartiles based upon the known distribution of MIC1 levels among previously matched controls without CRC within each cohort, with the lowest category used as reference group (Mehta 2014). The Kaplan-Meier method and log-rank test were performed for the survival analysis. In the time-to-event analysis, using CRC death as the outcome, patients dying from causes other than CRC were censored. To compare CRC-specific and overall mortality according to MIC1 levels, we estimated hazard ratios (HRs) for CRC with 95% confidence intervals (CIs) using Cox proportional hazards models. We stratified for age at the time of diagnosis, and conducted age and multivariate-adjusted analyses. Potential confounders adjusted in the model included cohort (sex); date of blood draw; race; body mass index (BMI) at the time of blood draw; physical activity (in MET-hours/week) at the time of blood draw; family history of CRC at time of blood draw; regular use of aspirin or NSAIDs at the time of blood draw; stage at diagnosis; history of polyps at the time of blood draw; grade (poor/unknown vs. well vs.

moderate) at diagnosis; anatomic tumor location (colon vs. rectum) at diagnosis; and the markers of inflammation, CRP and IL6.

We assessed collinearity of the model by using collinearity diagnostics; variance inflation factors showed no significant collinearity between each of our included covariates (all variance inflation factors < 2.0). We tested for trend by assigning each participant to the median value of their quartile and using these values as continuous terms in the hazards model. The proportional hazards assumption was tested by the Harrell and Lee test, which detects correlations between the Schoenfeld residuals of each covariate and the ranking of individual failure times (p for correlation coefficient for age at diagnosis = 0.02; all other p for correlation coefficients >0.05) (Kleinbaum 2010). In our secondary analyses, we assessed statistical interaction of plasma MIC1 (as a dichotomous variable defined according to the median based on the distribution among the controls) with subgroups, including tumor PTGS2 expression, by including cross-product terms in our models and assessing their statistical significance using the likelihood ratio test. We used SAS version 9.3 (SAS Institute, Inc, Cary, NC) for all analyses. All statistical tests were two-sided and P< 0.05 was considered statistically significant.

Results

Baseline cohort characteristics

Among the 618 eligible participants with CRC, we documented 266 total deaths among which 178 were due to CRC. For participants who were alive through the end of follow-up, the median time of follow-up from date of diagnosis was 9.2 years (interquartile range, 6.0-12.3). Plasma collection was performed at a median of 9.8 years for NHS and 6.3 years for HPFS (standard deviation, 5.0 years for NHS and 3.6 years for HPFS) before CRC diagnosis. Table 2.1 shows the baseline characteristics of the 618 CRC participants according to quartiles of MIC1 defined by the distribution among control individuals without CRC in our previous study (Mehta 2014). The proportion of colon versus rectal cancer, the distribution of stage at diagnosis, and the median age at diagnosis (including the proportion of postmenopausal women) appeared to vary significantly across quartiles of MIC1.

Plasma MIC1 and risk of CRC-specific and overall mortality

We examined the association of MIC1 with risk of CRC-specific and all-cause death according to quartile categories defined by the distribution among control individuals without CRC in our previous study (Table 2.2). Elevated levels of MIC1 were associated with a statistically significant higher in risk of overall mortality (log-rank p<0.0001; Figure 2.1) and CRC-specific mortality (log-rank p=0.003; Figure 2.2). For both cohorts together, the overall 5-year survival was 83% for those participants with the lowest levels of MIC1 compared with 62% for those with the highest levels of MIC1. The corresponding 10-year survival rates were 79% and 46%. Compared to men and women with levels of plasma MIC1 in the lowest quartile (Q1), the multivariate HRs for CRC were 1.67 (95% CI, 0.93 – 3.00) for those with MIC1 levels in the second quartile; 2.67 (95% CI, 1.49 – 4.80) for the third quartile; and 2.40 (95% 1.33 – 4.34) for the highest quartile after adjustment for other prognostic factors, including age, sex, stage at diagnosis, tumor grade, anatomic site, BMI, and family history of CRC as well as plasma CRP and IL6 (p for linear trend = 0.009). In contrast, there was no association observed with CRC-specific death when comparing extreme quartiles for CRP (multivariate HR, 1.09; 95% CI, 1.09 – 1.72; p for linear trend = 0.32) or IL6 (multivariate HR, 1.15; 95% CI, 0.72 – 1.84; p for linear trend = 0.59).

To assess if the association between MIC1 and CRC-specific death differed according to selected mortality-associated risk factors, we performed secondary analyses according to subgroups defined by age at diagnosis, sex, BMI, tumor stage, tumor grade, and smoking status (Table 2.3). No statistically significant interactions were identified between MIC1 levels and these factors. We also did not observe material differences between the association of MIC1 and risk of CRC-specific death according to anatomic site of tumor (colon vs. rectum) (p=0.92). Finally, we did not observe differences between the association of MIC1 and risk of CRC-specific death according to history of polyps at the time of blood draw (p_{int}=0.16).

To address the possibility that occult or subclinical CRC could influence levels of MIC1, we excluded cases diagnosed up to four years after blood draw from our analysis and observed similar results (multivariate HR for CRC-specific death 2.51, 95% CI, 1.20-5.24), comparing extreme quartiles of MIC1 (p_{trend} = 0.02). We also explored if the association between plasma MIC1 and CRC-specific mortality varied according to time between blood draw and diagnosis, but did not find material differences ($p_{interaction}$=0.44). Finally, given the measurement of MIC1 pre-diagnostically, we explored if stratification by tumor stage and adjustment for grade may

have attenuated our associations if the association of MIC1 and survival was mediated by differences in these prognostic factors at diagnosis. Comparing extreme quartiles, we observed a stronger association without additional adjustment for stage or grade (multivariate HR for CRC-specific death 3.10, 95% CI, 1.74-5.51) (p_{trend} = 0.0003).

Association between MIC1 levels and colorectal cancer-specific mortality according to PTGS2 status

In an exploratory analysis, we considered the possibility that MIC1 may be differentially associated with risk of CRC-specific death according to tumoral PTGS2 expression among the 245 cases in which there was available tumor tissue for analysis. We found evidence suggestive of an interaction between plasma MIC1 and tumor PTGS2 expression in survival analysis ($p_{interaction}$=0.04). Among individuals with PTGS2 positive cancers (n=161), we found that those with levels of plasma MIC1 at or above the median had a multivariate HR for CRC-specific mortality of 2.13 (95% CI 0.99 – 4.58) compared to those with levels of MIC1 below the median. In contrast, the corresponding HR for CRC-specific mortality among participants with PTGS2 negative cancer (n=84) was 0.61 (95% CI 0.13-2.93).

Association of aspirin/NSAIDs use and risk of CRC-specific death according to MIC1 levels

In an exploratory analysis, we also examined if aspirin/NSAIDs use may be differentially associated with the risk of CRC-specific mortality according to baseline levels of MIC1. Consistent with previous findings in these cohorts, regular use of aspirin compared to non-use after diagnosis was associated with a multivariate HR for CRC-specific death of 0.36 (95% CI, 0.22-0.60).[3] However, there did not appear to be a differential association of aspirin with CRC-specific survival according to prediagnostic levels of MIC1.

Commentary

In this prospective study, we observed that men and women with CRC in the highest two quartiles of prediagnostic plasma MIC1 had a nearly three-fold higher risk of CRC-specific mortality. Results were similar in men and women and persisted after adjusting for known prognostic factors, lifestyle risk factors for CRC, as well as other plasma markers of inflammation, including CRP and IL6. The associations did not materially change according to

subgroups defined by age, sex, BMI, tumor site, history of colonic polyps, tumor grade, or tumor stage. Our results, which included a large number of incident cases of CRC (n=618), are consistent with and significantly extend upon a prior smaller cross-sectional investigation demonstrating an association between MIC1 and overall survival and between MIC1 and relapse-free survival (n=224) (Brown 2003) Furthermore, they build upon our recent findings reporting an association between MIC1 and incident CRC (Mehta 2014).

Our results support the role of chronic inflammation in the development and progression of CRC. This is supported by considerable evidence demonstrating a specific mechanistic role for MIC1 in tumor development. MIC1, as a member of the TGFB1 superfamily, appears to have pleiotropic roles in carcinogenesis, showing anti-tumorigenic behavior in early stages but contributing to enhanced invasive and metastatic abilities by the cancer in later stages (Mimeault 2010). In early stages of neoplasia, in vitro evidence shows that MIC1 expression is up-regulated by several tumor suppressor genes, such as TP53 and early growth response protein 1 (EGR1) and may mediate endoplasmic reticulum-stress-induced apoptosis (Park 2012, Wang 2013). In animal models, MIC1-overexpressing transgenic mice bred to the Apc^{min} mouse (a model for human familial adenomatous polyposis) had a significantly reduced number of colonic polyps (Baek 2006)

However, MIC1 may promote neoplasia in other contexts. Experimental evidence shows enhancement of tumorigenic activity by MIC1 via the AKT and MAPK3/MAPK1 (ERK1/ERK2) pathways in human breast and gastric cancer cells (Kim 2008). In mouse xenograft models, MIC1 enhances tumor growth, stimulates cell proliferation, and promotes distant metastases (Wang 2013). Furthermore, in both mice and humans, MIC1 (GDF15) has been shown to be involved in cancer-related cachexia (Johnen 2007). Alternatively, it has been hypothesized that cancer cells may become resistant to the initial tumor suppressive effects of MIC1; thus, as these cells proliferate, ongoing expression of MIC1 may contribute to declining tumor immunity, leading to greater dissemination and/or earlier death (Wang 2013, Brown 2009). Nonetheless, regardless of the specific role of MIC1 in either inhibiting or stimulating carcinogenesis, our findings provide evidence that elevated prediagnostic circulating levels of MIC1 may have potential as a biomarker for risk of CRC death.

Although we had a limited number of cases with available tissue for PTGS2 analysis, we found evidence in exploratory analyses suggesting that individuals with PTGS2 positive tumors

and elevated plasma levels of MIC1 had a particularly higher risk of CRC-specific death compared to individuals with PTGS2 negative tumors. These findings offer a mechanistic corroboration of our results, suggesting that MIC1 is primarily associated with progression among cancers that arise and progress through a proinflammatory milieu. Our findings are also consistent with the interactive effect of physical activity and tumor PTGS2 expression in CRC survival (Yamamuchi 2013) supporting a potential immuno-modulatory role of physical activity in reducing systemic tumor-promoting inflammation. Furthermore, our exploratory results are supported by our recent findings that the association between MIC1 and incident CRC appeared more evident among PTGS2-positive tumors compared with PTGS2-negative tumors (Mehta 2014).

There are several strengths of our study, including its large size, its prospective design, and high follow-up rate. By measuring MIC1 levels before diagnosis of CRC, we were uniquely able to examine the association of an individual's pro-inflammatory status prior to diagnosis in relation to survival, minimizing the potential bias related to elevation of this marker by the cancer itself. Second, our findings were consistent between two independent cohorts. Third, we were able to examine a potential interaction between MIC1 levels and tumor expression of PTGS2 in our molecular pathological epidemiology database (Ogino 2011). Finally, we were able to examine associations after accounting for other markers of inflammation.

There are also some limitations to this study. Information on cancer treatment was limited. However, it is unlikely that differential receipt of chemotherapy could account for our findings. During the time period of the study, there was little variation in the type of chemotherapy available and the receipt of chemotherapy was likely defined by stage of disease for which we were able to adjust. Further, data on cancer recurrence and relapse are not available in these cohorts. However, because median survival for recurrent CRC was approximately 10-12 months during the time period of this study, colorectal-cancer specific survival should be a reasonable surrogate for recurrence (Meyerhardt 2005). Also, we only had a single measurement of MIC1. However, we found minimal intra-individual variation in MIC1 levels over a 10 year period among a subset of NHS participants, supporting MIC1 as a stable biomarker over time. Moreover, we found that the time between blood draw and diagnosis may have little effect on the association between MIC1 and CRC survival. Finally, because our participants were all health professionals and are largely Caucasian, our study may not be generalizable to other populations.

In summary, our data shows a higher risk of CRC-specific mortality associated with elevated prediagnostic MIC1 (GDF15) levels, even after accounting for other clinicopathological (e.g. stage) and lifestyle risk factors and other markers of inflammation. This association appeared to differ by CRC tumor PTGS2 status. These findings suggest that MIC1 may be a specific prognostic marker for risk of CRC death, and provide additional evidence for the role of chronic inflammation in the development and progression of CRC. Additional studies are needed to further explore the mechanistic role for MIC1 in the development and progression of CRC as well as a prognostic role of MIC1 after diagnosis.

Diet and Risk of CRC according to *F. nucleatum* enrichment

Methods

Study population

We used data drawn from two ongoing cohorts, the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). The NHS began in 1976 among 121,700 US female registered nurses aged 30 to 55 years at enrollment. The HPFS began in 1986 among 51,529 US male podiatrists, dentists, osteopathic physicians, veterinarians, pharmacists, and optometrists aged 40 to 75 years at enrollment. In both cohorts, participants have returned questionnaires every two years with greater than 90% follow-up to provide information about lifestyle and dietary factors, medication use, and diagnoses of CRC and other diseases. The Institutional Review Board at the Brigham and Women's Hospital and Harvard Medical School approved this study.

Assessment of dietary information

Participants self-reported average food intake over the preceding year through semi-quantitative food frequency questionnaires (FFQ), which have been previously validated and described in detail (Rimm 1992, Willett 2012). Briefly, to capture frequency of food consumption, nine response options were provided, ranging from "never or less than once per month" to "6 or more times per day." Total caloric and nutrient intake were calculated by summing up energy and nutrient intakes from all foods. For this analysis, we used information from the NHS FFQs administered in 1980, 1984, 1986, 1990, 1994, 1998, 2002, 2006, and 2010. For the HPFS, we used data from FFQs administered in 1986, 1990, 1994, 1998, 2002, and 2006. We excluded men with very high or very low caloric intakes (i.e., <600 or >3,500 kcal/d), those with a high number of blanks on their FFQ (>70 items blank) and men with a history of cancer (except for non–melanoma skin cancer) prior to baseline (1980 for the NHS and 1986 for the HPFS).

Assessment of colorectal cancer cases

In both cohorts, incident cases of CRC were reported by participants through 2010 follow-up for the HPFS and 2012 follow-up for the NHS. We identified deaths through next of-kin and the National Death Index. For all deaths, we sought information to determine the cause, including

death certificates and medical records. A study physician, blinded to exposure information, reviewed all records to confirm cases, as well as to extract data on histological type, anatomic location, and stage of the cancer. Among CRC cases with surgical resection and available tumor tissue, we retrieved formalin-fixed paraffin-embedded (FFPE) tissue blocks from hospitals throughout the U.S as previously detailed (Morikawa 2011). Normal and tumor sections from all CRC cases were reviewed by a pathologist (S.O.).

Fusobacterium nucleatum analysis

DNA was extracted from CRC tissue obtained from sections of FFPE tumor blocks using QIAamp DNA FFPE tissue kits (Qiagen). We performed a quantitative PCR (qPCR) assay to measure the amount of intratumoral *F. nucleatum* DNA, using custom TaqMan primer/probe sets (Applied Biosystems) for the *nusG* gene of *F. nucleatum* and for the reference human gene, *SLCO2A1*as previously described and validated (Castellarin 2011, Mima 2015). Amplification and detection of DNA was performed with the StepOnePlus real-time PCR Systems (Applied Biosystems). Levels of intratumoral *F. nucleatum* DNA in each specimen were calculated as a relative value normalized to levels measured of *SLCO2A1* using the $2^{-\Delta Ct}$ method (where delta cycle threshold (Ct)='the mean Ct value of *F. nucleatum*'—'the mean Ct value of *SLCO2A1*') (Mima 2015). The interassay coefficient of variation of Ct values from the same specimen in five different batches was 1% or less for all targets (Mima 2015). Each specimen was analyzed twice for each target in a single batch, and we used the mean of the two Ct values for each target. Tumors were classified as Fuso(+) if they had any detectable *F. nucleatum* DNA. Tumors were classified as Fuso(-) if they had no detectable *F. nucleatum* DNA.

Statistical Analysis

Two dietary patterns – named "Prudent" and "Western" -- were derived by principal component analysis (PCA) as previously described and validated with good reproducibility (Kleinbaum 2013, Hu 1999, Fung 2003). To briefly summarize, foods from the FFQ were classified into approximately 40 food groups based on nutrient profiles or culinary usage, with the numbers varying based on slight variation in FFQs from cycle to cycle. Foods that did not fit into any of the groups were left as individual categories (eg. pizza, French fries, and tea). Vitamin and mineral supplements were not included in the definition of the patterns. Factor

analysis was then performed using an orthogonal rotation procedure (the Varimax function in SAS) to produce two maximally uncorrelated factors, selected based upon the largest eigenvalues. Factor loadings were then derived based on the correlations between food groups and the two derived factors. Each participant was next assigned a factor score, determined by adding the reported frequencies of food group intakes, weighted by the factor loadings. These factor scores were then standardized using a z-score scale, with a mean of 0 and standard deviation of 1. Higher scores indicated stronger adherence to that particular pattern. To capture long-term habitual consumption, we calculated the cumulative average of factor scores, which we then categorized into quartiles.

Using Cox proportional hazards models, we examined the association between Western and Prudent dietary patterns and incidence of CRC. All analyses were adjusted for total caloric intake (kcal per day) and stratified by age (in months), sex (in the analysis using combined cohorts), and year of questionnaire return. In multivariable analysis, we adjusted for several risk factors for CRC including body mass index (kg/m^2), pack years of smoking (never, 0-<5 years, 5-<20 years, 20-<40 years, or ≥40 years), family history of CRC in any first-degree relative (yes or no), previous lower gastrointestinal endoscopy (yes or no), postmenopausal hormone use ((for women only) never, past current), physical activity [quintiles of metabolic-equivalent task (MET) hours per week],alcohol consumption (0-<5,5-<10,10-<15,15-<30, or ≥ 30 grams per day), and regular aspirin or NSAID use (≥2 tablets/wk).

To examine whether the association between dietary patterns and incidence of CRC differed according to intratumoral *F. nucleatum* status we then used Cox proportional hazards regression models with a duplication method for competing risks data and computed hazard ratios (HR) and 95% confidence intervals (CI). We tested for heterogeneity by using a likelihood ratio test, comparing a model that allows for separate associations of dietary patterns and risk of CRC according to *F. nucleatum* status with a model that assume a common association (Wang 2015). To test for trend, participants were assigned to the median value of their dietary pattern quartile and then this variable was entered into the models as a continuous term. In a secondary analysis, we examined associations with major macronutrients as well as dominant factor loadings for each pattern using cumulative average intakes of fruits, vegetables, red meat, and processed meat as well as energy-adjusted intakes of fat, fiber, and protein, all of which were categorized into quartiles.

Prior to pooling data from the two cohorts, we examined the possible heterogeneity between cohorts, using the Q statistic for the association between the Prudent and Western dietary patterns and overall incidence of CRC. We did not observe significant heterogeneity between cohorts (*P*=0.66 for the Prudent dietary pattern*; P*=0.84 for the Western dietary pattern). We used SAS software version 9.3 (SAS Institute Inc.) All statistical tests were two-sided and we consider $p<0.05$ as significant.

Results

Among 135,878 participants (89,768 women and 46,110 men), we documented 3,042 CRC cases that developed during 3,552,392 person-years of follow-up. Among CRC case patients in the NHS and HPFS with available tumor specimens, we classified tumors into either *F. nucleatum*-positive tumors [Fuso(+)] (n=123) or *F. nucleatum*-negative tumors [Fuso(-)] (n=867). Two major dietary patterns were identified by factor analysis that we labeled Prudent and Western. The Prudent dietary pattern was characterized by high intakes of vegetables, fruits, legumes, poultry, and fish. In contrast, the Western dietary pattern was characterized by red and processed meats, high-fat dairy products (such as whole milk and cream), refined grains, desserts, and French fries (Table 3.1). The characteristics of the patterns were stable across repeated FFQs and in both cohorts. Consistent with prior analysis, Western pattern scores were associated with less healthy behaviors whereas Prudent pattern scores were associated with healthier behaviors (Table 3.2) (Fung 2003, Wu 2004). For example, participants with high Western scores tended to smoke more, exercise less, and consume more alcohol, particularly among the men. In contrast, those with a high Prudent score were more likely to have a history of screening lower endoscopy and smoke less.

We found that Western dietary pattern scores was positively whereas and Prudent dietary pattern score was negatively associated with overall incidence of CRC (Table 3.3). Compared to those in the lowest quartile of Western dietary pattern scores, those in the highest quartile had a multivariate-adjusted HR for CRC of 1.30 (95% CI, 1.14-1.48; ptrend, 0.0001). In contrast, compared to those in the lowest quartile, those in the highest quartile of Prudent pattern had a HR of 0.87 (95% CI, 0.78-0.97; ptrend = 0.03).

We then examined the association of Western and Prudent dietary pattern scores with Fuso(+) and Fuso(-) CRC in the two cohorts (Table 3.4). Notably, we found that the association

between Prudent dietary pattern and risk of CRC differed by *F. nucleatum* enrichment status, with an inverse association restricted to Fuso(+) tumors. Comparing men and women with the highest Prudent dietary scores to those with the lowest Prudent dietary scores, the multivariate hazard ratio for Fuso(+) tumors was 0.41 (95% CI, 0.24-0.70, ptrend = 0.003) and was 0.98 (95% CI, 0.79-1.21, ptrend=0.73) for Fuso(-) tumors. The *P* value for heterogeneity was 0.005. For the Western dietary pattern, we did not observe significant heterogeneity (Phet=0.21). It appeared, however, that the association between increased Western dietary intakes and CRC was more evident among Fuso(+) tumors compared with Fuso(-) tumors.

In a secondary analysis, we sought to identify if specific food groups dominated the observed associations between Prudent and Western dietary patterns and risk of CRC according to *F. nucleatum* levels. We examined the top two dominantly contributing food groups for Western diet – red meat and processed meat and Prudent diet – fruits and vegetables in relation to risk of CRC by *F. nucleatum* status (Table 3.5). As with Prudent dietary pattern scores, we observed significant heterogeneity for vegetable intake (Phet=0.02) and heterogeneity approaching significance for fruit intake (Phet=0.15). In contrast, there did not appear to be heterogeneity for red or processed meat intake (Phet = 0.90 and 0.24, respectively). Mutual adjustment for the three other food groups and addition of micronutrients (folate, Vitamin D, and calcium) to the multivariate models did not appreciably change the associations presented (data not shown).

Finally, to further characterize which macronutrient components of the Western and Prudent dietary pattern dominated the observed differential associations according to *F. nucleatum* status, we explored associations of fat, fiber, and protein intake with CRC (Table 6). Notably, participants with greater intakes of fiber had a decreased risk of Fuso(+) CRC compared to participants with the lowest intakes of fiber (MV HR: 0.55, 95% CI 0.32-0.94) but not for Fuso(-) CRC (HR 1.19, 95% CI 0.96-1.48) (*P* for heterogeneity was 0.02). In contrast, we did not observe significant heterogeneity for fat or protein. In an exploratory analysis, we found that cereal-derived and vegetable-fiber, specifically, were the fiber subtypes most differentially associated with CRC according to *F. nucleatum* levels (Table 3.S1). Again, mutual adjustment for the two other macronutrients and addition of micronutrients (folate, Vitamin D, and calcium) to the multivariatee models did not appreciably change the associations presented (data not shown).

Commentary

In two large prospective cohorts, we found that participants with higher long-term Western dietary pattern scores were associated with a higher risk of CRC. In contrast, we observed that participants with higher long-term Prudent dietary pattern scores had a lower risk of CRC. This apparent risk reduction appeared to be confined to colorectal tumors enriched with *Fusobacterium nucleatum.* Similarly, we found evidence that higher intakes of vegetables and dietary fiber were associated with a decreased risk of Fuso(+) CRC tumors but not Fuso(-) tumors. These findings support the hypothesis that the CRC-preventative effects of healthy diets and dietary fiber may be mediated by modulation of specific species in the gut microbiome, such as *F. nucleatum*. To our knowledge, our study is the first to examine the intersection of diet and CRC according to microbial features of the tumor.

The potential role of diet in modulating the risk of CRC is widely recognized (Song 2015). Our finding that Western diets are associated with an increased risk of CRC builds on several prior studies and echoes recent conclusions by the World Health Organization that consumption of red and processed meat is carcinogenic or probably carcinogenic to humans (Wu 2004, Fung 2003, Bernstein 2015, Bouvard 2015). Our data are also consistent with a recent meta-analysis reporting that healthy or "Prudent" dietary patterns lead to a reduction in risk of CRC (Maghales 2012). Importantly, however, epidemiological data for associations of Prudent dietary patterns and its major components – including fruits, vegetables, and dietary fiber – with CRC have been more equivocal than for red and processed meat (Aune 2011; Aune 2011; Lee 2011).

Our results here suggest that some of the previously observed heterogeneity in the reduction of risk of CRC by Prudent dietary patterns, fiber, and vegetables may because due to differential modulation of the gut microbiome as characterized by tumor microbial profiles, including *F. nucleatum*. These data are supported by recent evidence showing that a two-week feeding intervention switching rural-dwelling South Africans from a high-fiber, low-fat diet to a low-fiber, high-fat diet was associated with an increase in *F. nucleatum* measured by PCR in the stool (O'Keefe 2015). In addition, some have hypothesized that the regional variation observed in *F. nucleatum* levels in CRC tumors collected from Spain, Vietnam, and the United States may be due to differences in dietary practices in these countries (Sears 2014; Kostic 2012).

Furthermore, in a retrospective cross sectional study, participants with advanced adenoma were found to have lower dietary fiber intakes as well as distinct fecal microbiome communities when compared to healthy controls (Chen 2013).

The precise mechanism by which Prudent diets as well as greater intakes of vegetables and fiber may differentially prevent *F. nucleatum*-enriched cancer development remains unclear. Accumulating evidence suggests that long-term dietary fiber intake has a profound impact on the gut microbiome (Wu 2011, DeFillipis 2015, Ou 2013, Sonnenburg 2016). Specifically consumption of fiber, fruit, and vegetables is associated with increased microbial taxonomic and genomic richness (Cotillard 2013, Chatelier 2013, Claesson 2012). It is plausible that abundance of microbiota-accessible carbohydrates from Prudent diets, vegetables, and dietary fiber may lead to increased bacterial fermentation of short-chain fatty acids (SCFA) which may alter pH, increase transit time, or lead to altered immune surveillance to make conditions less hospitable for non-native species such as *F. nucleatum* to invade colonic mucosa and promote CRC tumorigenesis (Sonnenburg 2016, O'Keefe 2015, David 2014, Garrett 2015, Smith 2013, Louis 2015).

There are several strengths to this study. First, our dietary data were prospectively collected, minimizing the possibility of recall or selection bias, and our dietary instruments have been well-validated (Rimm 1992; Willet 2012). Second, our data were detailed and updated such that we were able to explore associations with CRC according to broad dietary patterns, specific food groups, and macronutrients. Third, detailed data on a wide range of exposures allowed us to adjust for multiple confounding factors, although the possibility of residual confounding cannot be excluded. Finally, we had a large number of cases with tumor tissue in our molecular pathological epidemiology database collected from participants across the U.S (Ogino 2011).

We also acknowledge limitations to this study. With the use of FFPE tissue specimens, routine histopathology procedures may influence the sensitivity of our qPCR assay to detect microorganisms. However, in a limited validation study, we detected *F. nucleatum* in paired FFPE and frozen tissue specimens in 2 colorectal carcinoma cases by qPCR. Additionally, despite the large size of the two cohorts, the number of Fuso(+) cases was relatively small. Finally, because our participants were all health professionals and are mostly white, our study might not be generalizable to other populations.

In summary, we show that higher Prudent dietary pattern scores as well as greater intakes of vegetables and dietary fiber are associated with a lower risk of *Fusobacterium nucleatum*-enriched CRC tumors but not *Fusobacterium nucleatum*-negative tumors. Our data provide evidence and generate new hypotheses about how the microbiome may mediate the association between diet and risk of CRC. Further studies are needed to confirm these findings and determine the potential utility of using *F. nucleatum* detection in colonic mucosa as a biomarker for personalized nutritional interventions. In addition, our findings underscore the potential importance of future prospective studies examining the association of the gut microbiome as characterized by stool specimens to understand the complex intersection of diet, the gut microbiome, and colorectal carcinogenesis.

Genomic and Functional Stability of the Gut Microbiome

Methods

Study population

The 308 individuals sampled in this study were recruited into a Microbiome Correlative component of the Men's Lifestyle Validation Study (MLVS), a substudy nested within the Health Professionals Follow-up Study. The HPFS is an ongoing prospective cohort of 51,529 United States men described elsewhere (Wei 2005). The MLVS, starting in July 2012 and ending in July 2013, was established among 908 men who had completed the 2010 HPFS FFQ and had previously provided a blood sample in 1986. Men with coronary artery disease, stroke/TIA, cancer (does not include squamous or basal cell skin cancer), or major neurological disease (ALS, Alzheimer's, Parkinson's, epilepsy, MS) were excluded. Subject recruitment and protocols were approved by the Harvard School of Public Health Institutional Review Board.

Sample collection

Participants provided up to four stool samples – a pair of stool samples sampled from two consecutive bowel movements 24-72 hours apart followed by collection of a second pair of samples approximately 6 months later. The breakdown of the number of stool samples collected is detailed in Figure 4.1. Participants deposited each bowel movement into a commode plastic collection bowl, and then using a specially designed spoon attached to a collection tube cap, feces were scooped along with into a tube containing RNAlater (Franzosa 2014). At each stool collection, participants documented the date and time of defecation as well as a Bristol score (Lewis 1997). In addition, along with each pair of samples, participants completed a brief questionnaire collecting information regarding recent use of antibiotics, gastric acid reducing medications, laxatives, probiotics as well as other relevant exposures. At each time point, participants stored the specimen in the RNAlater fixative at ambient temperature until the specimen collected from the second bowel movement was produced. Each pair of stool specimens were then placed in a special mailing kit and returned at ambient temperatures by overnight mail. Upon receipt, the collection tubes were immediately placed into -20 degrees C freezers.

Laboratory Methods and Sequencing

Stool samples were removed from the freezers and thawed in a room-temperature water bath for 10 minutes. After homogenization with a spatula, 100 mg of stool was aliquoted and centrifuged at maximum speed. 110 uL of TE buffer with Proteinase K and lysozyme was added with incubation on a laboratory shaker for 10 minutes, followed by the addition of a 1 mL RLT and BME mixture for cell lysis and denaturation of RNases. We then bead beat each sample for 3 minutes and centrifuged each for 4 minutes. Qiagen AllPrep spin columns were then used for RNA and DNA extraction according to the manufacturer's protocol (Qiagen). We transferred the supernatant from the bead beating step to shredder columns and spun for 2 minutes. DNA was extracted from all 929 samples. We poured the eluent from all samples into DNA spin columns, centrifuged for 30 seconds, and transferred the column to the collection tube and incubated for fifteen minutes. We then added 500 uL of the AW1 wash buffer to DNA column, and centrifuged for 15 seconds, followed by 500 uL of the AW2 wash buffer and centrifuged for 2 min. Finally, we added 50 uL EB buffer and centrifuged for 1 min and then transferred the eluent to clean tubes.

RNA was extracted from 384 samples, derived from a subset of 96 individuals who provided all four stool samples spanning the 6 month study interval and who did not report use of antibiotics within the last year. We added 600 uL of 70% EtOH to the flow-through previously collected and then pipetted the mixture into a Qiagen RNA column and centrifuged for 30 seconds. We then added 350 uL of the RW1 buffer, centrifuged for 15 seconds, added 60 uL of DNAase solution and incubated at room temperature for 15 mins. We repeated with 350 uL of the RW1 buffer and centrifuged for 30 seconds. We then washed with 500 uL RPE buffer, added 50 uL of RNAse free H2O, centrifuged for 1 minute, and transferred to a clean tube. A NanoDrop 1000 (Thermo Fisher Scientific) was used to determine DNA and RNA concentration, quality and purity.

Sequencing & Taxonomic/Functional Profiling

Using Illumina HiSeq paired-end (2x101) shotgun sequencing, the 929 metagenomes and 378 metatranscriptomes were sequenced. Six metatranscriptomes that were extracted did not undergo sequencing because they were alone on a plate and did not meet the minimum number of samples to be sequenced. Before quality control, sequencing depth (mean +/- standard

deviation) for DNA was 3.8 +/- 1.6 Gnt and after was 1.8 +/- 0.7 Gnt. For RNA, mean sequencing depth was 2.8 +/- 2.4 Gnt before QC and 1.2 +/- 1.0 GNT after QC. RNA-seq libraries were depleted for ribosomal RNA using Ribo-Zero (Epicentre) as previously described (Franzosa 2014). Quality control entailed removal of human sequence contamination, quality trimming, and depletion of duplicate reads using HMP protocols (HMP1).

We then performed taxonomic profiling with MetaPhlAn2 and performed functional profiling of genes and transcripts using HUMAnN2 based on the KEGG database of gene families and on the Metacyc database of pathways (Segata 2015, Kegg, Uniref, Humann2). Outlier identification was performed and eight samples were removed because their ordination scores in either dimension (NMDS1 or NMDS2) were outside the range of the median ± 3 IQRs. Notably, four of these samples were provided by a participant who self-reported a history of colectomy. Six RNA samples were removed because they did not have matching DNA sequenced samples. We subsequently filtered all taxonomic features with a relative abundance less than 10−4 (0.01%) in 10% of all samples. Similarly, for DNA and RNA, we filtered all gene families with a relative abundance less than 10^{-5} (0.001%) in 10% of all samples.

Statistical Analysis

Variability in the relative abundance values for community composition, metagenomes and metatranscriptomes within and between samples was determined by calculating the Bray-Curtis dissimilarity metric for each individual over time (HMP1). Short-term intervals, as previously described, were defined as 24-72 hours apart whereas the intermediate interval was six months, the longest sampling interval in the study. Similarly, the Jaccard index was estimated within samples across time, where comparisons were drawn between the first sample and the remaining three (Faith 2013). With these metrics, higher values signify more variable features whereas lower values indicate more stable features.

To estimate the reliability or reproducibility in measuring relative abundance values of features over time, we calculated intraclass correlation coefficients (ICCs). The ICC is an index commonly used in epidemiological studies to assess the accuracy of measurements but also can be used to estimate the repeatability of a trait or feature over time (Kostopolous 2010). The ICC is different from an ordinary Pearson correlation coefficient because the estimated mean and variance are obtained by a pooled estimate over all individuals over all time points. It can be

interpreted as a measure of the percentage of total variation that is attributable to between-group variation. Relative abundance values were first arcsine square root-transformed to variance-stabilize data and better approximate normality, and then using linear mixed effects models, we divided the between-person variance by the sum of the within-and between-person variances (Nakagawa 2010).

We explored how several putative factors influenced global community, metagenome, and metatranscriptome stability across the longest interval in our study. Shannon index was determined by the mean of the within-sample diversity at time point 1 and time point 4; acid-lowering medication use was determined by use of proton-pump inhibitors or H2 receptor antagonists more than once a week for two months; bowel prep was determined as any use of bowel preparation in the two months prior to stool collection; and antibiotic use was defined as PO or IV administration of antibiotics in the 12 months prior to stool collection. Never users were set as the referent level, any users were defined as participants using antibiotics at any point in the last 12 months before second stool collection, and new users were defined as initiators of antibiotic use before second collection. Bristol category change was determined by categorization of participant-reported Bristol stool scores into hard (1,2), normal (3,4), and soft or liquid (5,6,7) groups and then by calculating the difference between the change in categories (Lewis 1997).

We examined associations between baseline mean relative abundance and prevalence of each feature versus feature stability, and how metagenomic variability correlated with metatranscriptomic variability per feature over time. As part of this, we investigated if dominant expression of stable genes versus expression of unstable genes varied according to species. For each transcript, a dominant contributing species was identified based upon the maximal average contribution to each transcript. We then ranked the species according to the total number of genes that they dominantly expressed and (for presentation purposes) selected those contributing to 30 or more, and overlaid them on the scatterplot in color (grouped by genus).

To test stability of regulation of gene expression over time, we identified 740 pathways found in the 340 samples for which we had DNA and RNA data across all four time points. 218 differentially regulated pathways were defined as having a mean log10(RNA/DNA) significantly different from zero at each time point using a linear model. We subjected these nominal P values to FDR correction following the Benjamini–Hochberg method with $\alpha = 0.05$. We then explored

the consistency of the differential regulation of each pathway by exploring the overlap at each time point. In addition, to identify the most stably regulated pathways, we computed the mean RNA/DNA ratio for each pathway across the 85 participants who gave 4 samples for transcriptomic analysis and calculated the CV for these means, and then ranked the pathways in descending order by their CV.

To assess the temporal variability of within sample diversity (alpha diversity), we calculated the coefficient of variation (CV) of the Shannon index for each individual over time, with higher values being indicative of more variable communities.

To determine if certain taxa or enzyme classes coded by DNA were more or less abundant in individuals with different levels of gut microbiome stability, individuals were assigned to stability classes based on quartiles (first = stable, second and third = average, fourth = variable) of the Bray-Curtis score for the 6 month period. We tested for differences between stability classes using the nonparametric Kruskal-Wallis analysis of variance. P values were adjusted for multiple comparisons using Benjamini-Hochberg correction.

Results

The gut metagenome is more stable than community structures and the metatranscriptome over time

We assessed the stability of microbiome communities, metagenomes, and metatranscriptomes among 308 participants (mean age 70±4) enrolled in the Men's Lifestyle Validation Study (MLVS), nested within the Health Professionals Follow-up Study (HPFS) cohort. Participants provided up to four stool samples – a pair of stool samples collected 24-72 hours apart followed by collection of a second pair approximately 6 months later. DNA was extracted from 929 stool samples collected from all 308 men. ; RNA was extracted and reverse-transcribed to cDNA only from 384 stool samples from participants who provided stool samples spanning 6 months and who did not report use of antibiotics within the last year. All samples were then sequenced using the Illumina HiSeq platform. Raw sequence data were filtered to remove low quality and human host reads and features with low overall relative abundance. Metagenomic and metatranscriptomic read data were profiled for functional and taxonomic composition using HUMAnN2 and MetaPhlAn2, respectively (Figure 4.1). We removed 8 samples that were outliers on the taxonomic level, including four samples collected from a participant who reported a history of a colectomy (Supp Figure 4.1).

We first compared the within-person stability of the gut metatranscriptome to the stability of organismal composition and the metagenome over short (24-72 hour) and intermediate (6 month) time intervals. This was assessed by estimating two beta-diversity metrics – Bray-Curtis dissimilarity scores and the Jaccard index – between an individual's samples collected over the study period. By both ecological metrics, over short and intermediate-term intervals, the metagenome appeared quite stable compared to species or the metatranscriptome (Figure 4.2A and Figure 4.2B). In addition, we considered if the stability of features was consistent over short- and intermediate-term intervals. For species, genes, and transcripts, stable features observed over a 24-72 hour period tended to be similarly stable over 6 months ($p < 2.2e-16$), suggesting the presence of a subset of features that can be consistently measured over time.

The gut metatranscriptome is stable but not highly personalized over time

We next explored if metatranscriptomes from participants' gut microbiomes appeared more similar to their own over short- and intermediate-term periods than to another person's. In the Human Microbiome Project, within-subject variation over time was consistently lower than between-subject variation, both in organismal composition and in metabolic function (HMP1). Although we found similar results for species and metagenomes ($p < 2.2e-16$), we found that within-subject variation over time was comparable to between-subject variation for metatranscriptomes (short-term: $p=0.30$; intermediate-term: $p=0.95$). These data were corroborated by an calculation of intraclass correlation coefficients (ICCs), which estimate the proportion of variation in relative abundance that can be explained by an individual. For the vast majority of species and genes (86.8% and 92.8%), over the longest sampling interval in our study, ICCs were greater than 0.40, generally considered to be associated with a fair-to-excellent repeatability of measurement (Rosner 2010). In contrast, only a small fraction of transcripts (0.79%) had an ICC greater than 0.40. Supp Table 4.1.

Feature abundance and prevalence are associated with feature stability over time

Consistency in feature stability over short- and intermediate-term periods led us to consider if there may be properties inherent to features that contribute to stability across all hosts. Across a range of body sites, overall taxonomic stability of the human microbiome has correlated with feature abundance and prevalence (Franzosa 2015). We extended these results by exploring

the correlations between these two ecological properties and stability of individual organisms in the gut microbiome as well as stability of genomic and transcriptomic levels of individual enzymes. Using the complement of within-person Bray-Curtis dissimilarity score (1-BC) over the longest interval in our study (time point 1 vs. time point 4, ~6 months), we similarly found that baseline feature prevalence and abundance were also strongly and positively correlated with not only organismal composition, but also genomic and transcriptional stability (Figure 4.3). Species that were highly abundant and prevalent, such as *Bacteroides uniformis*, were highly stable, whereas species that were of low prevalence and abundance, such as *Lactococcus lactis*, a Gram-positive bacteria used in making fermented foods, were less stable (Steidler 2000, McNulty 2011). Similarly, enzymes such as RNA polymerase (EC 2.7.7.6), essential for prokaryotic function, were universally prevalent, highly abundant, and stably expressed at the genomic and transcriptional level (Cooper 2000).

Notably, however, there were several features whose stability levels did not correspond with their prevalence or abundance. For example, *Sutterella wadsworthia*, in the bottom quintile of prevalence and bottom half according to abundance was in the top quintile for stability. This may owe to the role of this this species as the dominant contributor to over 30 gene families, including its lone production of the PAPS-independent sulfuryl transferase (EC 2.8.2.22), an important bacterial enzyme involved in detoxification of phenolic compounds (Malojcic 2008, Kim 1986). Additionally, as observed in the Human Microbiome Project, there were several genes that were universally prevalent at the DNA level but of relatively low genomic abundance (HMP1). Many of these enzymes had some of the most variable gene expression. For example, the pair of enzymes thiaminase I and II (EC 2.5.1.2 & 3.5.99.2), involved in bacterial degradation of thiamin and hypothesized to be overexpressed upon exposure to grain-rich diets, possessed DNA sequences detected in 100% of all participants, but were within the bottom 5% of stable transcripts (Karapinar 2010). Similarly, the enzyme ErmC (EC 2.1.1.184), which encodes one of the macrolide resistance genes, or the enzyme Superoxide Reductase (EC 1.15.1.2), responsible for hydrogen peroxide production by anaerobic bacteria in the gut and implicated in damage to coloncyte DNA, were universally prevalent at the genomic level but were unstable at the RNA level and of intermediate stability at the genomic level (Busierre 1998; Jenney 1999; Huycke 2001). Finally, since feature prevalence was correlated with abundance

(Supp Figure 4.2), after mutual adjustment, we found for species that only mean relative abundance was associated with stability (p=5.46e-06).

Relationship between richness and host factors that may influence the stability of human gut microbiome communities

In addition to examining determinants of stability for specific features, we also explored if there were lifestyle factors or habits at the time of stool collection that could explain global instability of the gut microbiome for each individual. We estimated Bray-Curtis dissimilarity values for each individual over the longest sampling interval (~6 months) for species, DNA, and RNA as our response variable. As shown in Table 4.1, there were few factors that explained instability. Common predictive factors observed across feature types included mean alpha diversity values (Shannon Index), and exposure to a bowel laxative preparation within two months prior to stool collection. Antibiotic use initiated between the collection time points seemed to affect genomic instability to a greater extent than species instability.

Relating metagenomic stability over time to metatranscriptomic stability over time

We next considered whether transcriptomic stability is related to metagenomic stability. Recent evidence from a pilot study involving a subset of these participants found that gene and transcript relative abundances were well correlated and that a substantial fraction of microbial transcripts were not differentially regulated (Franzosa 2014). Again using the complement of the within-person Bray-Curtis dissimilarity score over ~ 6 months as our metric of stability, we found that genomic instability appears to explain a large proportion of transcriptomic instability; the metatranscriptome can be no more stable than the metagenome ($R2 = 0.70$, $P< 2.2e-16$).

In addition, we investigated if dominant expression of stable genes versus unstable genes varied according to species (Turnbaugh 2010). For each transcript, a dominant contributing species was identified based upon the maximal average contribution to each transcript. We then ranked the species according to the total number of genes that they dominantly expressed and (for presentation purposes) selected those contributing to 30 or more, and overlaid them on the scatterplot in color (grouped by genus). As shown, *Bacteroides* species are dominant contributors to the greatest number of genes expressed as well as dominant contributors to the expression of the most stable transcripts. In contrast, *Streptococcus* species are dominant contributors to

expression of some of the most unstable transcripts, with several other species seen in between. This suggests that some species play a constant and major role in gene expression in the microbiome, while others play much more variable roles because they are rare, transitory, or more prone to be influenced by environmental factors.

Exploring stability of the regulation of gene expression

The previous section highlighted that much of the transcriptional variability over time for each feature can be explained by changes in the metagenome, or shifts in the underlying metagenomic code. Thus, we next queried what proportion of the remainder of transcriptional variation can be explained by differences in transcriptional regulation of features over time. In an approach similar to that performed in the pilot study, we identified significantly differentially regulated pathways, defined as having a mean log RNA/DNA abundance ratio >2, for all four time points, and then explored the overlap (Franzosa 2014). Of the 218 pathways that were significantly differentially regulated, 80% were consistently over- or underexpressed across three or four time points (Figure 4.5a). To better understand which pathways were most stably regulated, for each time point, we computed the mean RNA/DNA ratio for each pathway across the 85 participants who gave 4 samples for transcriptomic analysis. We then calculated the CV for these means, and then ranked the pathways by their CV. As shown in Figure 4.5b, the top 10 most stable pathways in terms of transcriptional regulation appear to be involved in cellular housekeeping, such as carbon metabolism (stably overexpressed) and amino acid synthesis (stably underexpressed).

Individuals with more stable gut microbiomes do not appear to have distinct taxa or gene families compared with those found in more variable individuals

We next examined if individuals who had more stable gut microbiome communities harbored taxonomic and metagenomically distinct features compared with those found in more variable individuals. Ecological theory maintains that, keystone species essential for preservation of community organization may exist in any ecosystem (Mills 1993). To test this, individuals were assigned to stability classes based on quartiles (first = stable, second and third = average, fourth = variable) of the Bray-Curtis score for the longest sampling interval and we tested for differences between stability classes using the nonparametric Kruskal-Wallis analysis of

variance. Although we found some features were nominally significant, they were no longer significant after FDR correction (Supp Table 4.2).

Temporal variability in diversity of species, metagenomes, and metatranscriptomes matches variability in communities

To quantify amount of variability in the richness of bacterial communities, metagenomes, and metaranscriptome over time, we estimated coefficients of variance (CV) for the alpha diversity Shannon Index for each individual. Low CV values indicate that an individual had stable alpha diversity levels, whereas high CV values indicate than an individual had relative variable alpha diversity levels over the 6 month study period. As evident for variability in beta-diversity, there was a wide range of variability in richness within the metatranscriptome and community structure, but not within the metagenome (Supp Figure 4.3).

Commentary

Our results broadly indicate that the gut microbiome's metagenome is generally stable over 24-72 hours and over 6 months within an individual. In contrast, an individual's organismal composition and metatranscriptome is considerably less stable. In addition, we have shown for the first time that an individual's metatranscriptome is not highly personalized over time, in contrast to his metagenomes and community structure. These findings suggest that although there are large-scale detectable differences among hosts on a compositional and metagenomic level in the gut microbiome, it appears that hosts are expressing a relatively conserved core functional set of genes that do not fluctuate dramatically over time. This may be reflected in our finding that the mostly highly expressed and stable genes were universally prevalent in the population; this may reflect the importance of overall homogeneity in gene expression during periods of generally stable health. However, substantial deviations from a central functional profile may be associated with shifts in disease state (Turnbaugh 2009).

In addition, we found the striking result that approximately 70% of the stability in the gut metatranscriptome can be explained by stability in the gut metagenome. This is best illustrated by our finding that some of the most highly stable (as well as highly abundant and prevalent) species, such as *Bacteroides uniformis*, were responsible for the dominant expression of a large number of stable genes. In contrast, *Streptococcus salivarius*, an unstable species that is likely

transiting from the oral microbiome, dominantly expresses many of the most variable genes (Franzosa 2014). Similarly findings were found in relation to regulation of gene expression. Of those genes we identified as significantly differentially regulated, 80% were consistently over- or underexpressed. Many of these stably expressed pathways appeared to be essential for cellular function, involved in glycolysis, synthesis of amino acids, and production of short-chain fatty acids. Taken together, it appears that much of the metatranscriptomic variation in the gut over time might reflect more enduring changes, such as proliferation of bacteria which leads to increases in DNA levels, introduction of new species, or host-induced shifts in community structures such as that induced by bowel preparation.

Of course, we did observe exceptions to this pattern, which may additionally, albeit to a lesser extent, explain some of the instability in the function of the gut microbiome. Expression of many genes, such as macrolide resistance, superoxide reductase, and thiaminase I&II, were not concordant with genomic stability, prevalence, or abundance but instead suggested short-term responses to environmental conditions. In addition, we observed that some species, such as *Sutterella wadsworthensis* and *Methanobrevibacter smithii*, despite their stable expression, appear to be dominantly contributing to expression of unstable genes, suggesting their specialized role associated with variable activity over time. These results are consistent with experimental evidence demonstrating that the transcriptome may be modulated in the short term by antibiotics or specific diets despite minimal changes in species composition (McNulty 2011, Maurice 2013). These outlying genes with high genomic prevalence but low abundance and unstable expression levels, or the stable bacteria leading to expression of unstable genes, are likely to play a critical role in understanding the impact of host behavior on the gut microbiome, and in deciphering the effect of microbiome on host phenotype. Identification of factors that either effect durable changes in the transcriptome or lead to more transient instability in gene expression is essential in understanding how to manipulate the microbiome for health purposes.

Our finding that gut microbial communities and metagenomes reveal a diversity-stability relationship echoes observations from recent studies (Flores 2014) and mirrors patterns observed in several other macro-ecosystems, where species-rich communities allow for greater compensation for disturbance in the ecosystem (McCann 2000; Schindler 2010). Based on prior work, we expected to see a more consistent effect of antibiotic initation on instability (Modi 2014, Jakobsson 2010, Dethefleson 2011, Perez-Cobas 2013, Maurice 2013). Our mixed results

could indicate that individuals' microbial community structures rebounded since antibiotic use (Dethlefson 2011). Lastly, while there has been inconsistent data on the effects of bowel preparation on the gut microbiome composition, it appears in this cohort it has a profound effect, particularly on the metagenome and metatranscriptome (Jalanka 2014, O'Brien 2013).

Finally, these results from our study are consistent with our prior work and underscore that a single self-collected stool sample can adequately represent species and metagenomic relative abundance to generate biomarkers to investigate exposure-disease relationships over at least intermediate-term follow-up (HMP1, Faith 2013, Caporaso 2011). Although our results suggest that transcriptomic data are less stable over time, our data demonstrate that ascertainment of metatranscriptomic data in combination with metagenomic data better characterizes the functional activity of the gut microbiome. As we have shown previously, transcriptomic levels may differ substantially from genomic data for a subset of enzymes (Franzosa 2014). Further work characterizing the microbiome in relation to environmental data collected in relation to stool collection will also be essential in mechanistically understanding the interactions between the gut microbiome and the host. Continuing to improve our understanding of the stability of the gut microbiome function is critical not only in determining the ecological dynamics of the human microbiome but also for the promotion of human health.

Discussion and Afterword

Through investigation leveraging decades of prospectively collected data from the Nurses' Health Study and the Health Professionals' Follow-up Study as well as molecular data from stool, plasma and CRC tissue, this thesis has 1) identified reliable and robust biomarkers for CRC; 2) investigated the causal mechanisms by which diet promotes and prevents CRC development; and 3) identified potential markers by which we can stratify participants for personalized interventions. Within the first two substudies comprising this thesis, we found that high plasma levels of MIC1 (GDF15) measured years before the diagnosis of CRC are associated with a greater risk of CRC as well as a greater risk of CRC-specific mortality, particularly in individuals with PTGS2-positive tumors. Our results also suggest that Aspirin/NSAID use lowers the risk of PTGS2-positive cancers, particularly among individuals with high levels of circulating MIC-1. In the third substudy, we found that higher Prudent dietary pattern scores as well as greater intake of vegetables and dietary fiber are associated with lower risk for Fuso(+) CRC tumors but not Fuso(-) CRC tumors, suggesting a potential role of the microbiome in mediating the association of diet with CRC. Finally, in the fourth substudy, in one of largest cohorts characterized metagenomically and metatranscriptomically to date, we described the major underpinnings of within- and between individual variation in gene expression in the gut microbiome. Human gut microbiome organismal composition and metagenomic profiles, but not metatranscriptomic data, remain highly personalized over time. These results suggest that prospective characterization of the gut microbiome at a single timepoint will provide useful information for long-term follow-up for incidence of chronic disease, like CRC. Taken together, the results presented here highlight the importance of taking an integrated molecular epidemiological approach to CRC. In addition, they provide additional evidence for the role of chronic inflammation in the development and progression of CRC. Finally, they underscore the need for future prospective studies examining the association of the gut microbiome as characterized by stool specimens to understand the complex intersection of diet, the gut microbiome, and colorectal carcinogenesis.

There are several strengths common to all of these studies. First, these are based on large prospective studies with high follow-up rates, limiting potential recall bias and allowing us to understand the temporal relationships between exposures and CRC development. Second, we were able to control for a wide array of confounders in each of our analyses, ranging from known

risk factors for CRC development, other markers of inflammation, and passing filter reads in each stool sample. Finally, these data offer the advantage of being relevant to humans, unlike animal or other experimental models which often require extrapolation to people.

We also acknowledge that there are limitations to our research. Most have reviewed in the commentary above, but they are similar to those of traditional epidemiology. Our studies are observational and despite adjustment for confounders, we cannot exclude residual confounding. We also were unable to obtain tumor tissue from all cases of confirmed CRC detected in the two cohorts, but the risk factors in these cases did not appreciably differ from those in cases with tumor tissue available. In addition, routine processing of blood, tumor tissue, and stool may have influenced the sensitivity of our assays. However, in a limited validation study, we detected *F. nucleatum* in paired FFPE and frozen tissue specimens in 2 colorectal carcinoma cases by qPCR; for MIC-1 we estimated the coefficient of reliability between two drift batches to be 0.96 (95% CI 0.87-0.99), suggesting excellent reproducibility in the assay's performance; and a previous pilot study by our group has shown that species, gene, and transcript abundances were highly concordant across fresh-frozen stool, ethanol-based stool, RNAlater-based stool (Franzosa 2014). Finally, because our participants were all health professionals and are mostly white, our study might not be generalizable to other populations.

Looking forward, integrated molecular epidemiological studies such as these offer tremendous opportunities to reduce CRC incidence and mortality. Major advancements have been made by traditional epidemiology for CRC prevention, but renewed efforts are needed. Molecular epidemiology has the potential to find preclinical signs of CRC development as well identify individuals with an increased susceptibility to CRC (Spitz 2010). Furthermore, integrated molecular epidemiological studies will allow us to better understand how a particular exposure influences the development of CRC and may reveal new relationships between risk factors and molecular subtypes of CRC (Ogino 2011). Finally, as increasing evidence supports the notion that CRC is a collection of many different diseases with unique genetic signatures, integrated molecular epidemiology studies may allow us to meet the call for personalized prevention and treatment of cancer (Collins 2015).

Our work also adds to the growing body of literature supporting a role of the gut microbiome and chronic inflammation in colorectal carcinogenesis (Garrett 2015). Our final substudy, in particular, may lay the groundwork for future prospective molecular

epidemiological studies to explore the intersection of the gut microbiome, chronic inflammation, and known risk factors for CRC. Several studies have hypothesized that the intestinal microflora may represent the link between inflammation and CRC tumorigenesis (Elinav 2013). Thus, there are many unanswered questions about which microbes are the drivers of tumorigenesis and the mechanisms by which they promote inflammation and CRC development. In addition, we are entering a new era in which these microbial protagonists of CRC can be prevented with dietary and probiotic interventions, and ultimately, how they can be eliminated.

Abbreviations:

BC, Bray-Curtis Score; CI, Confidence Interval; CV, coefficient of variance; COX-2, cyclooxygenase-2; CRC, colorectal cancer; CRP, C-reactive protein; ELISA, Enzyme-linked Immunosorbent Assay; FFQ, food frequency questionnaire; Fuso(+), colorectal tumors enriched with *Fusobacterium nucleatum;* Fuso(-), colorectal tumros without *Fusobacterium nucleatum;* HPFS, Health Professionals Follow-up Study; HR, Hazards Ratio; IL6, interleukin 6; MLVS, Men's Lifestyle Validation Study; MIC-1, macrophage inhibitory cytokine-1; NHS, Nurses' Health Study; PTGS2, prostaglandin-endoperoxide synthase 2; RR, relative risk; sTNFR-2, soluble tumor necrosis factor receptor 2; TGF, transforming growth factor

Acknowledgements:

Although only my name appears on this thesis, several people have contributed to its completion. I owe my gratitude to all those people who have made this thesis possible, which in many ways has defined my undergraduate medical education. My deepest gratitude is to my mentor, Dr. Andrew T. Chan. I have been incredibly fortunate to have an advisor who has given me unparalleled research opportunities, the freedom to explore on my own, endless and immediate feedback, and close guidance to help me recover when I found myself in challenging situations. He has shown me how to ask scientific questions, how to think like an epidemiologist, and how to communicate findings succinctly and effectively. Dr. Chan has also been an incredible role model and I hope that one day, I may follow in his footsteps, not just in terms of his scientific career, but also in his love of teaching, his passion for clinical medicine, and his dedication to his family and personal life. I am also thankful to Dr. Mingyang Song who is not only an inspiringly prolific molecular epidemiologist, but also a generous friend and patient teacher. Most importantly, none of this would have been possible without the love and support from Geneva and my family. Finally, I would like to acknowledge the funding I received from the Howard Hughes Medical Institute (HHMI), the American Gastroenterological Association (AGA), Horizon Pharma, the Eli & Edythe Broad Foundation, and the Harvard Medical School Scholars in Medicine Office.

References

Ahn, Jiyoung, Rashmi Sinha, Zhiheng Pei, Christine Dominianni, Jing Wu, Jianxin Shi, James J. Goedert, Richard B. Hayes, and Liying Yang. "Human Gut Microbiome and Risk of Colorectal Cancer." *Journal of the National Cancer Institute*, December 6, 2013, djt300. doi:10.1093/jnci/djt300.

Aune, Dagfinn, Doris S. M. Chan, Rosa Lau, Rui Vieira, Darren C. Greenwood, Ellen Kampman, and Teresa Norat. "Dietary Fibre, Whole Grains, and Risk of Colorectal Cancer: Systematic Review and Dose-Response Meta-Analysis of Prospective Studies." *BMJ* 343 (November 10, 2011): d6617. doi:10.1136/bmj.d6617.

Aune, Dagfinn, Rosa Lau, Doris S. M. Chan, Rui Vieira, Darren C. Greenwood, Ellen Kampman, and Teresa Norat. "Nonlinear Reduction in Risk for Colorectal Cancer by Fruit and Vegetable Intake Based on Meta-Analysis of Prospective Studies." *Gastroenterology* 141, no. 1 (July 2011): 106–18. doi:10.1053/j.gastro.2011.04.013.

Baek SJ, Okazaki R, Lee S-H, et al. Nonsteroidal Anti-Inflammatory Drug-Activated Gene-1 Over Expression in Transgenic Mice Suppresses Intestinal Neoplasia. *Gastroenterology*. 2006;131(5):1553–1560.

Bauskin AR, Brown DA, Kuffner T, et al. Role of macrophage inhibitory cytokine-1 in tumorigenesis and diagnosis of cancer. *Cancer Res*. 2006;66(10):4983–4986.

Belkina, Anna C., and Gerald V. Denis. "BET Domain Co-Regulators in Obesity, Inflammation and Cancer." *Nature Reviews Cancer* 12, no. 7 (July 2012): 465–77. doi:10.1038/nrc3256.

Bernstein CN, Blanchard JF, Kliewer E, et al. Cancer risk in patients with inflammatory bowel disease: a population-based study. *Cancer*. 2001;91(4):854–862.

Blumberg, Richard, and Fiona Powrie. "Microbiota, Disease, and Back to Health: A Metastable Journey." *Science Translational Medicine* 4, no. 137 (June 6, 2012): 137rv7–137rv7. doi:10.1126/scitranslmed.3004184.

Booijink, Carien C. G. M., Jos Boekhorst, Erwin G. Zoetendal, Hauke Smidt, Michiel Kleerebezem, and Willem M. de Vos. "Metatranscriptome Analysis of the Human Fecal Microbiota Reveals Subject-Specific Expression Profiles, with Genes Encoding Proteins Involved in Carbohydrate Metabolism Being Dominantly Expressed." *Applied and Environmental Microbiology* 76, no. 16 (August 2010): 5533–40. doi:10.1128/AEM.00502-10.

Bouvard, Véronique, Dana Loomis, Kathryn Z Guyton, Yann Grosse, Fatiha El Ghissassi, Lamia Benbrahim-Tallaa, Neela Guha, Heidi Mattock, and Kurt Straif. "Carcinogenicity of Consumption of Red and Processed Meat." *The Lancet Oncology* 16, no. 16 (December 2015): 1599–1600. doi:10.1016/S1470-2045(15)00444-1.

Breit SN, Johnen H, Cook AD, et al. The TGF-β superfamily cytokine, MIC-1/GDF15: a pleotrophic cytokine with roles in inflammation, cancer and metabolism. Growth Factors 2011; 29:187–195. Author names in bold designate shared co-first authors.

Brown DA, Breit SN, Buring J, et al. Concentration in plasma of macrophage inhibitory cytokine-1 and risk of cardiovascular events in women: a nested case-control study. *Lancet*. 2002;359(9324):2159–2163.

Brown DA, Hance KW, Rogers CJ, et al. Serum Macrophage Inhibitory Cytokine-1 (MIC-1/GDF15): A Potential Screening Tool for the Prevention of Colon Cancer? Cancer Epidem Biomar Prev 2012; 21:337–346. Author names in bold designate shared co-first authors.

Brown DA, Lindmark F, Stattin P, et al. Macrophage inhibitory cytokine 1: a new prognostic marker in prostate cancer. Clin. Cancer Res 2009; 15:6658–6664.

Brown DA, Ward RL, Buckhaults P, et al. MIC-1 serum level and genotype: associations with progress and prognosis of colorectal carcinoma. *Clin Cancer Res*. 2003;9(7):2642–2650.

Bussiere, D. E., S. W. Muchmore, C. G. Dealwis, G. Schluckebier, V. L. Nienaber, R. P. Edalji, K. A. Walter, U. S. Ladror, T. F. Holzman, and C. Abad-Zapatero. "Crystal Structure of ErmC', an rRNA Methyltransferase Which Mediates Antibiotic Resistance in Bacteria." *Biochemistry* 37, no. 20 (May 19, 1998): 7103–12. doi:10.1021/bi973113c.

Caporaso, J. Gregory, Christian L. Lauber, Elizabeth K. Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, et al. "Moving Pictures of the Human Microbiome." *Genome Biology* 12, no. 5 (2011): R50. doi:10.1186/gb-2011-12-5-r50.

Caspi, Ron, Hartmut Foerster, Carol A. Fulcher, Pallavi Kaipa, Markus Krummenacker, Mario Latendresse, Suzanne Paley, et al. "The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/Genome Databases." *Nucleic Acids Research* 36, no. suppl 1 (January 1, 2008): D623–31. doi:10.1093/nar/gkm900.

Castellarin, Mauro, René L. Warren, J. Douglas Freeman, Lisa Dreolini, Martin Krzywinski, Jaclyn Strauss, Rebecca Barnes, et al. "Fusobacterium Nucleatum Infection Is Prevalent in Human Colorectal Carcinoma." *Genome Research*, October 18, 2011. doi:10.1101/gr.126516.111.

Chan AT, Giovannucci EL, Meyerhardt JA, et al. Aspirin dose and duration of use and risk of colorectal cancer in men. *Gastroenterology*. 2008;134(1):21–28.

References

Chan AT, Giovannucci EL, Meyerhardt JA, et al. Long-term use of aspirin and nonsteroidal anti-inflammatory drugs and risk of colorectal cancer. *JAMA*. 2005;294(8):914–923.

Chan AT, Ogino S, Fuchs CS. Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. *N Engl J Med*. 2007;356(21):2131–2142.

Chan AT, Ogino S, Fuchs CS. Aspirin use and survival after diagnosis of colorectal cancer. *JAMA*. 2009;302(6):649–658.

Chan AT, Ogino S, Giovannucci EL, et al. Inflammatory markers are associated with risk of colorectal cancer and chemopreventive response to anti-inflammatory drugs. *Gastroenterology*. 2011;140(3):799–808.

Chan, Andrew T., and Edward L. Giovannucci. "Primary Prevention of Colorectal Cancer." *Gastroenterology* 138, no. 6 (June 2010): 2029–43.e10. doi:10.1053/j.gastro.2010.01.057.

Chatelier, Emmanuelle Le, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, et al. "Richness of Human Gut Microbiome Correlates with Metabolic Markers." *Nature* 500, no. 7464 (August 29, 2013): 541–46. doi:10.1038/nature12506.

Chen, Hui-Min, Ya-Nan Yu, Ji-Lin Wang, Yan-Wei Lin, Xuan Kong, Chang-Qing Yang, Li Yang, et al. "Decreased Dietary Fiber Intake and Structural Alteration of Gut Microbiota in Patients with Advanced Colorectal Adenoma." *The American Journal of Clinical Nutrition* 97, no. 5 (May 1, 2013): 1044–52. doi:10.3945/ajcn.112.046607.

Claesson, Marcus J., Ian B. Jeffery, Susana Conde, Susan E. Power, Eibhlís M. O'Connor, Siobhán Cusack, Hugh M. B. Harris, et al. "Gut Microbiota Composition Correlates with Diet and Health in the Elderly." *Nature* 488, no. 7410 (August 9, 2012): 178–84. doi:10.1038/nature11319.

Clarke R, Shipley M, Lewington S, et al. Underestimation of Risk Associations Due to Regression Dilution in Long-term Follow-up of Prospective Studies. *Am J Epidemiol*. 1999;150(4):341–353.

Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954*;10:*101–29.

Collins, Francis S., and Harold Varmus. "A New Initiative on Precision Medicine." *New England Journal of Medicine* 372, no. 9 (February 26, 2015): 793–95. doi:10.1056/NEJMp1500523.

Cooper, Geoffrey M. *The Cell*. 2nd ed. Sinauer Associates, 2000.

Costello, Elizabeth K., Christian L. Lauber, Micah Hamady, Noah Fierer, Jeffrey I. Gordon, and Rob Knight. "Bacterial Community Variation in Human Body Habitats Across Space and Time." *Science (New York, N.Y.)* 326, no. 5960 (December 18, 2009): 1694–97. doi:10.1126/science.1177486.

Cotillard, Aurélie, Sean P. Kennedy, Ling Chun Kong, Edi Prifti, Nicolas Pons, Emmanuelle Le Chatelier, Mathieu Almeida, et al. "Dietary Intervention Impact on Gut Microbial Gene Richness." *Nature* 500, no. 7464 (August 29, 2013): 585–88. doi:10.1038/nature12480.

Cottet, V., C. Bonithon-Kopp, O. Kronborg, L. Santos, R. Andreatta, M.-C. Boutron-Ruault, J. Faivre, and European Cancer Prevention Organisation Study Group. "Dietary Patterns and the Risk of Colorectal Adenoma Recurrence in a European Intervention Trial." *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)* 14, no. 1 (February 2005): 21–29.

Coyte, Katharine Z., Jonas Schluter, and Kevin R. Foster. "The Ecology of the Microbiome: Networks, Competition, and Stability." *Science* 350, no. 6261 (November 6, 2015): 663–66. doi:10.1126/science.aad2602.

David, Lawrence A., Arne C. Materna, Jonathan Friedman, Maria I. Campos-Baptista, Matthew C. Blackburn, Allison Perrotta, Susan E. Erdman, and Eric J. Alm. "Host Lifestyle Affects Human Microbiota on Daily Timescales." *Genome Biology* 15, no. 7 (July 25, 2014): R89. doi:10.1186/gb-2014-15-7-r89.

David, Lawrence A., Corinne F. Maurice, Rachel N. Carmody, David B. Gootenberg, Julie E. Button, Benjamin E. Wolfe, Alisha V. Ling, et al. "Diet Rapidly and Reproducibly Alters the Human Gut Microbiome." *Nature* advance online publication (December 11, 2013). doi:10.1038/nature12820.

Dethlefsen, Les, and David A. Relman. "Incomplete Recovery and Individualized Responses of the Human Distal Gut Microbiota to Repeated Antibiotic Perturbation." *Proceedings of the National Academy of Sciences* 108, no. Supplement 1 (March 15, 2011): 4554–61. doi:10.1073/pnas.1000087107.

Ding, Tao, and Patrick D. Schloss. "Dynamics and Associations of Microbial Community Types across the Human Body." *Nature* 509, no. 7500 (May 15, 2014): 357–60. doi:10.1038/nature13178.

Elinav, Eran, Roni Nowarski, Christoph A. Thaiss, Bo Hu, Chengcheng Jin, and Richard A. Flavell. "Inflammation-Induced Cancer: Crosstalk between Tumours, Immune Cells and Microorganisms." *Nature Reviews Cancer* 13, no. 11 (November 2013): 759–71. doi:10.1038/nrc3611.

Faith, Jeremiah J., Janaki L. Guruge, Mark Charbonneau, Sathish Subramanian, Henning Seedorf, Andrew L. Goodman, Jose C. Clemente, et al. "The Long-Term Stability of the Human Gut Microbiota." *Science* 341, no. 6141 (July 5, 2013): 1237439. doi:10.1126/science.1237439.

References

Filippis, Francesca De, Nicoletta Pellegrini, Lucia Vannini, Ian B. Jeffery, Antonietta La Storia, Luca Laghi, Diana I. Serrazanetti, et al. "High-Level Adherence to a Mediterranean Diet Beneficially Impacts the Gut Microbiota and Associated Metabolome." *Gut*, September 28, 2015, gutjnl – 2015–309957. doi:10.1136/gutjnl-2015-309957.

Flores, Gilberto E., J. G. Caporaso, Jessica B. Henley, Jai R. Rideout, Daniel Domogala, John Chase, Jonathan W. Leff, et al. "Temporal Variability Is a Personalized Feature of the Human Microbiome." *Genome Biology* 15, no. 12 (December 3, 2014): 531. doi:10.1186/s13059-014-0531-y.

Foxman, Betsy, and Emily T. Martin. "Use of the Microbiome in the Practice of Epidemiology: A Primer on -Omic Technologies." *American Journal of Epidemiology* 182, no. 1 (July 1, 2015): 1–8. doi:10.1093/aje/kwv102.

Franzosa, Eric A., Xochitl C. Morgan, Nicola Segata, Levi Waldron, Joshua Reyes, Ashlee M. Earl, Georgia Giannoukos, et al. "Relating the Metatranscriptome and Metagenome of the Human Gut." *Proceedings of the National Academy of Sciences of the United States of America* 111, no. 22 (June 3, 2014): E2329–38. doi:10.1073/pnas.1319284111.

Fung, Teresa, Frank B. Hu, Charles Fuchs, Edward Giovannucci, David J. Hunter, Meir J. Stampfer, Graham A. Colditz, and Walter C. Willett. "Major Dietary Patterns and the Risk of Colorectal Cancer in Women." *Archives of Internal Medicine* 163, no. 3 (2003): 309–14.

Garrett, Wendy S. "Cancer and the Microbiota." *Science* 348, no. 6230 (April 3, 2015): 80–86. doi:10.1126/science.aaa4972.

Giongo, Adriana, Kelsey A. Gano, David B. Crabb, Nabanita Mukherjee, Luis L. Novelo, George Casella, Jennifer C. Drew, et al. "Toward Defining the Autoimmune Microbiome for Type 1 Diabetes." *The ISME Journal* 5, no. 1 (January 2011): 82–91. doi:10.1038/ismej.2010.92.

Gur, Chamutal, Yara Ibrahim, Batya Isaacson, Rachel Yamin, Jawad Abed, Moriya Gamliel, Jonatan Enk, et al. "Binding of the Fap2 Protein of Fusobacterium Nucleatum to Human Inhibitory Receptor TIGIT Protects Tumors from Immune Cell Attack." *Immunity* 42, no. 2 (February 17, 2015): 344–55.

Haggar, Fatima A., and Robin P. Boushey. "Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors." *Clinics in Colon and Rectal Surgery* 22, no. 4 (November 2009): 191–97.

Hankinson SE, Willett WC, Manson JE, et al. Alcohol, height, and adiposity in relation to estrogen and prolactin levels in postmenopausal women. *J Natl Cancer Inst*. 1995;87(17):1297–1302.

Hsiao, Elaine Y., Sara W. McBride, Sophia Hsien, Gil Sharon, Embriette R. Hyde, Tyler McCue, Julian A. Codelli, et al. "Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders." *Cell* 155, no. 7 (December 19, 2013): 1451–63. doi:10.1016/j.cell.2013.11.024.

Hu, F. B., E. Rimm, S. A. Smith-Warner, D. Feskanich, M. J. Stampfer, A. Ascherio, L. Sampson, and W. C. Willett. "Reproducibility and Validity of Dietary Patterns Assessed with a Food-Frequency Questionnaire." *The American Journal of Clinical Nutrition* 69, no. 2 (February 1999): 243–49.

Huycke, Mark M., Victoria Abrams, and Danny R. Moore. "Enterococcus Faecalis Produces Extracellular Superoxide and Hydrogen Peroxide That Damages Colonic Epithelial Cell DNA." *Carcinogenesis* 23, no. 3 (March 1, 2002): 529–36. doi:10.1093/carcin/23.3.529.

Ito, Miki, Shinichi Kanno, Katsuhiko Nosho, Yasutaka Sukawa, Kei Mitsuhashi, Hiroyoshi Kurihara, Hisayoshi Igarashi, et al. "Association of Fusobacterium Nucleatum with Clinical and Molecular Features in Colorectal Serrated Pathway." *International Journal of Cancer*, February 1, 2015, n/a – n/a. doi:10.1002/ijc.29488.

Jakobsson, Hedvig E., Cecilia Jernberg, Anders F. Andersson, Maria Sjölund-Karlsson, Janet K. Jansson, and Lars Engstrand. "Short-Term Antibiotic Treatment Has Differing Long-Term Impacts on the Human Throat and Gut Microbiome." *PLoS ONE* 5, no. 3 (March 24, 2010): e9836. doi:10.1371/journal.pone.0009836.

Jalanka, Jonna, Anne Salonen, Jarkko Salojärvi, Jarmo Ritari, Outi Immonen, Luca Marciani, Penny Gowland, et al. "Effects of Bowel Cleansing on the Intestinal Microbiota." *Gut*, December 19, 2014, gutjnl – 2014–307240.

Jeffery, Ian B., Denise B. Lynch, and Paul W. O'Toole. "Composition and Temporal Stability of the Gut Microbiota in Older Persons." *The ISME Journal* 10, no. 1 (January 2016): 170–82. doi:10.1038/ismej.2015.88.

Jenney, Francis E., Marc F. J. M. Verhagen, Xiaoyuan Cui, and Michael W. W. Adams. "Anaerobic Microbes: Oxygen Detoxification Without Superoxide Dismutase." *Science* 286, no. 5438 (October 8, 1999): 306–9. =

Johnen H, Lin S, Kuffner T, et al. Tumor-induced anorexia and weight loss are mediated by the TGF-β superfamily cytokine MIC-1. Nat Med 2007. 13:1333–1340. Author names in bold designate shared co-first authors.

Kanavos, P. "The Rising Burden of Cancer in the Developing World." *Annals of Oncology* 17, no. suppl 8 (July 1, 2006): viii15–23. doi:10.1093/annonc/mdl983.

References

Kanehisa, Minoru, Susumu Goto, Yoko Sato, Miho Furumichi, and Mao Tanabe. "KEGG for Integration and Interpretation of Large-Scale Molecular Data Sets." *Nucleic Acids Research*, November 10, 2011, gkr988. doi:10.1093/nar/gkr988.

Karapinar, Tolga, Murat Dabak, and Omer Kizil. "Thiamine Status of Feedlot Cattle Fed a High-Concentrate Diet." *The Canadian Veterinary Journal* 51, no. 11 (November 2010): 1251–53.

Kim KK, Lee JJ, Yang Y, et al. Macrophage inhibitory cytokine-1 activates AKT and ERK-1/2 via the transactivation of ErbB2 in human breast and gastric cancer cells. Carcinogenesis 2008, 29:704–712.

Kim, D. H., L. Konishi, and K. Kobashi. "Purification, Characterization and Reaction Mechanism of Novel Arylsulfotransferase Obtained from an Anaerobic Bacterium of Human Intestine." *Biochimica Et Biophysica Acta* 872, no. 1–2 (July 25, 1986): 33–41.

Kleinbaum DG & Klein M. Survival Analysis: A Self-Learning Text. 3rd ed. London: Springer, 2010: 585-588.

Kleinbaum, David, Lawrence Kupper, Azhar Nizam, and Eli Rosenberg. *Applied Regression Analysis and Other Multivariable Methods*. Cengage Learning, 2013.

Koopmann J, Buckhaults P, Brown DA, et al. Serum Macrophage Inhibitory Cytokine 1 as a Marker of Pancreatic and Other Periampullary Cancers. Clin Cancer Res 2004; 10: 2386–92. Author names in bold designate shared co-first authors.

Kostic, Aleksandar D., Dirk Gevers, Chandra Sekhar Pedamallu, Monia Michaud, Fujiko Duke, Ashlee M. Earl, Akinyemi I. Ojesina, et al. "Genomic Analysis Identifies Association of Fusobacterium with Colorectal Carcinoma." *Genome Research* 22, no. 2 (February 2012): 292–98. doi:10.1101/gr.126573.111.

Kostic, Aleksandar D., Eunyoung Chun, Lauren Robertson, Jonathan N. Glickman, Carey Ann Gallini, Monia Michaud, Thomas E. Clancy, et al. "Fusobacterium Nucleatum Potentiates Intestinal Tumorigenesis and Modulates the Tumor-Immune Microenvironment." *Cell Host & Microbe* 14, no. 2 (August 14, 2013): 207–15. doi:10.1016/j.chom.2013.07.007.

Kotsopoulos, Joanne, Shelley S. Tworoger, Hannia Campos, Fung-Lung Chung, Charles V. Clevenger, Adrian A. Franke, Christos S. Mantzoros, et al. "Reproducibility of Plasma, Red Blood Cell, and Urine Biomarkers among Premenopausal and Postmenopausal Women from the Nurses' Health Studies." *Cancer Epidemiology Biomarkers & Prevention* 19, no. 4 (April 1, 2010): 938–46. doi:10.1158/1055-9965.EPI-09-1318.

Lagier, J.-C., F. Armougom, M. Million, P. Hugon, I. Pagnier, C. Robert, F. Bittar, et al. "Microbial Culturomics: Paradigm Shift in the Human Gut Microbiome Study." *Clinical Microbiology and Infection: The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 18, no. 12 (December 2012): 1185–93. doi:10.1111/1469-0691.12023.

Lee, Jung Eun, and Andrew T. Chan. "Fruit, Vegetables, and Folate: Cultivating the Evidence for Cancer Prevention." *Gastroenterology* 141, no. 1 (July 2011): 16–20. doi:10.1053/j.gastro.2011.05.020.

Lewis, S. J., and K. W. Heaton. "Stool Form Scale as a Useful Guide to Intestinal Transit Time." *Scandinavian Journal of Gastroenterology* 32, no. 9 (September 1997): 920–24.

Louis, Petra, Georgina L. Hold, and Harry J. Flint. "The Gut Microbiota, Bacterial Metabolites and Colorectal Cancer." *Nature Reviews Microbiology* 12, no. 10 (October 2014): 661–72. doi:10.1038/nrmicro3344.

Lozupone, Catherine A., Jesse I. Stombaugh, Jeffrey I. Gordon, Janet K. Jansson, and Rob Knight. "Diversity, Stability and Resilience of the Human Gut Microbiota." *Nature* 489, no. 7415 (September 13, 2012): 220–30. doi:10.1038/nature11550.

Magalhães, Bruno, Bárbara Peleteiro, and Nuno Lunet. "Dietary Patterns and Colorectal Cancer: Systematic Review and Meta-Analysis." *European Journal of Cancer Prevention: The Official Journal of the European Cancer Prevention Organisation (ECP)* 21, no. 1 (January 2012): 15–23. doi:10.1097/CEJ.0b013e3283472241.

Malojčić, Goran, Robin L. Owen, John P. A. Grimshaw, Maurice S. Brozzo, Hiang Dreher-Teo, and Rudi Glockshuber. "A Structural and Biochemical Basis for PAPS-Independent Sulfuryl Transfer by Aryl Sulfotransferase from Uropathogenic Escherichia Coli." *Proceedings of the National Academy of Sciences* 105, no. 49 (December 9, 2008): 19217–22. doi:10.1073/pnas.0806997105.

Markowitz, Sanford D., and Monica M. Bertagnolli. "Molecular Basis of Colorectal Cancer." *New England Journal of Medicine* 361, no. 25 (2009): 2449–60.

Maurice, Corinne Ferrier, Henry Joseph Haiser, and Peter James Turnbaugh. "Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome." *Cell* 152, no. 1–2 (January 17, 2013): 39–50. doi:10.1016/j.cell.2012.10.052.

Mazmanian, Sarkis K., June L. Round, and Dennis L. Kasper. "A Microbial Symbiosis Factor Prevents Intestinal Inflammatory Disease." *Nature* 453, no. 7195 (May 29, 2008): 620–25. doi:10.1038/nature07008.

McCann, Kevin Shear. "The Diversity–stability Debate." *Nature* 405, no. 6783 (May 11, 2000): 228–33. doi:10.1038/35012234.

References

McCoy, Amber N., Félix Araújo-Pérez, Andrea Azcárate-Peril, Jen Jen Yeh, Robert S. Sandler, and Temitope O. Keku. "Fusobacterium Is Associated with Colorectal Adenomas." *PLoS ONE* 8, no. 1 (January 15, 2013): e53653. doi:10.1371/journal.pone.0053653.

McNulty, Nathan P., Tanya Yatsunenko, Ansel Hsiao, Jeremiah J. Faith, Brian D. Muegge, Andrew L. Goodman, Bernard Henrissat, et al. "The Impact of a Consortium of Fermented Milk Strains on the Gut Microbiome of Gnotobiotic Mice and Monozygotic Twins." *Science Translational Medicine* 3, no. 106 (October 26, 2011): 106ra106. doi:10.1126/scitranslmed.3002701.

Mehta RS, Song M, Bezawada N, et al. A prospective study of macrophage inhibitory cytokine-1 (MIC-1/GDF15) and risk of colorectal cancer. J Natl Cancer Inst 2014; 106: dju016.

Meyerhardt JA & Mayer RJ. Systemic Therapy for Colorectal Cancer. N Eng J Med 2005; 352: 476–87.

Mills, L. Scott, Michael E. Soulé, and Daniel F. Doak. "The Keystone-Species Concept in Ecology and Conservation." *BioScience*, 1993, 219–24.

Mima, Kosuke, Reiko Nishihara, Zhi Rong Qian, Yin Cao, Yasutaka Sukawa, Jonathan A. Nowak, Juhong Yang, et al. "Fusobacterium Nucleatum in Colorectal Carcinoma Tissue and Patient Prognosis." *Gut*, 2015, gutjnl – 2015.

Mima, Kosuke, Yasutaka Sukawa, Reiko Nishihara, Zhi Rong Qian, Mai Yamauchi, Kentaro Inamura, Sun A. Kim, et al. "Fusobacterium Nucleatum and T Cells in Colorectal Carcinoma." *JAMA Oncology* 1, no. 5 (2015): 653–61.

Mimeault M, Batra SK. Divergent molecular mechanisms underlying the pleiotropic functions of macrophage inhibitory cytokine-1 in cancer. *J Cell Phys*. 2010;224(3):626–635.

Modi, Sheetal R., James J. Collins, and David A. Relman. "Antibiotics and the Gut Microbiota." *Journal of Clinical Investigation* 124, no. 10 (October 1, 2014): 4212–18. doi:10.1172/JCI72333.

Morikawa, Teppei, Aya Kuchiba, Mai Yamauchi, Jeffrey A. Meyerhardt, Kaori Shima, Katsuhiko Nosho, Andrew T. Chan, Edward Giovannucci, Charles S. Fuchs, and Shuji Ogino. "Association of CTNNB1 (beta-Catenin) Alterations, Body Mass Index, and Physical Activity with Survival in Patients with Colorectal Cancer." *JAMA* 305, no. 16 (April 27, 2011): 1685–94. doi:10.1001/jama.2011.513.

Mucci LA, Hjelmborg JB, Harris JR, and et al. "FAmilial Risk and Heritability of Cancer among Twins in Nordic Countries." *JAMA* 315, no. 1 (January 5, 2016): 68–76. doi:10.1001/jama.2015.17703.

Murata, Kiku. "Actions of Two Types of Thiaminase on Thiamin and Its Analogues." *Annals of the New York Academy of Sciences* 378, no. 1 (March 1, 1982): 146–56. doi:10.1111/j.1749-6632.1982.tb31193.x.

Nakagawa, Shinichi, and Holger Schielzeth. "Repeatability for Gaussian and Non-Gaussian Data: A Practical Guide for Biologists." *Biological Reviews* 85, no. 4 (November 1, 2010): 935–56.

Nakatsu, Geicho, Xiangchun Li, Haokui Zhou, Jianqiu Sheng, Sunny Hei Wong, William Ka Kai Wu, Siew Chien Ng, et al. "Gut Mucosal Microbiome across Stages of Colorectal Carcinogenesis." *Nature Communications* 6 (October 30, 2015): 8727. doi:10.1038/ncomms9727.

Nan H, Hutter CM, Lin Y, and et al. "ASsociation of Aspirin and NSAID Use with Risk of Colorectal Cancer according to Genetic Variants." *JAMA* 313, no. 11 (March 17, 2015): 1133–42. doi:10.1001/jama.2015.1815.

O'Brien, Claire L., Gwen E. Allison, Florian Grimpen, and Paul Pavli. "Impact of Colonoscopy Bowel Preparation on Intestinal Microbiota." *PLoS ONE* 8, no. 5 (May 1, 2013): e62815. doi:10.1371/journal.pone.0062815.

O'Keefe, Stephen J. D., Jia V. Li, Leo Lahti, Junhai Ou, Franck Carbonero, Khaled Mohammed, Joram M. Posma, et al. "Fat, Fibre and Cancer Risk in African Americans and Rural Africans." *Nature Communications* 6 (April 28, 2015): 6342. doi:10.1038/ncomms7342.

Ogino S, Brahmandam M, Kawasaki T, et al. Combined Analysis of COX-2 and p53 Expressions Reveals Synergistic Inverse Correlations with Microsatellite Instability and CpG Island Methylator Phenotype in Colorectal Cancer. Neoplasia 2006; 6:458–64.

Ogino S, Chan AT, Fuchs CS, Giovannucci E. Molecular Pathologic Epidemiology of Colorectal Neoplasia: An Emerging Transdisciplinary and Interdisciplinary Field. *Gut* 2011. 60:397-411.

Ou, Junhai, Franck Carbonero, Erwin G. Zoetendal, James P. DeLany, Mei Wang, Keith Newton, H. Rex Gaskins, and Stephen J. D. O'Keefe. "Diet, Microbiota, and Microbial Metabolites in Colon Cancer Risk in Rural Africans and African Americans." *The American Journal of Clinical Nutrition* 98, no. 1 (July 2013): 111–20. doi:10.3945/ajcn.112.056689.

Park SH, Choi HJ, Yang H, et al. Two in-and-out modulation strategies for endoplasmic reticulum stress-linked gene expression of pro-apoptotic macrophage-inhibitory cytokine 1. J Biol Chem 2012; 287:19841–19855. Author names in bold designate shared co-first authors.

References

Pérez-Cobas, Ana Elena, María José Gosalbes, Anette Friedrichs, Henrik Knecht, Alejandro Artacho, Kathleen Eismann, Wolfgang Otto, et al. "Gut Microbiota Disturbance during Antibiotic Therapy: A Multi-Omic Approach." *Gut*, December 12, 2012, gutjnl – 2012–303184. doi:10.1136/gutjnl-2012-303184.

Rajilić-Stojanović, Mirjana, Hans G. H. J. Heilig, Sebastian Tims, Erwin G. Zoetendal, and Willem M. de Vos. "Long-Term Monitoring of the Human Intestinal Microbiota Composition." *Environmental Microbiology*, October 15, 2012. doi:10.1111/1462-2920.12023.

Rimm, E. B., E. L. Giovannucci, M. J. Stampfer, G. A. Colditz, L. B. Litin, and W. C. Willett. "Reproducibility and Validity of an Expanded Self-Administered Semiquantitative Food Frequency Questionnaire among Male Health Professionals." *American Journal of Epidemiology* 135, no. 10 (May 15, 1992): 1114–26; discussion 1127–36.

Rosner, Bernard. *Fundamentals of Biostatistics*. Cengage Learning, 2010.

Rubinstein, Mara Roxana, Xiaowei Wang, Wendy Liu, Yujun Hao, Guifang Cai, and Yiping W. Han. "Fusobacterium Nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via Its FadA Adhesin." *Cell Host & Microbe* 14, no. 2 (August 14, 2013): 195–206. doi:10.1016/j.chom.2013.07.012.

Schindler, Daniel E., Ray Hilborn, Brandon Chasco, Christopher P. Boatright, Thomas P. Quinn, Lauren A. Rogers, and Michael S. Webster. "Population Diversity and the Portfolio Effect in an Exploited Species." *Nature* 465, no. 7298 (June 3, 2010): 609–12. doi:10.1038/nature09060.

Schloissnig, Siegfried, Manimozhiyan Arumugam, Shinichi Sunagawa, Makedonka Mitreva, Julien Tap, Ana Zhu, Alison Waller, et al. "Genomic Variation Landscape of the Human Gut Microbiome." *Nature* 493, no. 7430 (January 3, 2013): 45–50. doi:10.1038/nature11711.

Sears, Cynthia L., and Wendy S. Garrett. "Microbes, Microbiota, and Colon Cancer." *Cell Host & Microbe* 15, no. 3 (March 12, 2014): 317–28. doi:10.1016/j.chom.2014.02.007.

Shnaper S, Desbaillets I, Brown DA, et al. Elevated levels of MIC-1/GDF15 in the cerebrospinal fluid of patients are associated with glioblastoma and worse outcome. Int J Cancer 2009; 125:2624–2630.

Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. "Cancer Statistics, 2015." *CA: A Cancer Journal for Clinicians* 65, no. 1 (January 1, 2015): 5–29. doi:10.3322/caac.21254.

Smith, Patrick M., Michael R. Howitt, Nicolai Panikov, Monia Michaud, Carey Ann Gallini, Mohammad Bohlooly-Y, Jonathan N. Glickman, and Wendy S. Garrett. "The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic Treg Cell Homeostasis." *Science (New York, N.Y.)* 341, no. 6145 (August 2, 2013): 569–73. doi:10.1126/science.1241165.

Song M, Wu K, Ogino S, et al. A prospective study of plasma inflammatory markers and risk of colorectal cancer in men. *Br J Cancer*. 2013;108(9):1891–1898.

Song, M., K. Wu, S. Ogino, C. S. Fuchs, E. L. Giovannucci, and A. T. Chan. "A Prospective Study of Plasma Inflammatory Markers and Risk of Colorectal Cancer in Men." *British Journal of Cancer* 108, no. 9 (March 14, 2013): 1891–98. doi:10.1038/bjc.2013.172.

Song, Mingyang, Raaj S. Mehta, Kana Wu, Charles S. Fuchs, Shuji Ogino, Edward L. Giovannucci, and Andrew T. Chan. "Plasma Inflammatory Markers and Risk of Advanced Colorectal Adenoma in Women." *Cancer Prevention Research* 9, no. 1 (2016): 27–34.

Song, Mingyang, Wendy S. Garrett, and Andrew T. Chan. "Nutrients, Foods, and Colorectal Cancer Prevention." *Gastroenterology*, Food, the Immune System, and the Gastrointestinal Tract, 148, no. 6 (May 2015): 1244–60.e16. doi:10.1053/j.gastro.2014.12.035.

Sonnenburg, Erica D., Samuel A. Smits, Mikhail Tikhonov, Steven K. Higginbottom, Ned S. Wingreen, and Justin L. Sonnenburg. "Diet-Induced Extinctions in the Gut Microbiota Compound over Generations." *Nature* 529, no. 7585 (January 14, 2016): 212–15. doi:10.1038/nature16504.

Sonnenburg, Erica D., Samuel A. Smits, Mikhail Tikhonov, Steven K. Higginbottom, Ned S. Wingreen, and Justin L. Sonnenburg. "Diet-Induced Extinctions in the Gut Microbiota Compound over Generations." *Nature* 529, no. 7585 (January 14, 2016): 212–15. doi:10.1038/nature16504.

Staff A, Bock A, Becker C, et al. Growth differentiation factor-15 as a prognostic biomarker in ovarian cancer. Gynecol Oncol 2010; 118:237–243.

Steidler, Lothar, Wolfgang Hans, Lieven Schotte, Sabine Neirynck, Florian Obermeier, Werner Falk, Walter Fiers, and Erik Remaut. "Treatment of Murine Colitis by Lactococcus Lactis Secreting Interleukin-10." *Science* 289, no. 5483 (August 25, 2000): 1352–55.

Tahara, Tomomitsu, Eiichiro Yamamoto, Hiromu Suzuki, Reo Maruyama, Woonbok Chung, Judith Garriga, Jaroslav Jelinek, et al. "Fusobacterium in Colonic Flora and Molecular Features of Colorectal Carcinoma." *Cancer Research* 74, no. 5 (March 1, 2014): 1311–18. doi:10.1158/0008-5472.CAN-13-1865.

References

Terry, P., F. B. Hu, H. Hansen, and A. Wolk. “Prospective Study of Major Dietary Patterns and Colorectal Cancer Risk in Women.” *American Journal of Epidemiology* 154, no. 12 (December 15, 2001): 1143–49.

Tjalsma, Harold, Annemarie Boleij, Julian R. Marchesi, and Bas E. Dutilh. “A Bacterial Driver–passenger Model for Colorectal Cancer: Beyond the Usual Suspects.” *Nature Reviews Microbiology* 10, no. 8 (August 2012): 575–82. doi:10.1038/nrmicro2819.

Tjalsma, Harold, Annemarie Boleij, Julian R. Marchesi, and Bas E. Dutilh. “A Bacterial Driver–passenger Model for Colorectal Cancer: Beyond the Usual Suspects.” *Nature Reviews Microbiology* 10, no. 8 (August 2012): 575–82. doi:10.1038/nrmicro2819.

Truong, Duy Tin, Eric A. Franzosa, Timothy L. Tickle, Matthias Scholz, George Weingart, Edoardo Pasolli, Adrian Tett, Curtis Huttenhower, and Nicola Segata. “MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling.” *Nature Methods* 12, no. 10 (October 2015): 902–3.

Turnbaugh, Peter J., Christopher Quince, Jeremiah J. Faith, Alice C. McHardy, Tanya Yatsunenko, Faheem Niazi, Jason Affourtit, et al. “Organismal, Genetic, and Transcriptional Variation in the Deeply Sequenced Gut Microbiomes of Identical Twins.” *Proceedings of the National Academy of Sciences of the United States of America* 107, no. 16 (April 20, 2010): 7503–8.

Wang D & R. N. DuBois. The Role of COX-2 in Intestinal Inflammation and Colorectal Cancer. Oncogene 2010; 29:781–88.

Wang X, Baek SJ, Eling TE. The diverse roles of nonsteroidal anti-inflammatory drug activated gene (NAG-1/GDF15) in cancer. *Biochem Pharmacol*. 2013;85(5):597–606.

Wang X, Kingsley PJ, Marnett LJ, et al. The role of NAG-1/GDF15 in the inhibition of intestinal polyps in APC/Min mice by sulindac. *Cancer Prev Res*. 2011;4(1):150–160.

Wang, Molin, Donna Spiegelman, Aya Kuchiba, Paul Lochhead, Sehee Kim, Andrew T. Chan, Elizabeth M. Poole, et al. “Statistical Methods for Studying Disease Subtype Heterogeneity.” *Statistics in Medicine*, December 1, 2015. doi:10.1002/sim.6793.

Warren, René L., Douglas J. Freeman, Stephen Pleasance, Peter Watson, Richard A. Moore, Kyla Cochrane, Emma Allen-Vercoe, and Robert A. Holt. “Co-Occurrence of Anaerobic Bacteria in Colorectal Carcinomas.” *Microbiome* 1 (2013): 16. doi:10.1186/2049-2618-1-16.

Wei EK, Giovannucci E, Fuchs CS, et al. Low Plasma Adiponectin Levels and Risk of Colorectal Cancer in Men: A Prospective Study. *J Natl Cancer Inst*. 2005;97(22):1688–1694.

Wei EK, Giovannucci E, Fuchs CS, et al. Low plasma adiponectin levels and risk of colorectal cancer in men: a prospective study. *J Natl Cancer Inst*. 2005;97(22):1688–1694.

Wei EK, Ma J, Pollak MN, et al. C-peptide, insulin-like growth factor binding protein-1, glycosylated hemoglobin, and the risk of distal colorectal adenoma in women. *Cancer Epidemiol Biomarkers Prev*. 2006;15(4):750–755.

Willett, Walter. *Nutritional Epidemiology*. Oxford University Press, 2012.

Willett, Walter. *Nutritional Epidemiology*. Oxford University Press, 2012.

Wu, Kana, Frank B. Hu, Charles Fuchs, Eric B. Rimm, Walter C. Willett, and Edward Giovannucci. “Dietary Patterns and Risk of Colon Cancer and Adenoma in a Cohort of Men (United States).” *Cancer Causes & Control* 15, no. 9 (2004): 853–62.

Wu, Song, Scott Powers, Wei Zhu, and Yusuf A. Hannun. “Substantial Contribution of Extrinsic Risk Factors to Cancer Development.” *Nature* 529, no. 7584 (January 7, 2016): 43–47. doi:10.1038/nature16166.

Yamauchi M, Lochhead P, Imamura Y, Kuchiba A, et al. Physical Activity, Tumor PTGS2 Expression, and Survival in Patients with Colorectal Cancer. Cancer Epidemiol. Biomarkers Prev 2013. 22:1142-1152. Author names in bold designate shared co-first authors.

Yasuda, Koji, Keunyoung Oh, Boyu Ren, Timothy L. Tickle, Eric A. Franzosa, Lynn M. Wachtman, Andrew D. Miller, et al. “Biogeography of the Intestinal Mucosal and Lumenal Microbiome in the Rhesus Macaque.” *Cell Host & Microbe* 17, no. 3 (March 11, 2015): 385–91. doi:10.1016/j.chom.2015.01.015.

Zeevi, David, Tal Korem, Niv Zmora, David Israeli, Daphna Rothschild, Adina Weinberger, Orly Ben-Yacov, et al. “Personalized Nutrition by Prediction of Glycemic Responses.” *Cell* 163, no. 5 (November 19, 2015): 1079–94. doi:10.1016/j.cell.2015.11.001.

Zimmers TA, Gutierrez JC, Koniaris LG. Loss of GDF-15 abolishes sulindac chemoprevention in the ApcMin/+ mouse model of intestinal cancer. *J Cancer Res Clin Oncol*. 2010;136(4):571–576.

Zmora, Niv, David Zeevi, Tal Korem, Eran Segal, and Eran Elinav. “Taking It Personally: Personalized Utilization of the Human Microbiome in Health and Disease.” *Cell Host & Microbe* 19, no. 1 (January 13, 2016): 12–20. doi:10.1016/j.chom.2015.12.016.

References

Zoetendal, Erwin G., Antoon D. L. Akkermans, and Willem M. De Vos. “Temperature Gradient Gel Electrophoresis Analysis of 16S rRNA from Human Fecal Samples Reveals Stable and Host-Specific Communities of Active Bacteria.” *Applied and Environmental Microbiology* 64, no. 10 (October 1998): 3854–59.

Tables and figures: MIC-1 and Risk of Colorectal Cancer

Table 1.1 Baseline Characteristics of Cases and Controls at the Time of Blood Draw

	NHS (women)		HPFS (men)	
Characteristic	Cases (n=344)	Controls (n=419)	Cases (n=274)	Controls (n=531)
Age at blood draw (y, mean [SD])	58.9 [6.72]	58.9 [6.77]	65.8 [8.30]	65.8 [8.27]
Race				
Caucasian (%)	97.4	99.1	93.8	94.2
Other (%)	2.60	0.90	6.2	5.8
Smoking Status				
Current or Past (%)	57.3	56.1	55.5	51.3
Never (%)	42.7	43.9	45.5	48.7
Post-menopause (%)	88.5	88.4	-	-
Post-menopausal hormone use				
Never (%)	48.8	51.3	-	-
Past (%)	20.1	14.3		
Current (%)	31.1	34.4		
Regular use of aspirin or NSAIDs (≥ 2 tablets/wk, %)	45.9	53	48.9	54.1
Current use of multivitamins (%)	33.1	35	47.5	51.7
Colorectal cancer in a parent or sibling (%)	13.1	13	19.7	13.7
History of screening prior to dx (%)	11.9	12.7	56.2	67.3
History of colon polyp (%)	3.49	3.58	13.9	15.2
Body mass index (kg/m^2, mean[SD])	25.9 [4.94]	25.5 [4.92]	26.2 [3.05]	25.3 [2.71]
Physical activity (METs, mean [SD])	15.4 [16.1]	16.8 [20.3]	32.1 [29.8]	34.6 [34.9]
Alcohol intake (g/d, mean [SD])	5.63 [9.52]	5.61 [8.90]	12.1 [14.7]	11.9 [14.6]
Calcium (mg/d, mean [SD])	976 [406]	1025 [407]	917 [385]	928 [343]
Folate (mg/d, mean [SD])	398 [180]	422 [187]	494 [208]	520 [228]
Beef as main dish (servings/d, mean [SD])	0.30 [0.17]	0.29 [0.16]	0.27 [0.21]	0.26 [0.18]
MIC-1 (pg/mL, median [IQR])	754 [603-930]	748 [567-959]	900 [671-1132]	825 [610-1061]

HPFS, Health Professionals Follow-up Study; IQR, Interquartile Range; METs, metabolic equivalent task score hours per week; MIC-1, macrophage inhibitory cytokine-1; NHS, Nurses' Health Study; NSAID, non-steroidal anti-inflammatory drugs; SD, Standard Deviation.

Tables and figures: MIC-1 and Risk of Colorectal Cancer

Table 1.2. Relative risk of colorectal cancer according to plasma levels of MIC-1

	Quintile of plasma MIC-1*					
Analysis	1	2	3	4	5	P_{trend}†
Combined (n=1568)						
No. of cases/controls	93/191	116/190	133/191	138/189	138/189	
Age-adjusted RR	1	1.33	1.57	1.74	1.80	
(95% CI) ‡	(referent)	(0.94 - 1.88)	(1.09 -2.24)	(1.20 - 2.53)	(1.22- 2.64)	0.005
MV-adjusted	1	1.37	1.59	1.78	1.80	
RR(95% CI) §	(referent)	(0.96 – 1.94)	(1.10 - 2.28)	(1.22 - 2.60)	(1.22 – 2.67)	0.007
MV-adjusted	1	1.36	1.68	1.88	1.93	
RR(95% CI) ||	(referent)	(0.95 – 1.95)	(1.16 - 2.43)	(1.27 - 2.77)	(1.27 – 2.94)	0.004
Men (n=805)						
Median (pg/mL)	501	649	825	1000	1286	
Cutpoints (pg/mL)	0-575	575-747	747-890	890-1109	1109-3266	
No. of cases/controls	42/107	51/106	44/107	64/106	73/106	
Age-adjusted	1	1.36	1.29	2.00	2.40	
RR(95% CI) ‡	(referent)	(0.82 - 2.25)	(0.75 - 2.24)	(1.16 - 3.47)	(1.36 - 4.23)	0.001
MV-adjusted	1	1.39	1.19	2.01	2.28	
RR(95% CI) §	(referent)	(0.83 - 2.35)	(0.67 - 2.09)	(1.14 - 3.54)	(1.27 - 4.10)	0.003
MV-adjusted	1	1.44	1.26	2.21	2.76	
RR(95% CI) ||	(referent)	(0.85 - 2.43)	(0.71 - 2.23)	(1.24 - 3.95)	(1.48 – 5.17)	0.0006
Women (n=763)						
Median (pg/mL)	483	599	749	926	1218	
Cutpoints (pg/mL)	0-544	544-664	664-818	818-1014	1014-3256	
No. of cases/controls	51/84	65/84	89/84	74/84	65/83	
Age-adjusted	1	1.34	1.84	1.57	1.39	
RR(95% CI) ‡	(referent)	(0.82 - 2.17)	(1.13 – 2.99)	(0.94 - 2.62)	(0.82 - 2.35)	0.46
MV-adjusted	1	1.34	1.88	1.56	1.41	
RR(95% CI) §	(referent)	(0.82 - 2.18)	(1.15 - 3.07)	(0.93 - 2.63)	(0.82 - 2.41)	0.46
MV-adjusted	1	1.25	2.02	1.58	1.29	
RR(95% CI) ||	(referent)	(0.75 - 2.08)	(1.21 - 3.38)	(0.91 - 2.73)	(0.71 - 2.32)	0.66

* Quintiles of plasma MIC-1 were calculated separately within each cohort based on the distribution in the controls. For the combined analysis, we assigned participants to quintile categories based upon cohort-specific cutpoints. Two-sided $p_{interaction}$ for the association between MIC-1 and colorectal cancer among men compared with women was 0.13.

† Tests for linear trend were conducted using the median values for each quintile of analyte in a logistic regression model. P-values are two-sided

‡ Age-adjusted models include adjustment for matching factors (age at blood draw, date of blood draw, race, and gender).

§ Multivariate models were adjusted for age at blood draw, date of blood draw, race, gender, body mass index (BMI), physical activity (metabolic equivalent task score hours per week (METS)), current or past smoking (yes or no), prior/current use of postmenopausal hormones (yes or no), prior history of screening (yes or no), previous occurrence of adenoma, colorectal cancer in parent or sibling, regular use of multivitamins, regular use of aspirin or NSAIDs (≥ 2 tablets per wk), energy-adjusted intake (including supplements) of calcium and folate, servings of red meat as a main dish, and alcohol consumption.

||Multivariate model additionally adjusted for the plasma inflammatory markers CRP, TNFRSF1B, IL6, and total adiponectin as continuous measures as well as the factors listed in c.

Tables and figures: MIC-1 and Risk of Colorectal Cancer

Table 1.3. Relative risk of colorectal cancer according to plasma MIC-1 within selected subgroups

Subgroup	Analysis	Quintiles of plasma MIC-1					P_{trend}*	P_{int}†
		1	2	3	4	5		
Age <65 (n = 329)	No. of Cases/Controls	36/88	33/69	16/30	16/22	11/8		
	MV RR (95% CI)	1 (referent)	1.15 (0.60 – 2.17)	1.11 (0.48 – 2.57)	2.16 (0.88 – 5.29)	4.20 (1.31 – 13.5)	0.01	0.32
Age ≥ 65 (n = 1239)	No. of Cases/Controls	57/103	83/121	117/161	122/167	127/181		
	MV RR (95% CI)	1 (referent)	1.34 (0.85 – 2.11)	1.67 (1.08 – 2.61)	1.79 (1.13 - 2.84)	1.73 (1.07 – 2.81)	0.07	
No history of smoking (n = 676)	No. of Cases/Controls	48/107	48/90	67/91	53/83	53/72		
	MV RR (95% CI)	1 (referent)	1.19 (0.70 – 2.03)	2.03 (1.17 – 3.49)	2.08 (1.13-3.81)	2.39 (1.25– 4.57)	0.007	0.18
Past or current smoker (n = 852)	No. of Cases/Controls	45/84	68/100	66/100	85/106	85/117		
	MV RR (95% CI)	1 (referent)	1.45 (0.88 – 2.39)	1.38 (0.82 – 2.34)	1.81 (1.07 – 3.09)	1.61 (0.90 – 2.86)	0.17	
Non-user of aspirin or NSAIDs (n = 767)	No. of Cases/Controls	60/114	59/92	73/82	68/79	66/74		
	MV RR (95% CI)	1 (referent)	1.31 (0.80 – 2.12)	1.94 (1.17 – 3.22)	2.04 (1.19 – 3.51)	2.45 (1.34 – 4.45)	0.003	0.58
Regular user of aspirin or NSAID (n = 801)	No. of Cases/Controls	33/77	57/98	60/109	70/110	72/115		
	MV RR (95% CI)	1 (referent)	1.48 (0.85 – 2.57)	1.54 (0.87– 2.73)	1.75 (0.97 – 3.14)	1.78 (0.96 – 3.31)	0.15	
Low BMI‡ (n = 695)	No. of Cases/Controls	35/88	46/89	54/86	61/100	55/81		
	MV RR (95% CI)	1 (referent)	1.25 (0.71 – 2.23)	1.62 (0.89 – 2.97)	1.96 (1.05 – 3.64)	1.99 (1.01 -3.94)	0.04	0.23
High BMI§ (n = 873)	No. of Cases/Controls	58/103	70/101	79/105	77/89	83/108		
	MV RR (95% CI)	1 (referent)	1.39 (0.87 – 2.22)	1.62 (1.00 – 2.62)	1.91 (1.13 – 3.20)	1.90 (1.09 – 3.29)	0.04	
No family history of colorectal cancer‖ (n=224)	No. of Cases/Controls	79/169	103/168	112/163	114/171	111/154		
	MV RR (95% CI)	1 (referent)	1.44 (0.98 – 2.12)	1.77 (1.18 – 2.65)	1.79 (1.17 – 2.73)	2.23 (1.41 – 3.54)	0.002	0.78
Family history of colorectal cancer‖ (n=1344)	No. of Cases/Controls	14/22	13/22	21/28	24/18	27/35		
	MV RR (95% CI)	1 (referent)	0.96 (0.34 – 2.72)	1.41 (0.50 – 4.00)	2.44 (0.80 – 7.55)	1.01 (0.33 – 3.04)	0.97	

Tables and figures: MIC-1 and Risk of Colorectal Cancer

No history of colon polyps (n = 1422)	No. of Cases/Controls	90/178	108/175	124/165	121/167	125/169		
	MV RR (95% CI)	1 (referent)	1.28 (0.89 – 1.86)	1.66 (1.14 – 2.47)	1.67 (1.11 – 2.51)	1.81 (1.17 -2.81)	0.02	0.17
History of colon polyps (n = 146)	No. of Cases/Controls	3/13	8/15	9/26	17/22	13/20		
	MV RR (95% CI)	1 (referent)	3.18 (0.52 – 2.22)	1.68 (0.29 – 9.55)	6.06 (1.04 – 35.3)	3.78 (0.55 – 26.1)	0.19	
No history of prior screening (n = 959)	No. of Cases/Controls	68/121	86/103	99/105	90/102	80/105		
	MV RR (95% CI)	1 (referent)	1.45 (0.94 – 2.23)	1.73 (1.11 – 2.71)	1.58 (0.98 – 2.56)	1.30 (0.77 -2.17)	0.72	0.09
History of prior screening (n = 609)	No. of Cases/Controls	25/70	30/87	34/86	48/87	58/84		
	MV RR (95% CI)	1 (referent)	1.20 (0.61 – 2.35)	1.69 (0.84– 3.39)	2.72 (1.34 – 5.52)	4.18 (1.95 – 8.93)	<0.0001	
< 4 years after blood draw (n=360)	No. of cases/controls	15/41	22/31	32/44	32/56	42/47	0.03	
	MV RR (95% CI)	1 (referent)	2.32 (0.93-5.76)	2.64 (1.08-6.47)	2.48 (0.98-6.30)	4.10 (1.48-11.3)		0.98
≥ 4 years after blood draw (n=1208)	No. of cases/controls	78/150	94/159	101/147	106/133	96/142	0.02	
	MV RR (95% CI)	1 (referent)	1.28 (0.86-1.90)	1.57 (1.04-2.38)	1.90 (1.23-2.95)	1.70 (1.06-2.73)		

* Tests for linear trend were conducted using the median values for each quintile of analyte in a logistic regression model. P-values are two-sided

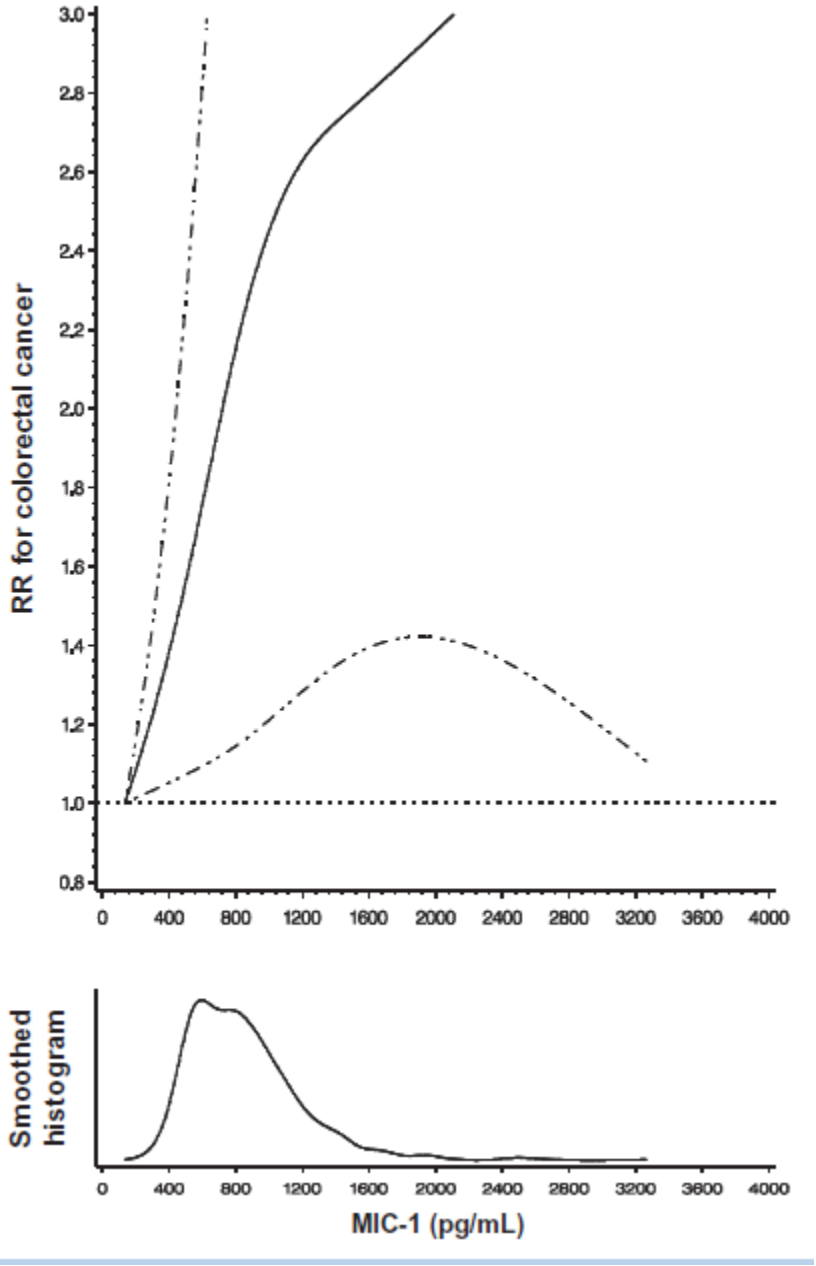
† Tests for interaction were conducted using cross-product terms between stratification factors and the continuous value of MIC-1 in logistic regression models. P-values are two-sided.

‡ Low BMI defined as below the median (<24.0 kg/m^2 for women and < 25.0 kg/m^{2}for men)

§ High BMI defined as above the median (≥ 24.0 kg/m^{2}for women and ≥ 25.0 kg/m^2 for men)

||Family history of colorectal cancer includes at least one sibling or parent with colorectal cancer

Figure 1.1. Restricted cubic spline plot for MIC-1 and risk of colorectal cancer. RR of colorectal cancer is plotted accordingly to serum MIC-1 level (pg/mL). *Hatched lines* represent 95% CIs. Spline was adjusted for the same factors as the multivariate model "e" in Table 3. A test for overall significance of the curve was statistically significant (p = 0.02). The test for curvature (i.e. non-linear relation) was 0.12. Test for linear relation was 0.01. All p-values listed here are two-sided. A smoothed histogram below shows the distribution of MIC-1 among cases and controls in both cohorts.



Supplementary Table 1.1. Spearman correlation coefficients between MIC-1, selected lifestyle factors and plasma inflammatory markers among NHS control participants*

Variable	MIC-1	Age	BMI	METs	TNFRSF1B	IL-6	Adiponectin	CRP	WHR
MIC-1	1.00	0.44[a]	-0.01	-0.09	0.43[a]	0.25[a]	0.14[b]	0.19[d]	0.16
Age		1.00	0.004	0.02	0.27[a]	0.09	0.19[d]	0.12[c]	0.24[a]
BMI			1.00	-0.16[d]	0.25[a]	0.28[a]	-0.32[a]	0.39[a]	0.37[a]
METs				1.00	-0.08	-0.11[c]	0.04	-0.16[b]	-0.14[c]
TNFRSF1B					1.00	0.22[a]	-0.11[c]	0.28[a]	0.21[d]
IL-6						1.00	-0.19[d]	0.50[a]	0.27[a]
Adiponectin							1.00	-0.25[a]	-0.30[a]
CRP								1.00	0.28[a]
WHR									1.00

BMI, body mass index; CRP, C-reactive protein; IL-6, interleukin-6; METs, metabolic equivalent task score hours; MIC-1, macrophage inhibitory cytokine-1; TNFRSF1B, tumor necrosis factor receptor superfamily member 1B ; WHR, waist-to-hip ratio

* Two-sided p-values calculated for Spearman correlation coefficients

†P< 0.0001

‡P< 0.01

§P < 0.05

||P < 0.001

Supplementary Table 1.2. Spearman correlation coefficients between MIC-1, selected lifestyle factors and plasma inflammatory markers among HPFS control participants*

Variable	MIC-1	Age	BMI	METs	TNFRSF1B	IL-6	Adiponectin	CRP	WHR
MIC-1	1.00	0.62[a]	0.02	-0.08	0.49[a]	0.37[a]	0.14[b]	0.27[a]	0.15[b]
Age		1.00	-0.07	0.008	0.35[a]	0.25[a]	0.20[a]	0.14[b]	0.10[c]
BMI			1.00	-0.12[b]	0.08	0.13[b]	-0.26[a]	0.19[a]	0.39[a]
METs				1.00	-0.11[b]	-0.13[b]	0.08	-0.10[c]	-0.15[b]
TNFRSF1B					1.00	0.32[a]	-0.04	0.26[a]	0.14[b]
IL-6						1.00	-0.07	0.50[a]	0.11[c]
Adiponectin							1.00	-0.15[d]	-0.26[a]
CRP								1.00	0.16[d]
WHR									1.00

BMI, body mass index; CRP, C-reactive protein; IL-6, interleukin-6; METs, metabolic equivalent task score hours; MIC-1, macrophage inhibitory cytokine-1; TNFRSF1B, tumor necrosis factor receptor superfamily member 1B ;; WHR, waist-to-hip ratio

* Two-sided p-values calculated for Spearman correlation coefficients

†P< 0.0001

‡P< 0.01

§P < 0.05

Tables and figures: MIC-1 and Colorectal Cancer Survival

Table 2.1. Baseline characteristics of study participants according to MIC1 levels.[a]

	Quartiles of MIC1[b]			
Characteristic	1 (n=119)	2 (n=160)	3 (n=174)	4 (n=165)
Age at diagnosis, y, mean (SD)	63.7 (8.0)	68.7 (7.2)	71.4 (7.6)	74.4 (7.1)
Sex				
Female, %	55	63	57	48
Body mass index, kg/m^2, mean (SD)	26.1 (4.1)	26.2 (4.5)	26.0 (4.4)	25.8 (3.8)
Race				
White, %	95	96	97	95
Other, %	5	4	3	5
Smoking status				
Current or past, %	51	43	45	38
Never, %	49	58	55	62
Physical activity, METs, mean (SD)	20.3 (20.1)	20.9 (22.2)	24.5 (25.7)	24.4 (22.7)
Postmenopausal[c], %	69	88	95	96
Menopausal hormone use[c]				
Current or past, %	42	54	71	53
Never, %	58	46	29	48
Regular use of aspirin or NSAIDs[d], %	40	45	50	52
Site of CRC, %				
Colon	83	86	90	93
Rectum	17	14	10	7
Stage of Disease, %				
I	37	26	21	29
II	23	24	27	18
III	18	23	21	18
IV	11	15	15	16
Other/unknown	12	16	16	19
Grade of differentiation, %				
Well	13	11	11	7
Moderate	61	53	56	55
Poor/undifferentiated	13	18	13	14
Unknown	13	19	20	24

[a] Data on lifestyle characteristics from the time of the blood draw. Data on tumor characteristics from the time of cancer diagnosis.

b. Quartiles of MIC1 were defined by the distribution among control individuals without CRC in our previous study.[9]

[c] The percentages of postmenopausal participants as well as hormone use is among women only.

[d] Regular use of aspirin or NSAIDs defined as ≥2 tablets/week

Abbreviations: CRC = colorectal cancer; METs = metabolic equivalent task score hours per week; MIC1 = macrophage inhibitory cytokine-1; NSAIDs = nonsteroidal anti-inflammatory drugs; SD = standard deviation.

Table 2.2. Colorectal cancer-specific mortality and overall mortality according to MIC1 levels.

	Quartile of plasma MIC1[a]				
Analysis	1	2	3	4	P_{trend}[b]
All-cause mortality (n=618)					
No. of events/No. at risk	25/119	62/160	83/174	96/165	
Age-adjusted HR (95% CI)	1 (referent)	1.88 (1.17-3.03)	2.51 (1.57-3.99)	2.85 (1.78-4.58)	<0.0001
MV-adjusted HR (95% CI)[c]	1 (referent)	1.77 (1.07-2.91)	2.55 (1.55-4.19)	2.63 (1.60-4.32)	0.0002
MV-adjusted HR (95% CI)[d]	1 (referent)	1.74 (1.06-2.88)	2.54 (1.54-4.19)	2.63 (1.60-4.32)	0.0002
CRC-specific mortality (n=618)					
No. of events/No. at risk	19/119	43/160	59/174	57/165	
Age-adjusted HR (95% CI)	1 (referent)	1.94 (1.12-3.37)	2.64 (1.54-4.51)	2.73 (1.57-4.75)	0.003
MV-adjusted HR (95% CI)[c]	1 (referent)	1.65 (0.92-2.94)	2.60 (1.45-4.67)	2.34 (1.30-4.21)	0.01
MV-adjusted HR (95% CI)[d]	1 (referent)	1.67 (0.93-3.00)	2.67 (1.49-4.80)	2.40 (1.33-4.34)	0.009

[a] Quartiles of plasma MIC1 were defined by the distribution among control individuals without CRC in our previous study.[9] We assigned participants to quartile categories based upon cohort-specific cutpoints (NHS: Q1 <565; Q2 565-747, Q3 748-953, Q4 >953 pg/mL) (HPFS: Q1 <610; Q2 610-825; Q3 826-1060; Q4 >1060 pg/mL).

[b] Tests for linear trend were conducted using the median values for each quartile of MIC1.

[c] Additionally adjusted for the following potential confounders: cohort (sex); date of blood draw; race; body mass index (BMI) at the time of blood draw; physical activity (in MET-hours/week) at the time of blood draw; family history of CRC at time of blood draw; regular use of aspirin or NSAIDs at the time of blood draw; stage at diagnosis; history of polyps at the time of blood draw; grade (poor/unknown vs. well vs. moderate) at diagnosis; and anatomic tumor location (colon vs. rectum) at diagnosis

[d] Multivariate models additionally adjusted for the plasma inflammatory markers CRP and IL6 as continuous measures as well as the factors listed in c

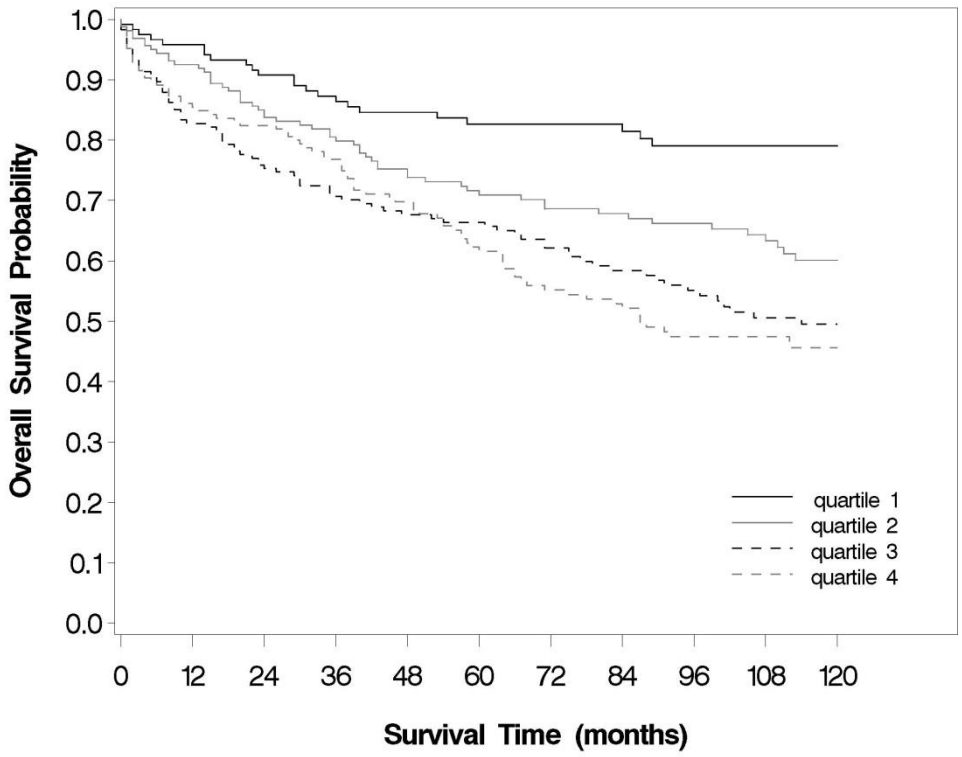
Table 2.3. Colorectal cancer-specific mortality according to plasma MIC1 within selected subgroups.

		Quartiles of plasma MIC1					
		1	2	3	4	P_{trend}	P_{int}
Younger Age[a] (n = 308)	No. of events/No. at risk	17/94	29/92	32/79	14/43		
	MV HR (95% CI)	1 (referent)	1.71 (0.93 – 3.16)	3.08 (1.66 – 5.69)	1.67 (0.81 – 3.45)	0.11	0.92
Older Age[b] (n = 310)	No. of events/No. at risk	2/25	14/68	27/95	43/122		
	MV HR (95% CI)	1 (referent)	1.35 (0.30 – 6.06)	2.38 (0.56 – 10.1)	2.94 (0.70 – 12.3)	0.003	
Male (n=274)	No. of events/No. at risk	9/54	11/60	23/75	28/85		
	MV HR (95% CI)	1 (referent)	1.56 (0.62 – 3.94)	2.08 (0.89 – 4.86)	2.16 (0.94-4.95)	0.11	0.96
Female (n=344)	No. of events/No. at risk	10/65	32/100	36/99	29/80		
	MV HR (95% CI)	1 (referent)	1.76 (0.83 – 3.69)	3.06 (1.45 – 6.49)	2.46 (1.14 – 5.28)	0.03	
Stage I/II/III tumors (n=533)	No. of events/No. at risk	9/106	27/141	34/148	31/138		
	MV HR (95% CI)	1 (referent)	2.87 (1.31-6.32)	3.54 (1.59-7.87)	3.23 (1.43-7.28)	0.04	0.42
Stage IV tumors (n=85)	No. of events/No. at risk	10/13	16/19	25/26	26/27		
	MV HR (95% CI)	1 (referent)	0.88 (0.38-2.06)	2.05 (0.91-4.60)	1.77 (0.80-3.89)	0.08	
Lower BMI[c] (n = 251)	No. of events/No. at risk	8/46	24/65	26/73	27/67		
	MV HR (95% CI)	1 (referent)	2.40 (1.03 – 5.57)	2.28 (0.98 – 5.30)	2.38 (1.02 – 5.55)	0.18	0.78
Higher BMI[d] (n = 367)	No. of events/No. at risk	11/73	19/95	33/101	30/98		
	MV HR (95% CI)	1 (referent)	1.20 (0.55 – 2.60)	2.88 (1.37 – 6.05)	2.27 (1.07 – 4.84)	0.03	
Well or moderate grade (n = 409)	No. of events/No. at risk	10/89	17/101	34/117	31/102		
	MV HR (95% CI)	1 (referent)	2.17 (0.94 – 5.01)	3.27 (1.49 – 7.20)	2.90 (1.30 – 6.46)	0.02	0.46
Poor or undifferentiated Grade (n = 88)	No. of events/No. at risk	4/15	8/28	7/22	7/23		
	MV HR (95% CI)	1 (referent)	1.15 (0.43 – 3.05)	1.69 (0.57– 4.96)	1.55 (0.53 – 4.56)	0.36	
Current/Past Smoker	No. of events/No. at risk	13/58	23/92	30/96	36/103		

(n = 349)	MV RR (95% CI)	1 (referent)	1.39 (0.67 – 2.87)	2.18 (1.08 – 4.38)	2.35 (1.17 – 4.72)	0.02	0.84
	No. of events/No. at risk	6/61	20/68	29/78	21/62		
Never Smoker (n = 269)	MV RR (95% CI)	1 (referent)	2.15 (0.83-5.56)	3.49 (1.35-9.00)	2.45 (0.92- 6.53)	0.16	

[a] Younger age defined as below the median at diagnosis (<68 for women and < 74 for men)
[b] Older age defined as above the median at diagnosis(≥ 68 for women and ≥ 74 for men)
[c] Lower BMI defined as below the median at blood draw(<24.0 kg/m^2 for women and < 25.0 kg/$m^2$for men)
[d] Higher BMI defined as above the median at blood draw (≥ 24.0 kg/$m^2$for women and ≥ 25.0 kg/m^2 for men)

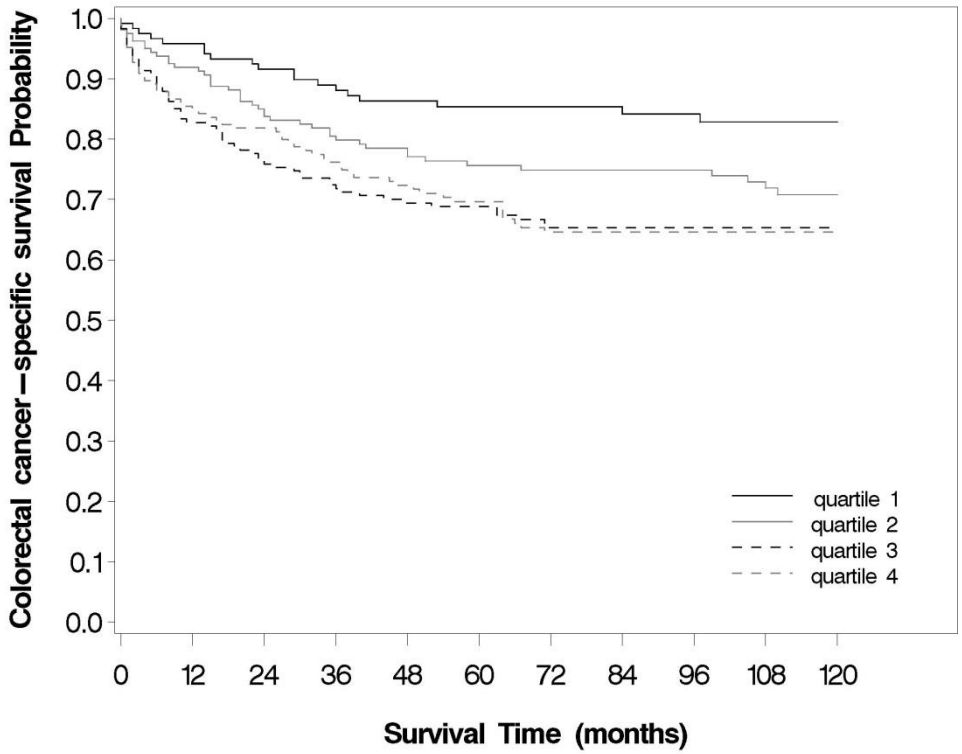
Figure 2.1. All-cause survival according to levels of MIC1 (in quartiles). Log rank test: p<0.0001.



Number at Risk

quartile 1	119	114	107	99	91	81	80	70	62	50	35
quartile 2	160	148	136	121	108	97	90	84	74	64	50
quartile 3	174	144	132	123	109	96	88	75	64	52	43
quartile 4	165	142	135	121	105	89	75	69	58	55	44

Figure 2.2. CRC-specific survival according to levels of MIC1 (in quartiles). Log rank test: p = 0.003.



Number at Risk

quartile 1	119	114	108	101	93	84	81	71	65	51	36
quartile 2	160	147	136	120	111	101	97	92	80	71	58
quartile 3	174	144	133	125	112	100	93	84	77	67	54
quartile 4	165	141	134	120	110	99	87	81	75	69	60

Tables and Figures: Diet and Risk of CRC According to *F. nucleatum* enrichment

Table 3.1. Factor loading matrix for dietary patterns from FFQs from 1990 for NHS and 1994 for HPFS.[1]

	HPFS		NHS	
	Prudent pattern	Western pattern	Prudent pattern	Western pattern
Other vegetables[2]	0.69		0.68	
Yellow Vegetables	0.64		0.65	
Leafy green vegetables	0.63		0.64	
Cruciferous Vegetables	0.63		0.61	
Fruit	0.62		0.58	
Legumes	0.62		0.57	
Tomatoes	0.52		0.48	
Fish	0.46		0.51	
Whole grain	0.38		0.41	
Potatoes	0.33			0.34
Poultry	0.32		0.41	
Water			0.32	
Salad dressing			0.33	
Low-fat dairy			0.32	
Red Meat		0.66		0.61
Processed meats		0.61		0.58
High fat dairy		0.51		0.50
French fries		0.49		0.46
Eggs		0.47		0.41
Desserts[3]		0.43		0.45
Condiments[4]		0.39		0.36
Refined grains		0.38		0.38
Butter		0.38		0.50
Mayo		0.36		0.34
Margarine		0.34		0.32
Snacks[5]		0.34		
Pizza		0.33		0.36
Creamy soups		0.31		0.32
Sugared-sweetened beverages		0.31		0.33

[1] With the orthogonal rotation used, correlations are identical to factor loading matrix. Only items with correlation coefficients >0.30 are presented.

[2] Other vegetables include corn, onion, eggplant, celery, green peppers and mixed vegetables.

[3] Desserts include chocolate, candy bars, cookies, brownies, cake, pie, and pastries.

[4] Condiments include soy sauce, non-dairy creamer, Worcestershire sauce, red chili sauce, and pepper.

[5] Snacks include chips, popcorn and crackers.

Table 3.2. Baseline age-standardized characteristics of HPFS and NHS participants according to Western and Prudent dietary pattern intake[1]

Characteristic	Western Dietary Pattern				Prudent Dietary Pattern			
	HPFS		NHS		HPFS		NHS	
	Q1	Q4	Q1	Q4	Q1	Q4	Q1	Q4
Participants, %	27.8	26.0	24.0	28.7	28.7	25.5	22.2	24.0
Age, years	54.01	53.75	46.88	45.87	53.83	53.96	46.19	46.54
Pack-years smoked	10.04	17.67	10.46	12.90	15.99	11.33	14.21	10.17
Current smoker, %	5.2	16.0	22.0	35.1	15.1	6.1	37.1	23.2
Family history of colorectal cancer, %	13.9	13.3	16.4	16.7	12.9	14.2	16.3	16.5
History of previous endoscopy, %	27.3	23.1	11.8	8.5	22.4	27.5	9.3	10.5
Current multivitamin use, %	46.2	37.1	41.9	27.7	37.7	46.7	28.7	38.5
Regular aspirin use, %[2]	30.8	34.5	34.9	40.5	31.7	32.8	40.1	36.5
Postmenopausal, %[3]			47.24	43.30			43.93	46.44
Menopausal hormone use, %[3]			20.26	16.19			17.43	18.31
Body mass index, kg/m^2	24.47	25.30	24.21	24.73	25.04	24.76	24.21	24.76
Physical activity, MET-hours/wk[4]	24.52	18.04	15.48	8.84	15.82	27.19	8.88	14.36
Dietary intake								
Unprocessed red meat, svg/d	0.27	0.99	0.55	1.00	0.64	0.55	0.77	0.80
Processed red meat, svg/d	0.11	0.70	0.16	0.61	0.42	0.29	0.42	0.33
Poultry, svg/d	0.37	0.34	0.33	0.23	0.24	0.48	0.15	0.41
Fruit, svg/d	2.73	2.17	2.75	1.72	1.38	3.63	1.42	2.85
Vegetable, svg/d	3.10	3.20	2.42	1.79	1.56	5.14	1.05	3.30
Alcohol, g/d	7.64	14.79	7.00	5.93	11.56	10.92	6.86	5.93
Folate, µg/d	561.55	408.96	471.48	290.01	402.71	573.83	285.51	451.75
Calcium, mg/d	1007.14	806.82	926.73	597.07	845.71	952.79	688.63	794.55
Vitamin D, IU/d	499.54	331.63	433.39	252.29	359.95	459.26	295.48	365.49
Dietary fiber, g/d	25.05	18.07	18.32	11.02	16.20	26.89	10.03	19.18

[1]Continuous variables are described as means. All variables were assessed in 1986 for the Health Professionals Follow-up Study and in 1980 for the Nurses' Health Study unless otherwise specified

[2]Regular aspirin users were defined as those who used ≥2 standard (325-mg) tablets of aspirin per week.

[3] The percentages of postmenopausal participants as well as hormone use are among women only.

[4]Physical activity is represented by the product sum of the metabolic equivalent (MET) of each specific recreational activity and hours spent on that activity per week. METs were assessed in 1986 for the NHS.

Table 3.3. Hazard ratios (HRs) and 95% confidence intervals (CIs) of overall colorectal cancer according to cumulative average Western and Prudent dietary patterns in the pooled Health Professionals' Follow-up Study and Nurses' Health Study.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Ptrend
Western Pooled:					
MV-adjusted[1]					
No. of cases (n=3042)	783	764	732	763	
Person-years	886,764	888,732	888,016	888,880	
HRs and 95% CIs	1.00 (referent)	1.08 (0.97-1.19)	1.11 (0.99-1.24)	1.30 (1.14-1.48)	0.0001
Prudent Pooled:					
MV-adjusted[1]					
No. of cases (n=3042)	780	708	773	781	
Person years	890,430	884,735	889,416	887,812	
HRs and 95% CIs	1.00 (referent)	0.89 (0.80-0.99)	0.92 (0.82-1.02)	0.87 (0.78-0.97)	0.03

[1] Cox proportional hazards models were used and stratified by age, calendar year, and gender and adjusted for family history of colorectal cancer in any first-degree relative, prior lower gastrointestinal endoscopy, menopausal hormone therapy status (never, past current), pack-years of smoking (never, 0-<5 years, 5-<20 years, 20-<40 years, or ≥40 years), body mass index (kg/m^2), physical activity, regular aspirin or NSAID use (≥2 tablets/wk), alcohol consumption, and total caloric intake (kcal/day).

Table 3.4. Association between Western and Prudent dietary patterns and CRC according to *F. nucleatum* status in the NHS and HPFS.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Ptrend	P het
Prudent dietary pattern						
F. nucleatum (+) CRC[1]						
No. of cases (n=123)	43	25	34	21		
Age-adjusted HR (95% CI)[2]	1.00 (referent)	0.52 (0.31-0.85)	0.66 (0.42-1.04)	0.38 (0.22-0.64)	0.0009	
Multivariable HR (95% CI)[3]	1.00 (referent)	0.54 (0.33-0.88)	0.70 (0.44-1.10)	0.41 (0.24-0.70)	0.003	
F. nucleatum (-) CRC						0.005
No. of cases (n=867)	201	212	226	228		
Age-adjusted HR (95% CI)[2]	1.00 (referent)	0.99 (0.82-1.21)	0.95 (0.78-1.16)	0.90 (0.74-1.11)	0.26	
Multivariable HR (95% CI)[3]	1.00 (referent)	1.03 (0.85-1.25)	1.01 (0.83-1.23)	0.98 (0.79-1.21)	0.73	
Western dietary pattern						
F. nucleatum (+) CRC[1]						
No. of cases (n=123)	24	33	33	33		
Age-adjusted HR (95% CI)[2]	1.00 (referent)	1.47 (0.87-2.50)	1.63 (0.95-2.78)	1.93 (1.12-3.35)	0.02	
Multivariable HR (95% CI)[3]	1.00 (referent)	1.41 (0.83-2.39)	1.51 (0.88-2.58)	1.67 (0.96-2.90)	0.07	
F. nucleatum (-) CRC						0.21
No. of cases (n=867)	212	238	206	211		
Age-adjusted HR (95% CI)[2]	1.00 (referent)	1.25 (1.04-1.51)	1.17 (0.95-1.43)	1.39 (1.10-1.75)	0.01	
Multivariable HR (95% CI)[3]	1.00 (referent)	1.20 (0.99-1.45)	1.08 (0.87-1.32)	1.20 (0.95-1.53)	0.23	

[1] Determined according to PCR detection of 16S rRNA sequence for F. nucleatum. Fuso positive was defined as any *Fusobacterium nucleatum* DNA and Fuso negative was defined no detection of Fusobacterium nucleatum DNA.
[2] Stratified by age, calendar year, and gender and adjusted for total caloric intake
[3] As above, and additionally adjusted for family history of colorectal cancer in any first-degree relative, prior lower gastrointestinal endoscopy, menopausal hormone therapy status (never, past current), pack-years of smoking (never, 0-<5 years, 5-<20 years, 20-<40 years, or ≥40 years), body mass index (kg/m^2), physical activity, regular aspirin or NSAID use (≥2 tablets/wk), alcohol consumption, and total caloric intake (kcal/day).

Table 3.5. Association between the top two factor loadings in the Prudent and Western dietary patterns and CRC according to *F. nucleatum* status in the NHS and HPFS.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Ptrend	P het*
Vegetables						
F. nucleatum (+) CRC[1] (n=123)	29	33	31	30		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.93 (0.56-1.54)	0.81 (0.48-1.35)	0.68 (0.40-1.16)	0.10	
F. nucleatum (-) CRC (n=867)	148	211	249	259		0.02
Multivariable HR (95% CI)[2]	1.00 (referent)	1.23 (0.99-1.53)	1.34 (1.09-1.66)	1.32 (1.06-1.64)	0.03	
Fruit						
F. nucleatum (+) CRC[1] (n=123)	31	3	36	25		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.78 (0.47-1.29)	0.84 (0.51-1.36)	0.54 (0.31-0.93)	0.03	
F. nucleatum (-) CRC (n=867)	188	221	226	232		0.15
Multivariable HR (95% CI)[2]	1.00 (referent)	1.00 (0.82-1.22)	0.91 (0.75-1.12)	0.85 (0.69-1.06)	0.07	
Red meat						
F. nucleatum (+) CRC[1] (n=123)	32	31	32	28		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.01 (0.61-1.66)	1.05 (0.63-1.72)	1.18 (0.70-1.99)	0.51	
F. nucleatum (-) CRC (n=867)	203	232	232	200		0.90
Multivariable HR (95% CI)[2]	1.00 (referent)	1.19 (0.98-1.44)	1.24 (1.02-1.51)	1.25 (1.00-1.56)	0.05	
Processed meat						
F. nucleatum (+) CRC[1] (n=123)	34	21	31	37		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.63 (0.36-1.08)	0.91 (0.56-1.49)	1.22 (0.76-1.97)	0.09	
F. nucleatum (-) CRC (n=867)	199	230	223	215		0.24
Multivariable HR (95% CI)[2]	1.00 (referent)	1.23 (1.01-1.49)	1.16 (0.95-1.41)	1.19 (0.97-1.46)	0.22	

[1] Determined according to PCR detection of 16S rRNA sequence for F. nucleatum. Fuso positive was defined as any *Fusobacterium nucleatum* DNA and Fuso negative was defined no detection of Fusobacterium nucleatum DNA.

[2] Stratified by age, calendar year, and gender and adjusted for family history of colorectal cancer in any first-degree relative, prior lower gastrointestinal endoscopy, menopausal hormone therapy status (never, past current), pack-years of smoking (never, 0-<5 years, 5-<20 years, 20-<40 years, or ≥40 years), body mass index (kg/m^2), physical activity, regular aspirin or NSAID use (≥2 tablets/wk), alcohol consumption, and total caloric intake (kcal/day).

* Phet did not change appreciably upon mutual adjustment and/or inclusion of micronutrients (calcium, folate, vitamin D)

Table 3.6. Association between three major macronutrients and CRC according to *F. nucleatum* status in the NHS and HPFS.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Ptrend	P het*
Fiber						
F. nucleatum (+) CRC[1] (n=123)	30	27	39	27		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.69 (0.40-1.16)	0.85 (0.52-1.39)	0.55 (0.32-0.94)	0.04	
F. nucleatum (-) CRC (n=867)	253	225	228	261		0.02
Multivariable HR (95% CI)[2]	1.00 (referent)	1.19 (0.97-1.47)	1.11 (0.90-1.38)	1.19 (0.96-1.48)	0.20	
Fat						
F. nucleatum (+) CRC[1] (n=123)	27	37	31	28		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.53 (0.93-2.52)	1.40 (0.83-2.36)	1.73 (1.00-2.98)	0.06	0.21
F. nucleatum (-) CRC (n=867)	232	225	224	186		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.03 (0.85-1.23)	1.11 (0.92-1.34)	1.12 (0.92-1.37)	0.18	
Protein						
F. nucleatum (+) CRC[1] (n=123)	32	36	35	20		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.05 (0.65-1.70)	1.03 (0.64-1.67)	0.66 (0.38-1.16)	0.19	0.59
F. nucleatum (-) CRC (n=867)	219	236	241	171		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.06 (0.88-1.27)	1.09 (0.91-1.31)	0.80 (0.65-0.98)	0.05	

[1] Determined according to PCR detection of 16S rRNA sequence for F. nucleatum. Fuso positive was defined as any *Fusobacterium nucleatum* DNA and Fuso negative was defined no detection of Fusobacterium nucleatum DNA.
[2] Stratified by age, calendar year, and gender and adjusted for family history of colorectal cancer in any first-degree relative, prior lower gastrointestinal endoscopy, menopausal hormone therapy status (never, past current), pack-years of smoking (never, 0-<5 years, 5-<20 years, 20-<40 years, or ≥40 years), body mass index (kg/m^2), physical activity, regular aspirin or NSAID use (≥2 tablets/wk), alcohol consumption, and total caloric intake (kcal/day).
* Phet did not change appreciably upon mutual adjustment and/or inclusion of micronutrients (calcium, folate, vitamin D)

Tables and figures: Diet and Risk of Colorectal Cancer according to Intratumoral *F. nucleatum* enrichment

Table 3.S1. Association between subtypes of dietary fiber and colorectal cancer enriched with *Fusobacterium nucleatum*.

	Quartile 1	Quartile 2	Quartile 3	Quartile 4	Ptrend	P het*
Cereal-derived Fiber						
F. nucleatum (+) CRC[1]						
No. of cases (n=123)	28	32	32	31		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.80 (0.48-1.35)	0.68 (0.40-1.14)	0.61 (0.35-1.04)	0.04	
F. nucleatum (-) CRC						0.01
No. of cases (n=867)	153	219	238	257		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.16 (0.94-1.44)	1.15 (0.93-1.42)	1.19 (0.96-1.48)	0.11	
Vegetable-derived Fiber						
F. nucleatum (+) CRC[1]						
No. of cases (n=123)	33	29	27	34		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.72 (0.43-1.19)	0.62 (0.37-1.04)	0.73 (0.45-1.19)	0.20	
F. nucleatum (-) CRC						0.06
No. of cases (n=867)	157	226	240	244		
Multivariable HR (95% CI)[2]	1.00 (referent)	1.28 (1.04-1.58)	1.29 (1.05-1.58)	1.28 (1.04-1.57)	0.07	
Fruit-derived Fiber						
F. nucleatum (+) CRC[1]						
No. of cases (n=123)	29	32	35	27		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.89 (0.53-1.47)	0.89 (0.54-1.47)	0.687 (0.40-1.18)	0.15	
F. nucleatum (-) CRC						0.38
No. of cases (n=867)	187	218	237	225		
Multivariable HR (95% CI)[2]	1.00 (referent)	0.99 (0.81-1.21)	0.99 (0.81-1.21)	0.879 (0.71-1.08)	0.16	

[1] Determined according to PCR detection of 16S rRNA sequence for F. nucleatum. Fuso positive was defined as any *Fusobacterium nucleatum* DNA and Fuso negative was defined no detection of Fusobacterium nucleatum DNA.

[2] Stratified by age, calendar year, and gender and adjusted for family history of colorectal cancer in any first-degree relative, prior lower gastrointestinal endoscopy, menopausal hormone therapy status (never, past current), pack-years of smoking (never, 0-<5 years, 5-<20 years, 20-<40 years, or ≥40 years), body mass index (kg/m^2), physical activity, regular aspirin or NSAID use (≥2 tablets/wk), alcohol consumption, and total caloric intake (kcal/day).
* Phet did not change appreciably upon mutual adjustment and/or inclusion of micronutrients (calcium, folate, vitamin D)

Figure 4.1. Study overview. A previously developed self-sampling method was used for metagenomic and metatranscriptomic sequencing of the human gut microbiome. 308 participants from the Men's Lifestyle Validation Study (MLVS) nested within the Health Professionals Follow-up Study (HPFS) cohort were recruited to assess the stability of microbiome communities, metagenomes, and metatranscriptomes. Participants provided up to four stool samples –a pair of stool samples collected 24-72 hours apart followed by collection of a second pair approximately 6 months later. DNA was extracted from all stool samples; RNA was extracted and reverse-transcribed to cDNA only from stool samples from participants who provided stool samples spanning 6 months and who did not report use of antibiotics within the last year. All samples were then sequenced using the Illumina HiSeq platform. Raw sequence data were filtered to remove low quality and human host reads and features with low overall relative abundance. Metagenomic and metatranscriptomic read data were profiled for functional and taxonomic composition using HUMAnN2 and MetaPhlAn2, respectively.

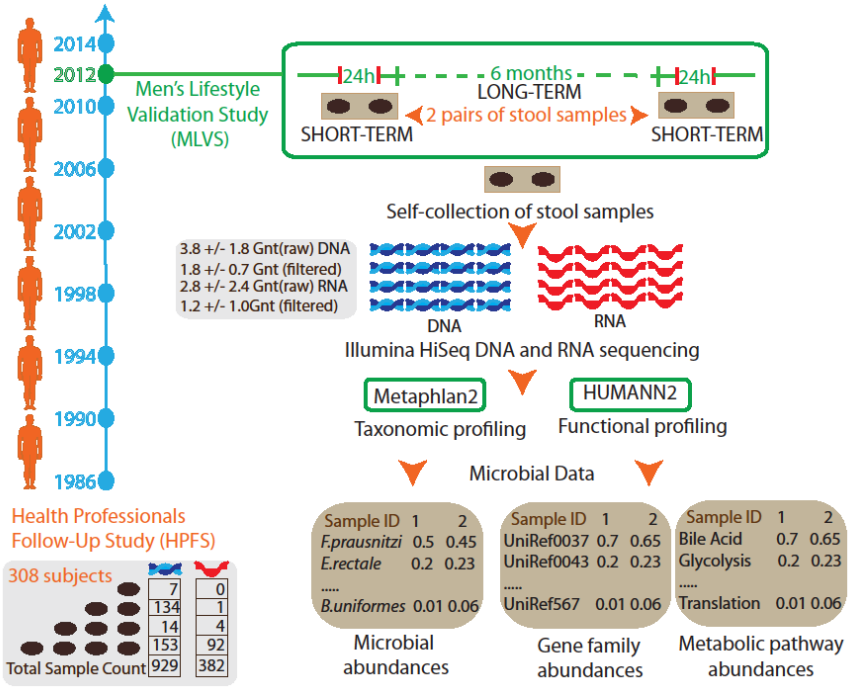


Figure 4.2. The gut metatranscriptome is stable but not highly personalized over time. A. Bray-Curtis beta-diversity scores within-people and between-people for short-term and intermediate-term intervals (6 months). Participants' gut microbiome organismal composition and metagenomes appear more similar to their own over short- and intermediate-term periods than to another person's, consistent with prior findings (HMP1). **B**. 1-Jaccard Index (fraction of shared strains) between all possible pairwise combinations between the first and other three fecal samples collected from each individual. 95% confidence intervals shown in gray.

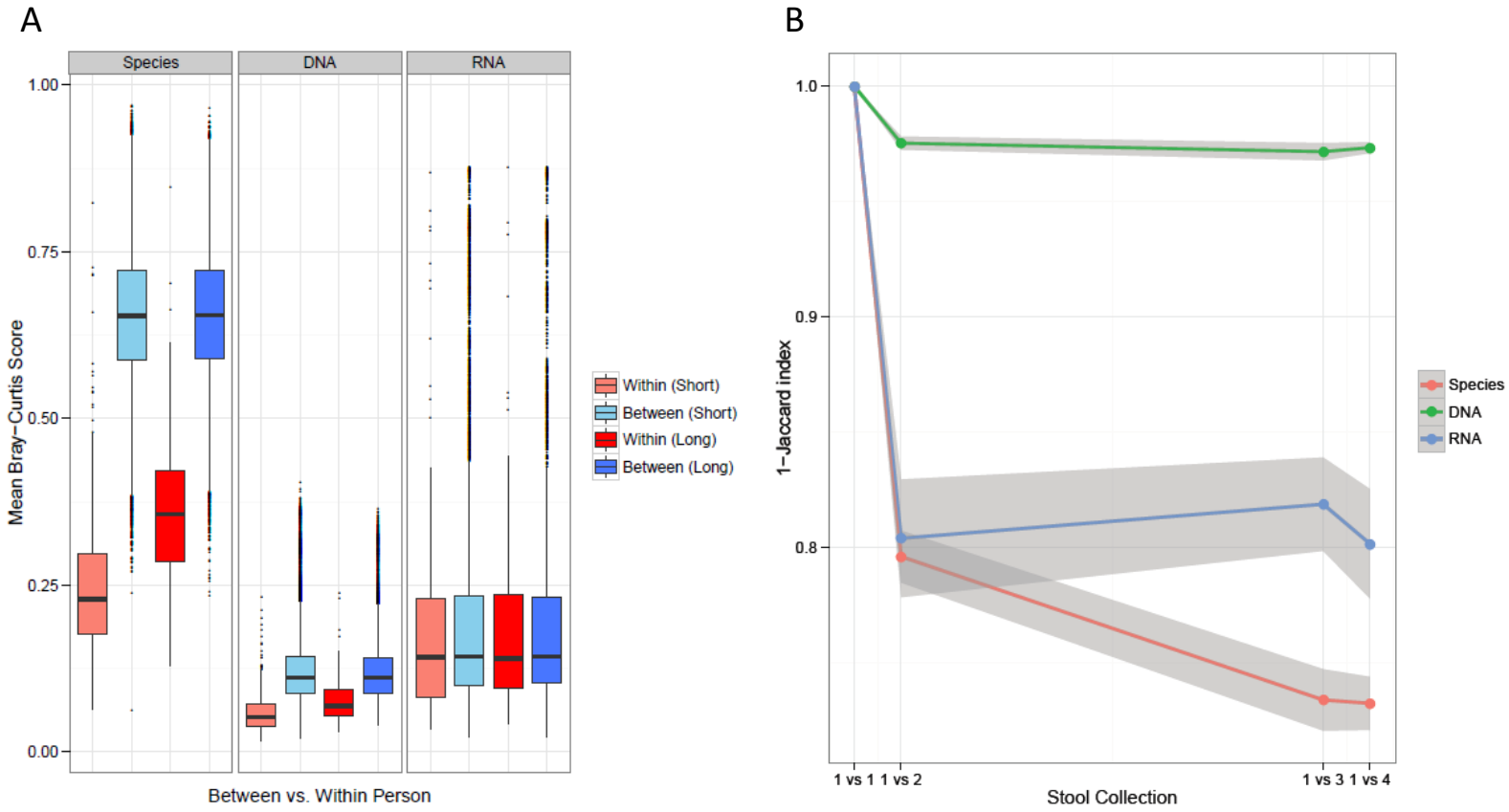


Figure 4.3. Stability of individual species, genes, and transcripts over 6 months is correlated with average baseline relative abundance and prevalence. For example, *Bacteroides uniformis* (purple) and RNA polymerase (red) are nearly 100% prevalent at baseline, highly abundant, and highly stable. Notably, however, features appear to be stable despite relative low prevalence or low abundance. For example, *Sutterella wadsworthia* (blue) is in the bottom quintile of prevalence but in the top quintile for stability. Similarly, some genes although universally prevalent at the DNA level, such as the thiaminases (yellow) are variably transcribed across time.

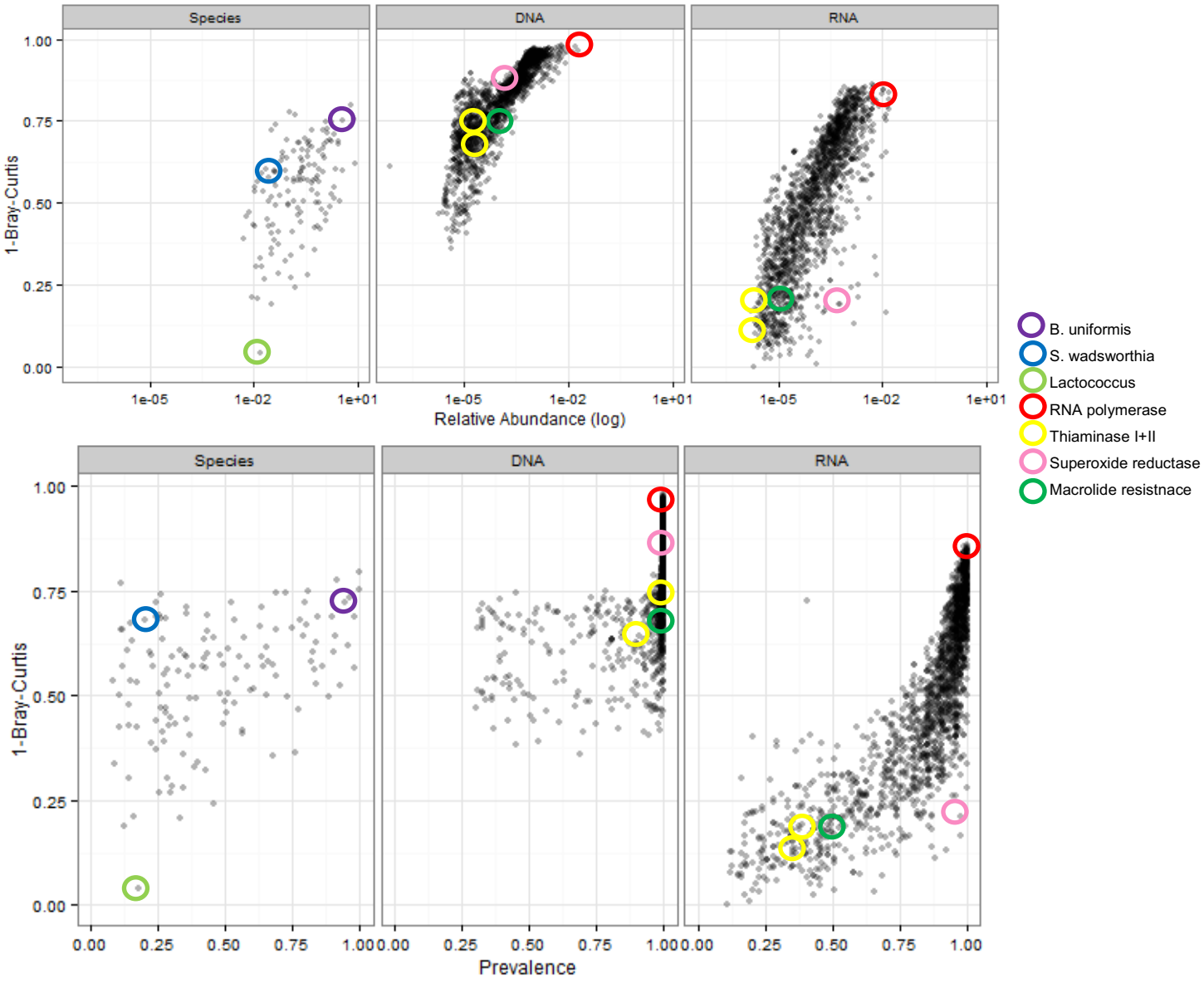


Table 4.1. Association of putative factors that may influence the stability of human gut microbiome communities and pathways over a six-month period. Linear models were used to determine which of the measured factors below predicted individuals' dissimilarity in gut microbiome communities and gut microbiome pathways between the first and the last sample. All models were adjusted by the mean total number of passing filter reads per ID.

	Species (n=160)		**DNA** (n=160)		**RNA** (n=93)	
	Parameter Estimate	p-value**	Parameter Estimate	p-value**	Parameter Estimate	p-value**
Median dissimilarity over 6 months*						
Mean Shannon index (α diversity)	**-0.096**	**0.006**	**-0.037**	**<0.0001**	-0.076	0.13
Acid-lowering medication use§	0.020	0.33	0.007	0.07	0.008	0.78
Bowel Prep¶	**0.087**	**0.01**	**0.015**	**0.04**	**0.13**	**0.03**
BMI in 2010 (kg/m^2)	0.002	0.51	0.0007	0.15	0.0035	0.33
Age	-0.0004	0.84	0.0002	0.69	0.005	0.08
Antibiotic use†						
Any Use	0.020	0.33	0.0004	0.92	NA	NA
New Use	0.033	0.34	**0.022**	**0.002**	NA	NA
Bristol Category Change‡	0.0004	0.56	-0.00001	0.52	0.0002	0.12

* As estimated by the Bray-Curtis dissimilarity score between the first and fourth stool sample.
** All models adjusted for median read depth (PF reads)
§ Use of proton pump inhibitors or H2 receptor antagonists >1 time per week in the past 2 months
¶ Any reported bowel prep (2 months prior to collection)
† Referent level is "never." "Any use" defined as any use in the last 12 months before second stool collection. "New use" defined as new use of antibiotics before second collection. We did not assess transcriptomics in individuals who reported antibiotic use.
‡ Bristol scores were assigned to three categories: 1, 2 = hard; 3, 4 = normal; 5,6,7 = soft or liquid

Figure 4.4. Relating metagenomic stability over time to metatranscriptomic stability over time. Genomic instability appears to explain a large proportion of transcriptomic instability over intermediate-term time (6 month) periods (R2=0.70); the metatranscriptome can be no more stable than the metagenome. In addition, we investigated if dominant expression of stable genes versus expression of unstable genes varied according to species. For each transcript, a dominant contributing species was identified based upon the maximal average contribution to each transcript. We then ranked the species according to the total number of genes that they dominantly expressed and (for presentation purposes) selected those contributing to 30 or more, and overlaid them on the scatterplot in color (grouped by genus). As shown, *Bacteroides* species are dominant contributors to expression of stable transcripts. In contrast, *Streptococcus* species are dominant contributors to expression of unstable transcripts, with several other species seen in between.

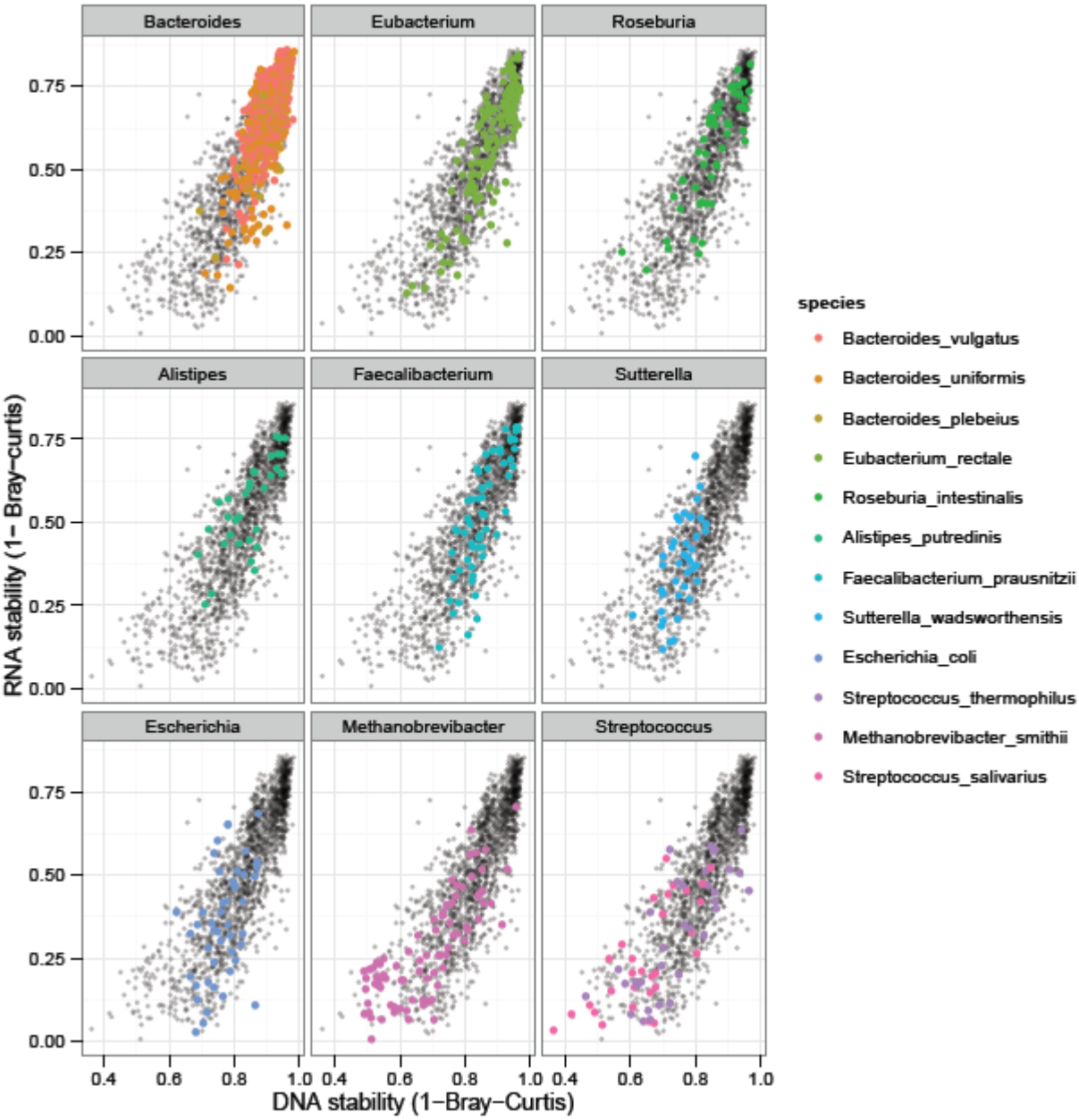
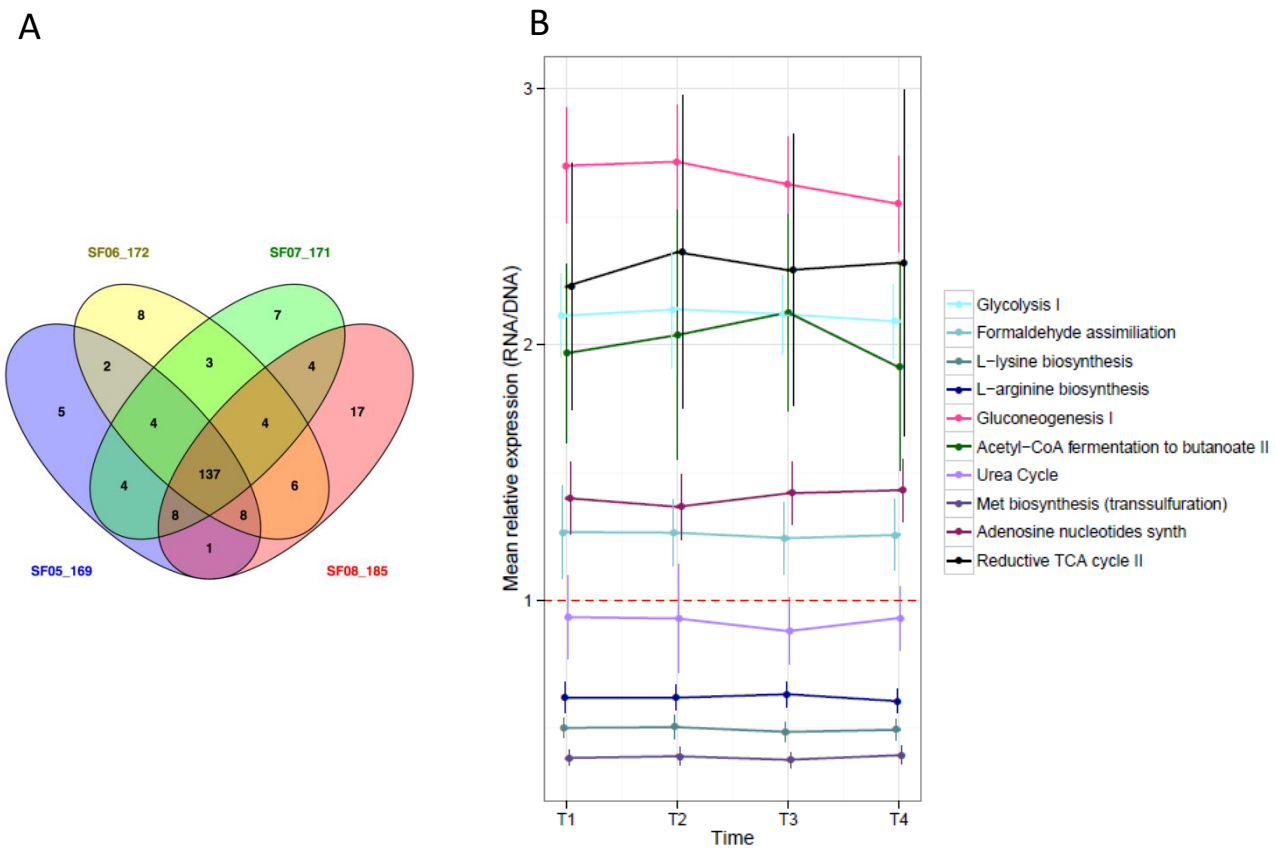
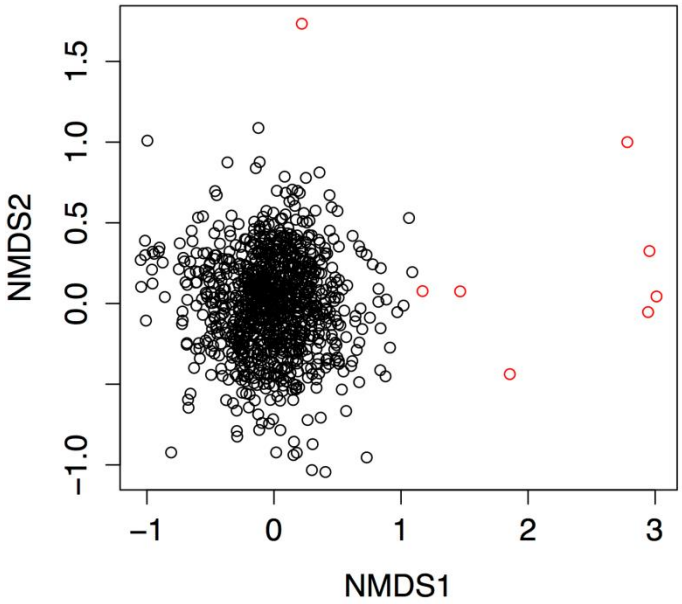


Figure 4.5. Exploring stability of the gene expression regulation. (Left) Transcriptional regulation may be stable for a large proportion of pathways over time. The middle panel shows consistency of differentially regulated pathways (mean RNA/DNA ratios ≥ 2 across at least one time point) across 4 stool samples. (Right) A subset of stably regulated pathways is highlighted, according to those with the lowest coefficient of variance of mean RNA/DNA expression across the four time points. Many of the stable pathways appear to be involved in cellular housekeeping, such as carbon metabolism (overexpressed) and amino acid synthesis (underexpressed).



Supplementary figures:

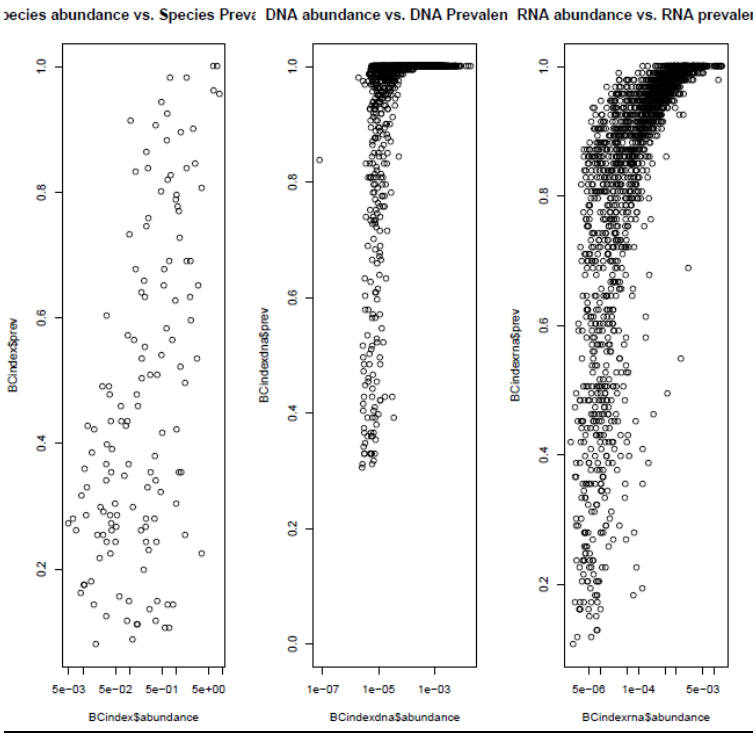
Supplementary Figure 4.1: Ordination plot of taxonomic profiles. 8 outliers (samples) were removed. 4 of these came from one person who reported a colectomy at the time of participation in the study.



Supplementary Table 4.1. Reliability in measuring relative abundance levels over short-term (24-72 hours) and long-term (~6 months) intervals for the majority of species and gene families. Genes were most highly repeatable, followed by species, and then distantly followed by RNA pathways. An ICC of <0.40 suggests poor repeatability, 0.40 to 0.75 indicates fair to good repeatability, and ≥ 0.75 indicates excellent repeatability (Rosner 2006).

	Percent of features with an ICC > 0.40	
	Short-term	Long-Term
Species (n=146)	96.8%	86.8%
DNA (n=1951)	99.9%	92.8%
RNA (n=3566)	1.3%	0.79%

Supplementary Figure 4.2. Feature abundance is highly correlated with feature prevalence.



Supplementary Table 4.2. Individuals with more stable gut microbiomes do not appear to have distinct taxa or gene families compared with those found in more variable individuals. Individuals were assigned to stability classes based on quartiles (first = stable, second and third = average, fourth = variable) of median within person Bray-Curtis distances for species data. Significant differences were observed and are presented below. However, after correction for multiple testing by the Benjamini-Hochberg method, these did not reach significance.

Species	**P-value***	**DNA EC**	**P value***
Turicibacter_sanguinis	0.0049	Carboxybiotin decarboxylase	0.0003
Alistipes_onderdonkii	0.0074	vanc D-Ala-D-Ala dipeptidase	0.0003
Dorea_formicigenerans	0.0112	malyl-CoA synthetase	0.0006
Anaerotruncus_unclassified	0.0152	morphine 6-dehydrogenase	0.0012
Peptostreptococcaceae_unclassified	0.0201	GMP synthase	0.0015
Turicibacter_unclassified	0.0219	phosphopyruvate hydratase	0.0017
Holdemania_unclassified	0.0333	butyrate kinase	0.0023
Lactococcus_lactis	0.0412	methionine sulfoxide reductase	0.0023
Desulfovibrio_piger	0.0425	uroporphyrinogen methyltransferase	0.0025
Veillonella_parvula	0.0439	methylated-DNA---[protein]-cysteine S-methyltransferase	0.0027

*According to the Kruskal-Wallis test

Supplementary Figure 4.3. Temporal variability in diversity of species, metagenomes, and metatranscriptomes matches variability in communities. To quantify amount of variability in the richness of taxa, metagenomes, and metaranscriptome over time, we estimated coefficients of variance (CV) for the alpha diversity Shannon Index for each individual. Low CV values indicate that an individual had stable alpha diversity levels, whereas high CV values indicate than an individual had relative variable alpha diversity levels over the 6 month study period.

